

Statistik

Paul Ruppen

6. Oktober 2021

Dieses Dokument wurde mit L^AT_EX gesetzt.

Im Allgemeinen werden die Abkürzungen, Schreibweisen und Redewendungen des Deutschen Instituts für Normung verwendet (DIN-Taschenbuch 202, Formelzeichen, Formelsatz, Mathematische Zeichen und Begriffe, Berlin, Beuth, 1994). In Abweichung von diesen Normen wird bei Dezimalzahlen statt des Kommas der Punkt verwendet.

Literatur

- Bleymüller, J., Gehlert, G. & Gülicher, H. (2000). *Statistik für Wirtschaftswissenschaftler*. München: Vahlen.
- Fahrmeir, L., Künstler, R., Pigeat, I. & Tutz, G. (2004). *Statistik: der Weg zur Datenanalyse*. Berlin: Springer.
- Hartung, J. (1999). *Statistik: Lehr- und Handbuch der angewandten Statistik*. München: Oldenbourg.
- Hippmann, H.-D. (1997). *Statistik für Wirtschafts- und Sozialwissenschaftler* (2. Aufl.). Stuttgart: UTB Schäffer-Poeschel.
- Wewel, M. C. (2006). *Statistik im Bachelor-Studium der BWL und VWL: Methoden, Anwendung, Interpretation*. München: Pearson.

Paul Ruppen
HES-SO Valais Wallis
Betriebsökonomie
CH-3960 Siders
<http://math.logik.ch>

Inhaltsverzeichnis

1	Einleitung	1
1.1	Begriff und Aufgaben der Statistik	1
1.2	Wichtigkeit und Bild der Statistik	2
I	Beschreibende Statistik	5
2	Variablen und Skalen	7
2.1	Variablen und Ausprägungen	7
2.2	Datenvektoren und Datenmatrizen	16
2.3	Lernziele	22
3	Univariate Häufigkeitsverteilungen	23
3.1	Absolute Häufigkeitsverteilungen	23
3.2	Relative Häufigkeitsverteilungen	34
3.3	Absolute kumulative Häufigkeitsverteilung	35
3.4	Relative kumulative Häufigkeitsverteilung	37
3.5	Empirische Verteilungsfunktion	38
3.6	Lernziele	42
4	Univariate statistische Kennwerte	43
4.1	Kennwerte der mittleren Lage	43
4.2	Streuungskennwerte	57
4.3	Zeitreihen und Indizes	79
4.4	Lernziele	89
5	Bivariate beschreibende Statistik	91
5.1	Häufigkeiten und graphische Darstellungen	91
5.2	Streudiagramme	93
5.3	Kennwerte	96
5.4	Lernziele	138
II	Kombinatorik	141
6	Etwas Kombinatorik	143
6.1	Einleitung	143
6.2	Vier Grundprobleme der Kombinatorik	143
III	Einführung in die Wahrscheinlichkeitstheorie	157
7	Einleitung	159

8	Grundbegriffe	161
8.1	Diskrete Ergebnismengen	161
8.2	Begriff der Wahrscheinlichkeit	165
8.3	Das Rechnen mit Wahrscheinlichkeiten	171
8.4	Lernziele	196
9	Zufallsvariablen	197
9.1	Zufallsvariablen und Bildmasse	197
9.2	Erwartungswert einer Zufallsvariable	201
9.3	Varianz und Standardabweichung einer Zufallsvariable	206
9.4	Produkte von Bildmassen	209
9.5	Statistiken	214
9.6	Das Schwache Gesetz der Grossen Zahlen	224
9.7	Lernziele	235
10	Diskrete Verteilungsfamilien	237
10.1	Familie der Bernoulli-Verteilungen	237
10.2	Familie der Binomialverteilungen	238
10.3	Familie der Poisson-Verteilungen	243
10.4	Familie der hypergeometrischen Verteilungen	250
10.5	Diskrete uniforme Verteilungen	256
10.6	Lernziele	269
11	Stetige Wahrscheinlichkeitsräume	271
11.1	Verteilungsfunktionen; Definition und Beispiele	272
11.2	Die Dichtefunktion	274
11.3	Kennzahlen	279
11.4	Übersicht stetige Wahrscheinlichkeitsräume	280
11.5	Lernziele	280
12	Stetige Verteilungsfamilien	283
12.1	Die Familie der stetigen, uniformen Verteilungen	283
12.2	Familie der Exponentialverteilungen	288
12.3	Lernziele	293
13	Schätzen von stetigen Verteilungen	295
13.1	Lernziele	304
14	Die Familie der Normalverteilungen	305
14.1	Einleitung	305
14.2	Eigenschaften	306
14.3	Berechnung der Werte von $P(a < X \leq b)$	309
14.4	Verteilungsfunktion	312
14.5	Zusammenhang von Normalverteilung und Binomialverteilung	313
14.6	Lernziele	321
15	P-P-Plots und Q-Q-Plots	323
15.1	P-P-Plots	323
15.2	Trendbereinigte P-P-Plots	325
15.3	Q-Q-Plots	326
15.4	Trendbereinigte Q-Q-Plots	327
15.5	Lernziele	333
16	Zufallszahlen stetiger Verteilungen	335
16.1	Lernziele	339

17 Der Zentrale Grenzwertsatz	341
17.1 Verteilung der Statistik Mittelwert	341
17.2 Anwendungen des zentralen Grenzwertsatzes	344
17.3 Lernziele	346
18 Weitere stetige Verteilungsfamilien	347
18.1 Die Familie der χ_n^2 -Verteilungen	347
18.2 Die Familie der t_n -Verteilungen	352
18.3 Die Familie der $F_{m,n}$ -Verteilungen	355
18.4 Lernziele	357
 IV Schliessende Statistik - Testtheorie	 359
19 Testen von Hypothesen	363
19.1 Problemlage	363
19.2 Nullhypothese und Alternativhypothese	364
19.3 Signifikanzniveau	365
19.4 Zweiseitige und einseitige Tests	366
19.5 p-Wert	368
19.6 Fehler erster und zweiter Art	369
19.7 Macht eines Testes	372
19.8 Lernziele	378
20 Einstichprobentests	379
20.1 z-Test	380
20.2 t-Test	384
20.3 Wilcoxon-Rang-Test	386
20.4 Vorzeichentest	392
20.5 Mächtigkeitsvergleich der vier Tests	393
20.6 Test auf abweichende Varianz	393
20.7 Anteilstest bei zwei Ausprägungen	395
20.8 Anteilstest bei mehr als zwei Ausprägungen	395
20.9 Tests mit Computerexperimenten	401
20.10 Lernziele	407
21 Verbundene Zweistichprobentests	409
21.1 t-Test	409
21.2 Wilcoxon-Test	411
21.3 Vorzeichentest	413
21.4 McNemar-Test	416
21.5 Lernziele	419
22 Unverbundene Zweistichprobentests	421
22.1 Testen von Varianzen	421
22.2 t-Test bei nicht signifikant verschiedenen Varianzen	422
22.3 t-Test bei signifikant verschiedenen Varianzen (Welch-Test)	423
22.4 Rangtest von Mann-Whitney	425
22.5 Test auf Anteile bei zwei Ausprägungen	433
22.6 Lernziele	440

23 Vertrauensintervalle bei Testwerten	441
23.1 Vertrauensintervalle für den Einstichproben z-Test	441
23.2 Vertrauensintervalle für den Einstichproben t-Test	444
23.3 Vertrauensintervalle bei poisson-verteilter Teststatistik	445
23.4 Lernziele	447
 V Bivariate Einstichprobentests	 449
24 Testen auf Unabhängigkeit	453
24.1 χ^2 -Unabhängigkeitstests	453
24.2 Fishers exakter Test	457
24.3 Vergleich mit vorgegebener Häufigkeitsverteilung	460
24.4 Lernziele	468
 25 Gamma	 469
25.1 Lernziele	473
 26 Varianzanalyse und Kruskal-Wallis-Test	 475
26.1 Einfaktorielle Varianzanalyse	475
26.2 Kruskal-Wallis-Test	482
26.3 Lernziele	486
 27 Tests für Korrelationskoeffizienten	 487
27.1 Test für den Pearson-Korrelationskoeffizienten	487
27.2 Test für die Spearman-Rangkorrelation	488
27.3 Test für die partielle Korrelation	490
27.4 Lernziele	496
 28 Regression	 497
28.1 Schätztheorie	497
28.2 Testtheorie für die lineare Regression	498
28.3 Lernziele	513
 VI Multivariate schliessende Statistik	 515
29 Multiple Regression	519
29.1 Vorbemerkungen zu Matrizen und Vektoren (Wiederholung)	519
29.2 Berechnung der Geraden	521
29.3 Testtheorie	526
29.4 Überprüfung der Voraussetzungen	531
29.5 Lernziele	536
 VII Schliessende Statistik - Schätztheorie und Stichprobentheorie	 537
30 Ein paar Bemerkungen zur Schätztheorie	539
30.1 Parametrisches Schätzen	540
30.2 Nichtparametrisches Schätzen von Verteilungen	546
30.3 Lernziele	559

31 Stichprobentheorie	561
31.1 Mittelwert und Total	564
31.2 Schätzen von Anteilen	570
31.3 Schätzen des Medians und von Quartilen	572
31.4 Geschichtete Zufallsstichproben	573
31.5 Schlussbemerkungen	577
31.6 Beweise	578
31.7 Lernziele	585
 VIII Datenerhebung	 587
32 Datenerhebung und -Bearbeitung	589
32.1 Planung der Datenerhebung	589
32.2 Methoden der Stichprobenerhebung	590
32.3 Behandlung fehlender Daten	591
32.4 Auswertung	593
 A Liste der Symbole	 595
 Literaturverzeichnis	 597
Literatur	597
 Index	 598

Kapitel 1

Einleitung

1.1 Begriff und Aufgaben der Statistik

In der *beschreibenden Statistik* geht es um die informative Darstellung von unübersichtlichem Datenmaterial. Oft liegen viele Daten vor (z.B. Verkaufszahlen pro Produkt, Spesen pro Verkaufsstelle, etc.), die unbearbeitet keinen Informationsgehalt aufweisen. Durch die Bearbeitung mit Hilfe der Methoden beschreibender Statistik reduzieren wir die komplexe Vielfalt auf einfache, übersichtliche Darstellungen und Kennzahlen, die für uns (und unser Handeln, unsere Entscheidungen) bedeutsam werden können. Beschreibende Statistik ist überall dort unverzichtbar, wo viel Zahlenmaterial anfällt, ohne dessen Analyse kein situationsgemässes (z.B. marktgemässes, betriebsinternes, staatliches, wirtschaftspolitisches) Handeln möglich ist.

Beispiel 1.1.1. *für die Unübersichtlichkeit von Datensätzen*

Die folgenden Daten (s. Tabelle 1.1.1) wurden von Immobilienmaklern in den USA verwendet, um den ungefähren Preis von Immobilien in einer Stadt abzuschätzen (kleiner Ausschnitt aus der Gesamttabelle). ◇

In der *schliessenden Statistik* geht es einerseits um Methoden, um mit möglichst wenig Aufwand Schätzungen bezüglich einer Gesamtpopulation anzustellen. Wollen wir wissen, was die Schweizer von einem bestimmten Produkt halten, wäre es sehr teuer, alle Schweizer zu befragen. Billiger und schneller ist ein Verfahren, das eine „gute“ Stichprobe aus allen Schweizern zieht. Die Verfahren der schliessenden Statistik liefern uns Kriterien für die zu ziehende Stichprobengrösse sowie die zu treffende Auswahl. Schliessende Statistik wird andererseits auch bei Experimenten eingesetzt (z.B. um die Wirkung eines Medikamentes zu überprüfen, einer Werbeaktion, von Zusätzen in der Kälberzucht, von Stoffen in der Materialforschung, usw.). Die schliessende Statistik hat fast in allen Wissensgebieten als unverzichtbare Methode Einzug gehalten.

In der *explorativen Statistik* geht es um Methoden der Klassifizierung von Daten und der Hypothesengenerierung. Die Hypothesen können anschliessend durch Methoden der schliessenden Statistik getestet werden.

Eine Statistik: konkretes Datenmaterial bezüglich eines Bereiches, in Form von Tabellen, Grafiken. (z.B. Zivilstands-Statistik der Schweiz).

Die Statistik: die Wissenschaft, die sich u.a. mit der Sammlung, Aufbereitung, der Analyse und der Interpretation von grossen Datenmengen beschäftigt. Die Statistik in diesem Sinne des Wortes ist eine Zweig der (angewandten) Mathematik.

Wir werden noch weitere Bedeutungen des Wortes „Statistik“ kennenlernen.

Ursprung des Wortes: Vielfältiges und reiches Datenmaterial fiel in der Geschichte zuerst beim Staat an. Es galt, dieses für politische Entscheidungen aufzubereiten. Das Wort „Statistik“ ist mit dem Wort „Staat“ verwandt.

Betriebliche Statistik: Die betriebswirtschaftliche Statistik analysiert betriebliches Datenmaterial, um den interessierten Personen Informationen für ihre Entscheidungen und Handlungen

PRICE	SQFT	AGE	FEATS	NE	COR	TAX
2050	2650	13	7	1	1	1639
2080	2600	.	4	1	1	1088
2150	2664	6	5	1	1	1193
2150	2921	3	6	1	1	1635
1999	2580	4	4	1	1	1732
1900	2580	4	4	1	0	1534
1800	2774	2	4	1	0	1765
1560	1920	1	5	1	1	1161
1450	2150	.	4	1	0	.
1449	1710	1	3	1	1	1010
1375	1837	4	5	1	0	1191
1270	1880	8	6	1	0	930
1250	2150	15	3	1	0	984
1235	1894	14	5	1	1	1112
1170	1928	18	8	1	1	600
1180	1830	.	3	1	0	733
1155	1767	16	4	1	0	794
739	970	4	4	0	0	541

Tabelle 1.1.1: Immobilienpreise und Eigenschaften von Wohnungen (Quelle: AlbuquerqueHomePrices; Internet: <http://www.stat.cmu.edu/DASL/>)

Erläuterungen zu den Variablen der Tabelle:

- PRICE=Sellingprice(\$hundreds)
- SQFT=Squarefeet of livingspace
- AGE=Age of home (years)
- FEATS=Number out of 11 features (dishwasher, refrigerator, microwave, disposer, washer, intercom, skylight(s), compactor, dryer, handicapfit, cableTVaccess)
- NE=Located in north east sector of city(1) or not(0)
- COR=Cornerlocation(1)or not(0)
- TAX=Annualtaxes(\$)

bereitzustellen (z.B. Verkaufszahlen von Produkten, Käuferprofile, optimale Produktionsbedingungen, Qualitäts-Kontrolle, Qualitäts-Sicherung). Die statistischen Methoden, die in der Betriebswirtschaft zur Anwendung kommen, unterscheiden sich abgesehen vom Anwendungsgebiet nicht von den Methoden, die in anderen Bereichen gebraucht werden.

1.2 Wichtigkeit und Bild der Statistik

„Wir schlagen morgens unsere Zeitung auf - und ehe das letzte Blatt gewendet ist, haben wir mehr Statistiken gesehen als Goethe oder Schiller solange sie lebten. Die Arbeitslosen werden je nach Quelle weniger oder mehr, die Krebsgefahr und das Ozonloch nehmen zu, ein Drittel aller Menschen leben unter dem Existenzminimum, ein Bundesbürger verzehrt im Jahr 5,8 Liter Speiseeis, Musiker leben länger als andere, Landluft ist gesund, Landluft ist ungesund, im Jahr 2020 gibt es 10 Milliarden Menschen auf der Welt, der Dow-Jones-Index hat 30 Punkte zugelegt, ... 25 Prozente mehr Frauen als Männer biegen falsch in Einbahnstrassen ein “ (aus *Wie lügt man mit Statistik* (Krämer, 1997) oder auch *Warum dick nicht doof macht und Genmais nicht tötet: über Risiken und Nebenwirkungen der Unstatistik* (Bauer, Gigerenzer & Krämer, 2014); empfehlenswerte, leichte Lektüre).

So omnipräsent die Statistik im medialen Alltag und in den Wissenschaften ist, so wenig schmeichelhaft ist oft das Bild, das man sich von ihr macht - auf Grund oft vorkommender missbräuchlicher und falscher Anwendung der statistischen Instrumente. Die folgenden saloppen

Sprüche zur Statistik zeugen vom Bild, das manche von der Statistik hegen:

- „Ich glaube nur an Statistiken, die ich selbst gefälscht habe.“, Sir Winston Churchill, britischer Premierminister (1874 -1965)
- „Ich könnte sogar Gott statistisch nachweisen.“, George Gallup, amerikanischer Meinungsforscher (geb. 1901)
- „Ich stehe Statistiken etwas skeptisch gegenüber. Denn laut Statistik haben ein Millionär und ein armer Kerl jeder eine halbe Million.“, Franklin Delano Roosevelt, 32. US-Präsident (1882 -1945)
- „Es gibt drei Arten von Lügen: Lügen, infame Lügen und Statistiken.“, Benjamin Disraeli, englischer Staatsmann und Schriftsteller (1804-1881)
- „Eine Statistik ist wie ein spanisches Gasthaus: Jeder findet darin, was er sucht.“, Jean-Claude Riber, deutsch-französischer Opernintendant (geb.1934)
- „Wenn man Zahlen richtig foltert, gestehen sie, was man will.“, Unbekannt
- „Mit Statistiken kann man alles beweisen, auch das Gegenteil.“, Baron James Callaghan of Cardiff, britischer Politiker (geb. 1912)
- „Die Statistik ist eine sehr gefällige Dame. Nähert man sich ihr mit entsprechender Höflichkeit, dann verweigert sie einem fast nie etwas.“, Edouard Herriot, französischer Politiker (1872 -1957)
- „Stört die Wahrheit, nütze listig die Produkte der Statistik.“, Manfred Rommel, deutscher Politiker (geb. 1928)
- „Es wird so leicht dahingesagt, Zahlen sprächen für sich; ich zweifle an dieser Feststellung. Grosse Zahlen verdecken eher, als dass sie offenbaren.“, Heinrich Büll, deutscher Schriftsteller und Nobelpreisträger (1917 -1985)
- „Wenn Sie mit den Füßen im kalten Eiswasser stehen und mit dem nackten Hintern auf einer heißen Herdplatte sitzen, dann haben Sie im statistischen Durchschnitt eine angenehme Körpertemperatur.“, Peter Paterna, deutscher Politiker (geb. 1936)
- „Warum heisst es ‚Statistischer Fragebogen‘? Statistisch gebogen werden doch erst die Antworten!“, Gabriel Laub, polnisch-tschechischer Satiriker (1928 -1998)

Allerdings kann man missbräuchliche und falsche Anwendung eines Instrumentes nicht diesem anlasten. Dafür sind allein die Anwender verantwortlich. Wichtig bei der Bewertung statistischer Ergebnisse sind Transparenz (z.B. Zugänglichkeit der anonymisierten Daten) und exakte Angabe der angewandten Verfahren.

Teil I

Beschreibende Statistik

Kapitel 2

Variablen und Skalen

2.1 Variablen und Ausprägungen

In der beschreibenden Statistik legen wir für eine bestimmte Untersuchung jeweils eine Menge von Objekten als *Grundmenge* (oder Population, manchmal auch „Statistische Masse“ genannt) fest. Wir bezeichnen alles als Objekt, was Eigenschaften hat. Ein wesentlicher Begriff der Statistik ist der Begriff der *statistischen Variable*. Beispiele für statistische Variablen sind die Variable „Geschlecht“, die Variable „Einkommen“, die Variable „Ausgaben für Unterhaltungselektronik“, die Variable „Zufriedenheit mit einem Produkt“. Variablen weisen sogenannte *Ausprägungen* auf, die auch „Werte“ genannt werden. So weist die Variable „Geschlecht“ die Ausprägungen „männlich“ und „weiblich“ auf. Die Variable „Zufriedenheit mit einem Produkt“ kann z.B. die Ausprägungen „sehr zufrieden“, „zufrieden“, „müßig zufrieden“ und „unzufrieden“ aufweisen.

Formal gesehen sind statistische Variablen *Zerlegungen* der untersuchten Population. Eine Zerlegung (= Partition) einer Menge G ist eine Menge Z von nicht-leeren Mengen A_i (= Äquivalenzklassen) so dass gilt:

$$\begin{aligned} (i) : A_i &\subset G \\ (ii) : \bigcup_{i=1}^n A_i &= G \\ (iii) : A_i \cap A_j &= \emptyset \text{ für } i \neq j. \end{aligned}$$

Definition 2.1.1. „statistische Variable“: Sei eine Grundmenge G gegeben: Jede Zerlegung der Grundmenge G nennen wir „(statistische) Variable“. Die (nichtleeren) Elemente (= Äquivalenzklassen) von Z nennen wir „Ausprägungen der Variable“ oder „Werte der Variablen“ (die Elemente von Z sind Mengen).

Variablen sowie deren Ausprägungen benennen wir gewöhnlich mit Namen.

Beispiel 2.1.2. Population $G = \{\text{Anton, Beat, Anna, Ruth, Christine}\}$
Mögliche Variable: $V = \{\{\text{Anton, Beat}\}, \{\text{Anna, Ruth, Christine}\}\}$ mit den Ausprägungen $\{\text{Anton, Beat}\}$ und $\{\text{Anna, Ruth, Christine}\}$. Wir benennen die Variable V mit „Geschlecht“, die Ausprägung $\{\text{Anton, Beat}\}$ mit „männlich“ und die Ausprägung $\{\text{Anna, Ruth, Christine}\}$ mit „weiblich“. Die Grundmenge G wird im Beispiel in zwei Äquivalenzklassen zerlegt: die Menge der Männer und die Menge der Frauen. Wir beschreiben die Situation auch mit der folgenden Ausdrucksweise: „die Variable 'Geschlecht' weist zwei Ausprägungen auf, nämlich 'männlich' und 'weiblich'“ oder mit „die Variable 'Geschlecht' weist zwei Werte auf, nämlich 'männlich' und 'weiblich'“. (s. Abbildung 2.1.1)

◇

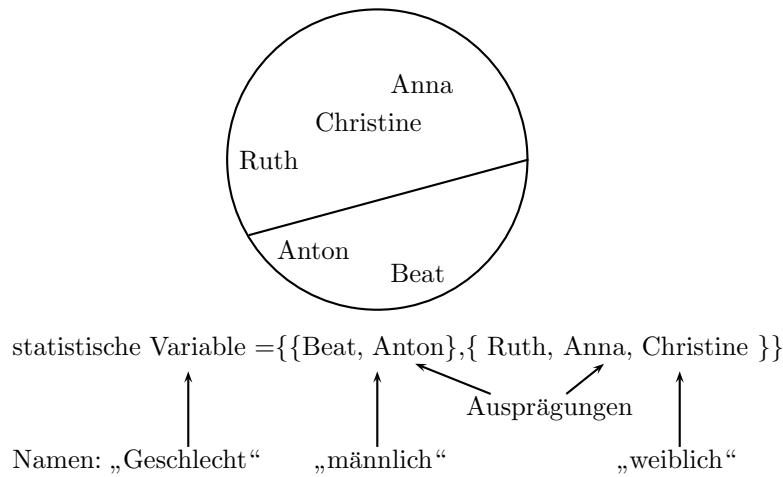


Abbildung 2.1.1: Graphische und mengentheoretische Darstellung einer Variablen; Namensgebung für die Variable und deren Ausprägungen

Der Variable V entspricht eine Äquivalenzrelation, die wir wie folgt erhalten:

$$R_V := \bigcup_{i=1}^n (A_i \times A_i)$$

(A_i sind Äquivalenzklassen von Z , n ist die Anzahl der Ausprägungen der Variable). Dabei ist

$$A_i \times A_i = \{(x, y) : x, y \in A_i\}$$

das kartesische Produkt von A_i und A_i und (x, y) ein geordnetes Paar. Am obigen Beispiel erhalten wir:

$R_V = \{(\text{Anton}, \text{Anton}), (\text{Beat}, \text{Beat}), (\text{Anton}, \text{Beat}), (\text{Beat}, \text{Anton}), (\text{Ruth}, \text{Ruth}), (\text{Anna}, \text{Anna}), (\text{Christine}, \text{Christine}), (\text{Ruth}, \text{Anna}), (\text{Anna}, \text{Ruth}), (\text{Anna}, \text{Christine}), (\text{Christine}, \text{Anna}), (\text{Ruth}, \text{Christine}), (\text{Christine}, \text{Ruth})\}$. Diese Äquivalenzrelation können wir sprachlich beschreiben, indem wir in „hat das (die, der) selbe ... wie“ den Namen der Variable einsetzen. Im Beispiel erhalten wir: „hat das selbe Geschlecht wie“.

Umgekehrt erzeugt jede Äquivalenzrelation auf eine Menge G eine Zerlegung dieser Menge. Wir wählen der Reihe nach alle Komponente x in einem der geordneten Paare der Äquivalenzrelation und bilden $M_x := \{y : (x, y) \in R\}$. Die Menge der M_x bildet eine Zerlegung von G . Äquivalenzrelation und Zerlegung legen sich gegenseitig genau fest.

Beispiel 2.1.3. Der Ausdruck „hat dieselbe Augenfarbe wie“ legt eine Äquivalenzrelation fest. Unterscheiden wir nur 4 Augenfarben (blau, grau, grün, braun), so erzeugt diese Äquivalenzrelation eine Zerlegung der betrachteten Grundmenge in 4 Äquivalenzklassen - sofern es Objekte gibt, die unter die vier Klassen fallen. Wir sagen in diesem Beispiel, dass die Variable „Augenfarbe“ vier Ausprägungen hat, nämlich, „blau“, „grau“, „grün“ und „braun“. Dabei müssen wir alle Augenfarben diesen vier Kategorien zuordnen können! $\diamond$

Für jede Variable muss gelten, da eine Variable eine Zerlegung ist: jedes Objekt der betrachteten Grundmenge muss unter genau eine der Ausprägungen der Variable fallen und kein Objekt darf unter zwei Ausprägungen der Variable fallen. So ist z.B. „ist männlich“ (als biologisches Kriterium) i.A. zweifelsfrei anwendbar. Demgegenüber wäre bezüglich „ist männlich“ (als Kriterium, dass manche Männer vor anderen „auszeichnet“) kaum zweifelsfrei entscheidbar, ob es auf ein Objekt anwendbar ist.

In der Praxis fehlen oft Daten und deshalb sind die entsprechenden Objekte bezüglich der betrachteten Variablen keinen Ausprägungen zuzuordnen. Dies ist kaum zu verhindern. Wichtig ist, dass vom Konzept her mit wirklichen Variablen gearbeitet wird.

Die Forderung, dass jedes Objekt der Grundmenge unter eine der Eigenschaften der Variable fallen muss, wird manchmal durch die Ausprägung „andere“ erfüllt (z.B. „ist verheiratet“, „ist ledig“, „andere“).

Die Grundmenge G und die Ausprägungen der Objekte von G werden gewöhnlich durch Tabellen dargestellt:

Beispiel 2.1.4. Variable „Bezug von Überstunden“ mit den Ausprägungen „Bezug durch Geld“ und „Bezug durch Ferien“.

Wir nehmen an, wir würden die folgenden Daten erhalten (s. Tabelle 2.1.1):

Person	Bezug von Überstunden
1	Bezug durch Geld
2	Bezug durch Ferien
3	Bezug durch Ferien
4	Bezug durch Ferien
5	Bezug durch Geld
6	Bezug durch Geld
7	Bezug durch Ferien
8	Bezug durch Geld
9	Bezug durch Geld
10	Bezug durch Ferien
11	Bezug durch Ferien
12	Bezug durch Geld
13	Bezug durch Geld
14	Bezug durch Ferien
15	Bezug durch Ferien
16	Bezug durch Geld
17	Bezug durch Geld

Tabelle 2.1.1: Beispiel für eine Urliste von Daten

Eine solche Tabelle nennen wir „Urliste der Daten“. In der ersten Spalte werden die Objekte der Grundmenge durchnummeriert. In der Kopfzeile rechts der ersten Spalte stehen die Namen der Variablen. Im Beispiel gibt es nur einen Variablennamen. Jeder Variable entspricht eine Spalte. In der Spalte unter den Variablennamen stehen Namen der Ausprägungen, zu denen die Objekte bezüglich der jeweiligen Variable gehören.

Aus der Urliste der Daten kann man folgende Angaben herauslesen: die Gesamtpopulation ist: $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17\}$ (statt uns mit Namen auf die Personen zu beziehen, verwenden wir hier Zahlen. Dabei sind nicht die Zahlen Elemente der Grundmenge, sondern die durch sie bezeichneten Personen).

Für die Relation R_V „bezieht Überstunden auf dieselbe Art wie“ erhalten wir die folgende Zerlegung: $V = \{\{1, 5, 6, 8, 9, 12, 13, 16, 17\}, \{2, 3, 4, 7, 10, 11, 14, 15\}\}$

Die Schnittmenge der Elemente der Zerlegung ist leer. Die Vereinigungsmenge dieser beiden Elemente ist mit der Gesamtpopulation identisch. $\diamond$

Objekte können verschiedene Eigenschaften aufweisen und je nach Eigenschaften können sie bezüglich unterschiedlicher Variablen mit verschiedenen Methoden gemessen werden. Wir betrachten in der Folge die für die Statistik relevanten Messarten.

2.1.1 Nominal skalierte Variablen

Oft verwenden wir Ausdrücke, um Objekte zu klassifizieren, ohne weitere Relationen wie z.B. Größenverhältnisse zwischen ihnen auszudrücken. Solche Ausdrücke stehen für sogenannte qualitative Merkmal oder Eigenschaft. Z.B. „ist weiblich“, „ist männlich“, „ist ledig“, „ist verheiratet“,

„ist geschieden“, „fährt mit dem Auto zur Arbeit“, „geht zu Fuss zur Arbeit“, „ist zuverlässig“, „ist krank“, „ist unberechenbar“, „ist zerbrechlich“, „ist teuer“, „ist ein Kunde“, „ist ein Mitarbeiter“.

Eine Variable, welche mit ihren Ausprägungen eine Grundmenge nur klassifiziert, ohne dass die Objekte so gemessen werden können, dass weitere Relationen zwischen den Objekten der verschiedenen Ausprägungen berücksichtigt werden, wird in der Statistik „qualitative Variable“, „nominal skalierte Variable“ oder „kategoriale Variable“ genannt, z.B. Variable „berufliche Tätigkeit“ mit den Ausprägungen „ist berufstätig“, „ist nicht berufstätig“.

Messen besteht im Zuordnen von Zahlen zu Objekten. Eine primitive und nicht sehr aussagekräftige Art des Messens wird oft bei qualitativen Variablen vorgenommen, indem den Objekten einer Ausprägung eine spezifische Zahl zugeordnet wird: jedem Objekt einer spezifischen Ausprägung wird dieselbe Zahl zugeordnet und Objekten verschiedener Ausprägungen werden andere Zahlen zugeordnet. Man spricht bei dieser primitiven Form des Messen oft von „Kodierung einer Variablen“. z.B. In einer Tabelle wird statt „ist männlich“ die Zahl 1 gesetzt, statt „ist weiblich“ die Zahl 2. Die Zahlen drücken somit folgendes aus: weisen zwei Objekte dieselbe Zahl auf, so gehören sie zur selben Ausprägung. Weisen zwei Objekte verschiedene Zahlen auf, so gehören sie zu verschiedenen Ausprägungen. Somit spielt nur die mathematische Eigenschaft der Gleichheit und Ungleichheit für Zahlen bei nominalskalierten Variablen eine Rolle. Die übrigen mathematischen Eigenschaften von Zahlen spielen bei dieser Art des Messen keine Rolle: der Umstand, dass $2 > 1$, drückt nichts über das Verhältnis zweier Objekte aus, denen im Rahmen einer nominal skalierten Variable 2 und 1 zugeordnet ist. Messskalen mit den eben erwähnten Eigenschaften werden „Nominalskalen“ genannt.

Formaler können wir eine Nominalskala (bezüglich der Variable V) wie folgt definieren: Sei R_V die durch die Variable V definierte Äquivalenzrelation auf die Grundmenge G . Wir können nun definieren:

Definition 2.1.5. Eine Abbildung $f : G \rightarrow \mathbb{R}$ ist eine Nominalskala auf die Grundmenge G bezüglich der Variable V genau dann, wenn:

1. $f(x) = f(y)$ genau dann, wenn $(x, y) \in R_V$

(R_V ist die durch die Variable V festgelegte Äquivalenzrelation auf G . R_V ist eine Relation auf G genau dann, wenn alle Elemente von G mindestens einmal als Komponente in einem Element der Relation auftauchen).

Wir werden folgende Redewendungen verwenden: Sei f eine Skala auf G bezüglich V . Dann sagen wir, dass V eine f -skalierte Variable ist. Ist f eine Nominalskala, sprechen wir im Besonderen von einer „nominalskalierten Variable“.

Aus der Definition folgt, dass zwei Objekten x und y durch f dieselbe Zahl zugeordnet wird, wenn x und y zur selben Äquivalenzklasse (= Ausprägung) von V gehören. x und y werden durch f verschiedene Zahlen zugeordnet, wenn x und y nicht zur selben Äquivalenzklasse von V gehören.

Beispiel 2.1.6. Wir führen das obige Beispiel (Bezug von Überstunden) fort: Wir legen als Funktion, die die Bedingung „ $f(x) = f(y)$ genau dann, wenn $(x, y) \in R_V$ “, erfüllt fest:

für $x \in \{1, 5, 6, 8, 9, 12, 13, 16, 17\} : f(x) = 1$

für $x \in \{2, 3, 4, 7, 10, 11, 14, 15\} : f(x) = 2$

In tabellarischer Form (1 = Bezug durch Geld, 2 = Bezug durch Freizeit, s. Tabelle 2.1.2):

Diese Tabelle ist mit der oben gelieferten eng verwandt. Es wurden nur die Bezeichnungen für die Ausprägungen der Variable durch die entsprechenden Skalenwerte ersetzt. Auch solche Tabellen werden wir „Urliste der Daten“ nennen. $\diamond$

2.1.2 Ordinal skalierte Variablen

Während wir bei nominalskalierten Variablen auf Grund der Zahlen, die den Elementen einer Ausprägung zugeordnet sind, nur wissen können, ob zwei Objekte einer selben Ausprägung zugehören oder nicht, stellen wir bei ordinalskalierten Variablen zwischen den Objekten verschiedener

Person	Bezug von Überstunden
1	1
2	2
3	2
4	2
5	1
6	1
7	2
8	1
9	1
10	2
11	2
12	1
13	1
14	2
15	2
16	1
17	1

Tabelle 2.1.2: Beispiel für eine Urliste von Daten mit Skalenwerten

Ausprägungen eine Ordnung (= ordo) her. Wir betrachten die Variable „Qualität eines Produktes“ mit den Ausprägungen „sehr gut“, „gut“, „zufriedenstellend“, „nicht zufriedenstellend“. Eine Grundmenge von Produkten würde den Ausprägungen entsprechend klassifiziert. Zwischen den klassifizierten Produkten besteht nun eine Ordnung: ein als sehr gut klassifiziertes Produkt ist besser als ein als gut, zufriedenstellend oder nicht zufriedenstellend klassifiziertes Produkt. Ein als gut klassifiziertes Produkt ist besser als ein als zufriedenstellend oder nicht zufriedenstellend klassifiziertes Produkt. Ein als zufriedenstellend klassifiziertes Produkt ist besser als ein als nicht zufriedenstellend klassifiziertes Produkt.

Auch bei ordinalskalierten Variablen muss die Zuordnung zu den Ausprägungen der Variable eindeutig sein. Zudem muss jedes Objekt der Grundmenge durch die Ausprägungen einer Variable klassifiziert werden.

Messen erfolgt bei ordinalskalierten Variablen ebenfalls durch die Zuordnung von Zahlen zu den Objekten der verschiedenen Ausprägungen. Wie bei nominalskalierten Variablen müssen die Zahlen die Zugehörigkeit zu denselben Ausprägungen durch dieselbe Zahl ausdrücken. Verschiedene Zahlen drücken die Zugehörigkeit zu verschiedenen Ausprägungen aus. Im Gegensatz zu nominalskalierten Variablen müssen bei ordinalskalierten Variablen die Zahlen jedoch noch die Ordnung zwischen den Objekten der verschiedenen Ausprägungen ausdrücken. Im Beispiel könnten wir etwa den Objekten, die als sehr gut klassifiziert wurden, die Zahl 5 zuordnen, den Objekten, die als gut klassifiziert wurden, die Zahl 3, den Objekten, die als zufriedenstellend klassifiziert wurden, die Zahl 2.5, den Objekten, die als nicht zufriedenstellend klassifiziert wurden, die Zahl 1. Die mathematische Relation „>“ zwischen den Zahlen spiegelt nun die Relation „besser sein als“. Jede Zuordnung von Zahlen, die diese Relation spiegelt, stellt eine angemessene Messung der Objekte einer ordinalskalierten Variable dar. So wäre etwa die folgende Zuordnung zu den Objekten der oben verwendeten Ausprägungen ebenso angemessen: 1000, 100, 10, 1 oder 1000, 3, 2, 1, usw.

Wir stellen das eben gesagte nun noch etwas systematischer dar. Bei Objekten, die man mit einer Ordinalskala messen kann, muss eine Relation T definierbar sein, für die folgende Eigenschaften gelten:

Definition 2.1.7. Die Abbildung $f : G \rightarrow \mathbb{R}$ ist eine Ordinalskala auf die Grundmenge G bezüglich der Variable V genau dann, wenn es eine Relation T auf G gibt, so dass gilt:

1. $(x, y) \in R_V$ genau dann, wenn $f(x) = f(y)$.
2. T ist transitiv und asymmetrisch.
3. $(x, y) \in T$ genau dann, wenn $f(x) > f(y)$

(R_V ist die durch die Variable V festgelegte Äquivalenzrelation auf G).

Bemerkung 2.1.8. • eine Relation ist asymmetrisch, wenn für alle $(x, y) \in R$ gilt, $(y, x) \notin R$. Somit ist z.B. die Relation des Vaterseins oder des „Strikt-Grösser-Seins-Als-Jemand-Anders“ asymmetrisch

- Wird die obige Definition von einer Variable V erfüllt, sagen wir, dass V eine ordinalskalierte Variable ist.
- z.B. für „ T “ „ist besser als“, für „ R “ „ist gleich gut wie“. Wir legen eine beliebige Funktion f fest, für die gilt: f ordnet x genau dann eine höhere Zahl zu als y , wenn x in der Relation T zu y steht (z.B. wenn x besser als y ist). Die Zahl, die x zugeordnet ist, ist genau dann gleich der Zahl, die y zugeordnet ist, wenn x in der Relation R_V zu y steht (z.B. wenn x gleich gut ist wie y).

◇

Beispiel 2.1.9. Die Angestellten einer Unternehmung werden befragt, ob Sie die Personalführung gut finden. Auf die Frage „Wie finden Sie die Personalführung“ konnten Sie zwischen den folgenden Fragen wählen: „Sehr gut“, „Gut“, „Recht gut“, „Schlecht“. Wir nehmen an, 20 Personen wären befragt worden. Wir ordnen den Personen, die die Personalführung sehr gut finden eine 4 zu, denen, die sie gut finden eine 3, denen die sie recht gut finden eine 2 und den restlichen eine 1. Wir nehmen an, wir hätten folgendes Resultat erhalten (s. Tabelle 2.1.3):

◇

2.1.3 Metrisch skalierte Variablen

Wenn die körperlichen Grössen von Personen in Metern gemessen wird, so stellt die Variable „Körpergrösse“ eine metrisch skalierte Variable dar. Zeitmessung, Gewichtsmessung, Geschwindigkeitsmessung, usw. erfolgen im Allgemeinen mit metrischen Skalen. Metrische Skalen zeichnen sich durch folgende Eigenschaften aus:

Definition 2.1.10. Die Abbildung $f : G \rightarrow \mathbb{R}$ ist eine metrische Skala auf die Grundmenge G bezüglich der Variable V genau dann, wenn es Relationen T und $\circ$ auf G gibt, so dass gilt:

1. $(x, y) \in R_V$ genau dann, wenn $f(x) = f(y)$.
2. T ist transitiv und asymmetrisch.
3. $(x, y) \in T$ genau dann, wenn $f(x) > f(y)$
4. $f(x \circ y) = f(x) + f(y)$ (Additionsprinzip oder Linearität)

(R_V ist die durch die Variable V festgelegte Äquivalenzrelation auf G).

Bemerkung 2.1.11. • Wir sagen, dass V eine metrisch skalierte Variable ist.

- „ $\circ$ “ ist irgend eine empirische Operation des Aneinanderfügens. Beim Messen von Längen besteht dieses etwa darin, auf einer Geraden zwei Gegenstände aneinanderzufügen. Beim Gewichtsmessen besteht dieses darin, auf eine Waage zu einem Gewicht ein weiteres hinzuzufügen. Für f muss dabei laut obiger Bedingung z.B. gelten: Wenn f dem Objekt x die Zahl 5 zuordnet und f dem Objekt y die Zahl 3 zuordnet, dann muss f den zusammengefügteten Objekten $x \circ y$ die Zahl 8 zuordnen. Konkret: wenn ein Gegenstand 3 Kilo wiegt und ein weiterer

Personen	Personalführung (4 = „sehr gut“; 3 = „gut“; 2 = „recht gut“; 1 = „schlecht“)
1	4
2	3
3	4
4	1
5	2
6	3
7	3
8	4
9	3
10	4
11	3
12	1
13	2
14	2
15	1
16	2
17	3
18	3
19	4
20	3

Tabelle 2.1.3: Mögliche Daten zu Beispiel 11.1.3 in Form einer Urliste.

Gegenstand 5 Kilo wiegt, so müssen sie zusammen 8 Kilo wiegen. Wenn ein Gegenstand 3 Meter lang ist und ein weiterer 5 Meter, so sind sie auf einer Geraden zusammengefügt 8 Meter lang.

- Offenbar kann man Geld metrisch messen. Da Geld etwas Abstraktes ist, stellt sich die Frage, wie hier eine entsprechende empirische Operation aussehen könnte. Wir können in der Tat ein empirisches Äquivalent für Geld finden, auf das eine solche Operation definiert werden kann. Wir können z.B. für Geldbeträge die entsprechende Menge an neuen Einfrankenstücken betrachten und dann die Gewichtsmessung verwenden. Für diese gibt es eine empirische Operation, welche die obige Anforderung erfüllt. Dies überträgt sich dann auf das Geld. Beachten Sie allerdings: obwohl man Geld metrisch messen kann, gilt dies nicht für den Nutzen, den Geld für eine Person darstellen kann!
- Bezüglich einer Variablen, für die eine metrische Skala definierbar ist, kann immer auch eine Ordinalskala definiert werden. Eine Ordinalskala ist jedoch weniger informativ. Nur die „>“ Relation ist auf ihr sinnvoll interpretierbar. Bei metrischen Skalen ist jedoch zusätzlich die Operationen der Addition interpretiert, bei manchen metrischen Skalen zusätzlich die Operation der Multiplikation. So ist es sinnvoll zu sagen, eine Stange x sei dreifach so lang wie eine andere y (dies ist genau dann der Fall, wenn gilt: $f(x) = 3f(y)$). Haben wir eine ordinalskalierte Variable vor uns mit den Ausprägungen „sehr gut“, „gut“, „genügend“, „ungenügend“, so dass wir den entsprechenden Objekten 1000, 10, 5 und 1 zuordnen, hat es hingegen keinen Sinn zu sagen, ein als „sehr gut“ klassifiziertes Objekt x sei 1000 mal besser als ein als ungenügend klassifiziertes Objekt y , weil $f(x) = 1000f(y)$.
- Metrisch skalierte Variablen können stetiger oder diskreter Natur sein. Stetige oder kontinuierliche Variablen kommen etwa bei der Längenmessung vor: in der Theorie - aber nicht

in der Praxis - sind zwischen zwei Werten überabzählbar unendlich stetiger oder diskreter Natur viele Werte möglich (überabzählbar unendlich = gleichviele wie das entsprechende, beidseitig beschränkte Intervall reeller Zahlen enthält). Bei metrisch, stetigen Variablen tritt pro Ausprägung oft nur ein Objekt auf. Bei diskreten Variablen sind in einem beidseitig beschränkten Intervall nur endlich viele Skalenwerte oder abzählbar unendlich viele Werte verfügbar (abzählbar unendlich = gleichviele wie das entsprechende Intervall rationaler Zahlen). Werden nur natürliche Zahlen zugeordnet, sprechen wir von einer Absolutskala, die damit eine Unterart der diskret metrischen Skala darstellt. Absolutskalen kommen etwa vor, wenn man die Anzahl der Eigentumswohnungen oder die Anzahl der Kinder einer Population erfasst. Streng genommen gibt es nur diskrete Skalen, da wir in der Wirklichkeit nur auf endlich viele Stellen genau messen können. Bei Geld etwa messen wir höchstens zwei Stellen nach dem Komma. Trotzdem würden wir Einkommen als eine stetig metrisch skalierte Variable betrachten - ausser wir würden dieses z.B. in ganzen Tausendern messen. Allgemein betrachtet man eine metrisch skalierte Variable dann als stetig, wenn es viele Ausprägungen hat und die Anzahl der Skalenwerte fast oder ganz der Anzahl der Daten entspricht.

- *Bei stetig metrisch skalierten Variablen enthalten die Äquivalenzklassen jeweils nur ein Objekt, wenn wir genügend genau messen.*
- *Bei den metrischen Skalen sind weitere Unterscheidungen üblich (Intervallskalen, z.B. Temperaturmessung; Verhältnisskalen, z.B. Gewichtsmessung, Längenmessung. Verhältnisskalen haben einen absoluten Nullpunkt und die Multiplikation ist sinnvoll interpretierbar. z.B. ein Tisch ist dreimal so lang wie ein anderer). Es handelt sich um in unserem Rahmen überflüssige Unterscheidungen.*
- *Zu beachten ist, dass bezüglich der Skalenwerte die Operation der Addition und der Multiplikation unabhängig vom Skalenniveau möglich ist - sie ist aber nur bezüglich metrischer Skalenniveaus sinnvoll. So kann man zwar rechnerisch den Mittelwert bezüglich einer ordinalskalierten Variable berechnen. Dies macht aber wenig Sinn. Entscheidend ist deshalb nicht die Durchführbarkeit der Operationen, sondern das Auffinden einer empirischen Operation des Zusammenfügens, die den Operationen auf der Ebene der Zahlen entsprechen.*
- *Wer legen fest: die Ordinalskala ist eine höhere Skala als die Nominalskala und die metrische Skala ist eine höhere Skala als die Ordinalskala. Es gilt: Objekte, die auf einer höheren Skala messbar sind, sind auch auf einer tieferen Skala messbar. Dies ergibt sich durch den Umstand, dass jede höhere Skala alle Voraussetzungen von tieferen Skalen aufweist (siehe Zusammenstellung unten). Die Umkehrung gilt nicht. Nicht alle Objekte, die bezüglich einer Variable auf einer Nominalskala messbar sind, sind bezüglich dieser Variable ordinal oder metrisch messbar. Nicht alle Objekte, die bezüglich einer Variablen ordinal messbar sind, sind bezüglich dieser Variable metrisch messbar.(s. Tabelle 2.1.4)*
- *Eine Messmethode wird objektiv genannt, wenn verschiedene Personen, die die Methode kennen, bei deren Anwendung zum selben Resultat gelangen. So ist etwa das Messen einer Tischlänge eine objektive Angelegenheit - wenn man eine nicht allzu grosse Genauigkeit verlangt. Die Benotung von Deutschaufsätzen oder von Kooperationsfähigkeit im Betrieb ist demgegenüber kaum objektiv, da verschiedene Spezialisten zu unterschiedlichen Beurteilungen kommen. Bezüglich der Bewertung von Deutschaufsätzen ist die mangelnde Objektivität durch empirische Untersuchungen gut bestätigt. Beachten Sie, dass Objektivität in der Statistik nicht als Übereinstimmung von Meinungen und Fakten interpretiert wird, sondern als die Produzierbarkeit identischer Resultate durch verschiedene Spezialisten (Kenner der Materie).*
- *Eine Messmethode wird valide genannt, wenn durch die Methode das gemessen wird, was man messen will. Die Messung der Schuhgrösse in Zentimeter ist recht objektiv. Verwenden wir die Schuhgrösse jedoch zur Messung der Intelligenz der Schuhträger, so ist die Methode kaum valide. Wir messen mit der Methode nicht das Gewünschte. Für ein weiteres Beispiel siehe in den Lösungen zu den Übungen.*

Eine Abbildung $f : G \rightarrow \mathbb{R}$ ist eine Skala auf die Grundgesamtheit G bezüglich der Variable V genau dann, wenn es Relationen T und $\circ$ auf G gibt, so dass:	
(1) $(x, y) \in R_V$ genau dann, wenn $f(x) = f(y)$ (R_V ist die durch die Variabale V festgelegte Äquivalenzrelation auf G).	nominale
(2) T ist transitiv und asymmetrisch. (3) $(x, y) \in T$ genau dann, wenn $f(x) > f(y)$. (4) $(x, y) \in T$ oder $(y, x) \in T$ genau dann, wenn $f(x) \neq f(y)$.	ordinale
(5) $f(x \circ y) = f(x) + f(y)$ (Additionsprinzip oder Linearität).	metrische

Tabelle 2.1.4: Zusammenstellung über den Zusammenhang zwischen den drei behandelten Skalen

◇

Beispiel 2.1.12. Eine Unternehmung untersucht die Kundenstruktur. Dazu werden die Umsätze der Unternehmung mit den Kunden erhoben (in Franken). Wir erhalten z.B. die folgende Tabelle 2.1.5.

Unternehmung	Umsatz in Fr.
1	54907.43
2	25187.51
3	61348.31
4	25001.79
5	84718.87
6	62441.16
7	48867.09
8	74686.72
9	36554.36
10	42621.76
11	46823.83
12	24619.05
13	14209.21
14	37737.41
15	71026.32
16	73230.15
17	119375.88
18	57365.19

Tabelle 2.1.5: Beispiel für eine metrisch skalierte Variable; Urliste von Daten

Die durch „hat gleichviel Umsatz wie“ beschriebene Relation ist eine Äquivalenzrelation R_V auf G

und sie induziert eine Zerlegung von G . Die Elemente dieser Zerlegung enthalten jeweils nur ein Element.

Die durch „hat mehr Umsatz als“ beschriebene Relation T ist asymmetrisch und transitiv. Wir ordnen hier unmittelbar die Skalenwerte zu.

Zwei Kunden haben zusammen als Umsatz die Summe der Einzelumsätze. ◇

2.2 Datenvektoren und Datenmatrizen

Die Spalte der Urliste der Daten mit den Skalenwerten einer Variablen können wir auch als ein n -Tupel oder einen Vektor von Daten auffassen. Solche n -Tupel nennen wir künftig auch „Datenvektor“. Formellere Definition:

Definition 2.2.1. „Datenvektor der Grundmenge G “: Sei $|G| = n$. Sei $y = (a_1, \dots, a_n)$ ein n -Tupel der n Objekte der Grundmenge G so dass $a_i \neq a_j$ für $i \neq j$. $x = (f(a_1), \dots, f(a_n))$ wird „Skalenwertvektor (der Grundmenge)“ genannt.
(G = Grundmenge, a_i Elemente der Grundmenge. f irgend eine Messskala).

Definition 2.2.2. Redewendungen: Statt „Skalenwertvektor“ verwenden wir gewöhnlich „Datenvektor“ oder „Datensatz“. Die Skalenwerte $f(a_i)$ nennen wir gewöhnlich „Daten“. Statt „Skalenwerte einer nominal- oder ordinalskalierten oder metrisch skalierten Variable“ verwenden wir auch die kürzeren Redewendungen „nominal- oder ordinalskalierte Daten oder metrisch skalierte Daten“.

Betrachten wir eine Urliste von Daten mit mehreren Variablen, so können wir die Tabelle als eine Matrix von Daten auffassen. Wir nennen diese oft „Datenmatrix“.

2.2.1 Übungen

1. Wählen Sie zwei Variablen (eine mit zwei und eine mit drei Ausprägungen) und klassifizieren Sie damit die Menge der Mitstudentinnen und -studenten, indem Sie eine Urliste der Daten erstellen. Überprüfen Sie, ob die obigen Kriterien einer Variable beachtet wurden.
2. Überprüfen Sie, ob die Variable „Freizeitaktivitäten“ mit den vier Ausprägungen „führt Ski“, „führt Snowboard“, „führt Schlittschuh“, „andere“ die obigen Kriterien an eine Variable erfüllt: Wenn wir im Rahmen einer Marketingstudie Datenmaterial bezüglich des Freizeitverhaltens sammeln, wie müssten wir dann vorgehen, wenn wir den obigen Kriterien genügen möchten?
3. Definieren Sie für Variablen unter Übung 2.2.1.1 eine Nominalskala und stellen Sie die dadurch erreichten Zuordnungen in Tabellenform dar.
4. Legen Sie bezüglich Ihrer Klasse eine Variable fest, bezüglich der eine Ordinalskala definierbar ist. Bestimmen Sie die Ausprägungen der Variable. Bestimmen Sie die Elemente der Mengen, die den verschiedenen Ausprägungen zugeordnet sind. Bestimmen Sie die beiden Relationen T und R , die zur Festlegung der Ordinalskala dienen könnten. Überprüfen Sie, ob T transitiv und asymmetrisch ist. Überprüfen Sie, ob R eine Äquivalenzrelation ist. Legen Sie ein passende Skala f fest.
5. Legen Sie eine stetig und eine diskret metrisch skalierte Variable für die Studentinnen und Studenten Ihrer Klasse fest und überprüfen Sie die Voraussetzungen der Skala am Beispiel.
6. Wir überprüfen mit Hilfe eines Testes (Kenntnis von Wörtern), wer wie gut Französisch kann. Dabei möchten wir die Französischkenntnisse mit Hilfe der Anzahl gewussten Wörter messen. Ist diese Messmethode objektiv? Ist diese Messmethode valide? Die Ausprägungen der Variable „Französisch können“ sei die Anzahl Wörter, die jemand kann. Es ergeben sich

Person	Anzahl Wörter
1	5
2	1
3	8
4	2
5	9
6	10
7	3
8	4
9	8
10	1
11	2
12	11
13	3
14	4
15	2
16	8
17	7
18	5

Tabelle 2.2.6: Beispiel für Urliste der Daten einer (scheinbar) metrisch skalierten Variable.

die Daten der Tabelle [2.2.6](#).

- Geben Sie bezüglich des Beispiels mit den Immobiliendaten an, welche Skalenniveaus bezüglich der einzelnen Variablen vorliegen.
- Untersuchen Sie, ob Notenskalen in der Schule als metrische, ordinale oder nominale Skalen zu betrachten sind.

2.2.2 Lösungen und Lösungshinweise

- Sei $g(x)$ die Körpergrösse von x . Sei K die Menge der Studentinnen und Studenten x für die gilt: $g(x) \leq 1.70m$. Sei T die Menge der Studentinnen und Studenten x für die gilt: $g(x) > 1.70m$. Sei G die Gesamtmenge der Studentinnen und Studenten. Es gilt nun offenbar: $K \cup T = G$ und $K \cap T = \emptyset$.

Sei

C die Menge der Studentinnen und Studenten x für die gilt: x hat grüne Augen,

B die Menge der Studentinnen und Studenten x für die gilt: x hat blaue Augen und

A die Menge der Studentinnen und Studenten x für die gilt, x hat eine andere Augenfarbe.

Es gilt nun offensichtlich: $C \cup B \cup A = G$ und $C \cap B = \emptyset, C \cap A = \emptyset, B \cap A = \emptyset$

Wir nehmen an, wir würden folgende Daten erhalten, die wir in der Urliste [2.2.7](#) der Daten festhalten.

Im Beispiel gibt es zwei Variablen, die je eine Spalte beanspruchen.

Die Gesamtpopulation ist damit: $G = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$

Person	Augenfarbe	Körpergröße
1	grün	< 1.70
2	blau	< 1.70
3	grün	< 1.70
4	andere	> 1.70
5	andere	> 1.70
6	grün	< 1.70
7	blau	> 1.70
8	blau	< 1.70
9	andere	< 1.70
10	grün	> 1.70
11	andere	< 1.70
12	andere	> 1.70
13	grün	> 1.70

Tabelle 2.2.7: Beispieldaten für eine Urliste von Daten zweier Variablen mit drei respektive zwei Ausprägungen.

Der zur Relation R passende Ausdruck ist „hat dieselbe Augenfarbe wie“. Durch R_V wird die Zerlegung der Grundmenge in:

$V = \{\{1, 3, 6, 10, 13\}, \{2, 7, 8\}, \{4, 5, 9, 11, 12\}\}$ festgelegt.

Die Elemente der Zerlegung sind paarweise disjunkt und deren Vereinigungsmenge ist mit der Gesamtpopulation identisch.

- Die Kriterien werden nicht erfüllt, da jemand Schlittschuhfahren und Snowboard fahren kann. Die Schnittmengen der beiden Ausprägungen wären somit nicht leer. Man könnte z.B. unterscheiden zwischen „Treibt Sport“, „Treibt nicht Sport“ oder z.B. „fährt nur Schlittschuh“, „fährt nur Snowboard“, „fährt nur Schlittschuh und Snowboard“, „fährt nur Ski“, „fährt nur Ski und Snowboard“, „fährt nur Ski und Schlittschuh“, „andere“.
- Wir führen das Lösungsbeispiel mit der Augenfarbe und der Körpergröße fort: Wir legen für die obigen Beispieldaten und die Variable „Augenfarbe“ folgende Funktion, welche die Bedingung (1) der Definition 2.1.5 erfüllt, fest:

für $x \in \{1, 3, 6, 10, 13\} : f(x) = 1$

für $x \in \{2, 7, 8\} : f(x) = 2$

für $x \in \{4, 5, 9, 11, 12\} : f(x) = 3$

Durch die Relation R_V , zu der der sprachliche Ausdruck „hat dieselbe Körpergröße wie“ passt, wird die Zerlegung der Grundmenge in: $\{\{1, 2, 3, 6, 8, 9, 11\}, \{4, 5, 7, 10, 12, 13\}\}$ festgelegt. Die Elemente der Zerlegung sind paarweise disjunkt und deren Vereinigungsmenge ist mit der Gesamtpopulation identisch. Wir legen für die Variable „Körpergröße“ die folgende Funktion, welche die Bedingung (1) der Definition 2.1.5 erfüllt fest:

für $x \in \{1, 2, 3, 6, 8, 9, 11\} : f(x) = 1$

für $x \in \{4, 5, 7, 10, 12, 13\} : f(x) = 2$

In tabellarischer Form erhält man die Tabelle 2.2.8.

Am Beispiel sehen wir, dass dieselbe Grundmenge durch verschiedene Variablen zerlegt werden kann. In diesem Fall überlappen sich die Zerlegungen im Allgemeinen. (z.B. Variable Geschlecht mit den Ausprägungen „ist männlich“ und „ist weiblich“; Variable Sportverhal-

Person	Augenfarbe	Körpergrösse
1	1	1
2	2	1
3	1	1
4	3	2
5	3	2
6	1	1
7	2	2
8	2	1
9	3	1
10	1	2
11	3	1
12	3	2
13	1	2

Tabelle 2.2.8: Beispiel für eine Urliste von Daten mit zwei Variablen und Angabe der Ausprägungen durch Skalenwerte

ten mit den Ausprägungen „treibt Sport“ und „treibt keinen Sport“. Die beiden Zerlegungen würden nur zusammenfallen, wenn alle Männer und nur die Männer Sport treiben, oder wenn alle Frauen und nur die Frauen Sport treiben).

- Wir möchten wissen, wie zufrieden die Personen einer Klasse mit einer spezifischen Unterrichtsstunde sind. Dazu wird ein Fragebogen ausgeteilt mit den Ausprägungen „unzufrieden“, „zufrieden“, „sehr zufrieden“. Wir sammeln den Fragebogen ein und erstellen z.B. die folgende Aufstellung (Urliste der Daten 2.2.9):

Person	Zufriedenheitsgrad
1	zufrieden
2	zufrieden
3	unzufrieden
4	unzufrieden
5	sehr zufrieden
6	sehr zufrieden
7	sehr zufrieden
8	sehr zufrieden
9	zufrieden
10	zufrieden

Tabelle 2.2.9: Beispiel für Urliste der Daten einer ordinalskalierten Variable

Die Gesamtpopulation ist damit: $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

Wir erhalten eine Zerlegung V durch die Relation R „hat denselben Zufriedenheitsgrad wie“:
 $V = \{\{3, 4\}, \{1, 2, 9, 10\}, \{5, 6, 7, 8\}\}$.

Für die Zerlegung gilt: die Elemente der Zerlegung sind paarweise disjunkt und deren Vereinigungsmenge ist mit der Gesamtpopulation identisch.

Als T legen wir fest „ist zufriedener als“ (T ist transitiv und asymmetrisch).

Wir können nun die folgende Skala oder Funktion f festlegen:

für $x \in \{3, 4\} : f(x) = 1$

für $x \in \{1, 2, 9, 10\} : f(x) = 2$

für $x \in \{5, 6, 7, 8\} : f(x) = 3$

Offenbar erfüllt, wie man leicht kontrollieren kann, f die Anforderungen an eine Ordinalskala.

In tabellarischer Form ergibt sich die Tabelle 2.2.10.

Person	Skalenwert
1	2
2	2
3	1
4	1
5	3
6	3
7	3
8	3
9	2
10	2

Tabelle 2.2.10: Beispiel für Urliste von Daten einer ordinalskalierten Variable (Angabe von Skalenwerten).

5. stetig, metrisch skaliert. Wir untersuchen z.B. die Körpergrösse der Studentinnen und Studenten. Wir nehmen an, wir erhalten die folgenden Daten (s. Tabelle 2.2.11):

Person	Grösse
1	176.25
2	179.59
3	163.31
4	189.95
5	161.55
6	169.23
7	188.37
8	183.16
9	178.41
10	190.11
11	183.14
12	171.65
13	179.31
14	175.57
15	180.03

Tabelle 2.2.11: Beispieldaten für metrisch skalierte Variable. Urliste der Daten.

Wir haben die Relation „ R “ mit „gleich gross wie“ und die Relation „ T “ mit „grösser als“. R ist bei der vorgenommenen Interpretation symmetrisch, transitiv und reflexiv (Ist x gleichgross wie y und y gleichgross wie z , dann ist x gleichgross wie z . Ist x gleichgross wie y , so ist y gleichgross wie x . x ist gleichgross wie x). Es gibt keine zwei Personen, die genau dieselbe Grösse haben. Bei stetig, metrisch skalierten Variablen fallen bei genügend genauen Messungen kaum je zwei Objekte in dieselbe Kategorie.

„grösser als“ ist offensichtlich transitiv und asymmetrisch (wenn x grösser als y ist und y grösser als z , dann ist x grösser als z ; wenn x grösser als y ist, dann ist y nicht grösser als

x).

Durch das Messen der Grüssen ordnen wir den Personen unmittelbar Skalenwerte zu. Wir können entsprechend die Ausprägungen mit den Skalenwerten identifizieren. Für die Funktion f gilt offensichtlich: $f(x) = f(y)$ genau dann, wenn $(x, y) \in R$

$f(x) > f(y)$ genau dann, wenn $(x, y) \in T$

$f(x \circ y) = f(x) + f(y)$ (hat eine Person z.B. eine Grüsse von 1.76 und eine andere von 1.82, dann haben Sie zusammen - wenn man sie übereinanderstellt - eine Grüsse von 3.58).

6. 1 Wort können die Elemente der Menge $\{2, 10\}$
- 2 Wörter können die Elemente der Menge $\{4, 11, 15\}$
- 3 Wörter können die Elemente der Menge $\{7, 13\}$
- 4 Wörter können die Elemente der Menge $\{8, 14\}$
- 5 Wörter können die Elemente der Menge $\{1, 18\}$
- 7 Wörter können die Elemente der Menge $\{17\}$
- 8 Wörter können die Elemente der Menge $\{3, 9, 16\}$
- 9 Wörter können die Elemente der Menge $\{5\}$
- 10 Wörter können die Elemente der Menge $\{6\}$
- 11 Wörter können die Elemente der Menge $\{12\}$

Die Relationen sind: „ T “ für „kann mehr Wörter als“ und „ R “ für „kann ebenso viele Wörter wie“.

Die paarweisen Schnittmengen dieser Mengen sind leer. Die Vereinigungsmenge dieser Menge ergibt die Gesamtpopulation. Zudem ist R transitiv, symmetrisch und reflexiv (Kann x ebenso viele Wörter wie y und y ebenso viele Wörter wie z , so kann x ebenso viele Wörter wie z . Kann x ebenso viele Wörter wie y , so kann y ebenso viele Wörter wie x . x kann ebenso viele Wörter wie x).

T ist offenbar transitiv: Wenn eine Person mehr Wörter kann als eine zweite und diese zweite mehr Wörter kann als eine dritte, so kann die erste mehr Wörter als die dritte. T ist auch asymmetrisch: Es gibt keine Person, die mehr Wörter kann als eine andere, so dass letztere mehr Wörter kann als die erste.

Wir ordnen den Objekten der Elemente der Zerlegung als Skalenwerte die Anzahl der gekonnten Wörter zu:

$$\begin{array}{ll}
 \text{für } x \in \{2, 10\} : f(x) = 1 & \text{für } x \in \{4, 11, 15\} : f(x) = 2 \\
 \text{für } x \in \{7, 13\} : f(x) = 3 & \text{für } x \in \{8, 14\} : f(x) = 4 \\
 \text{für } x \in \{1, 18\} : f(x) = 5 & \text{für } x \in \{17\} : f(x) = 7 \\
 \text{für } x \in \{3, 9, 16\} : f(x) = 8 & \text{für } x \in \{5\} : f(x) = 9 \\
 \text{für } x \in \{6\} : f(x) = 10 & \text{für } x \in \{12\} : f(x) = 11
 \end{array}$$

Es gilt nun: ist $f(x) > f(y)$, dann gilt $(x, y) \in T$, d.h. x kann mehr Wörter als y .
und ist $f(x) = f(y)$, dann gilt $(x, y) \in R$, d.h. x kann ebenso viele Wörter wie y .

Sei $\circ$ die Operation des Zusammenzählens von gekonnten Wörtern. Dann gilt: $f(x) + f(y) = f(x \circ y)$

D.h. Hat eine Person 3 gekonnte Wörter und eine andere 4 gekonnte Wörter, haben sie zusammen 7 gekonnte Wörter. Eine Tabelle mit den Skalenwerten statt der Anzahl Wörter unterscheidet sich nicht von der obigen Tabelle.

Bezüglich des Beispiels kann man ein paar kritische Fragen stellen: ist wirklich die Kenntnis aller Wörter gleichgewichtig für die Messung der Französischkenntnisse. Sollten die Wörter unterschiedliches Gewicht haben, müsste man das Gewicht metrisch messen können, sonst sind die Französischkenntnisse mit Hilfe von Wörterkenntnissen nicht mal ordinal messbar

(die Kenntnis der Wörter a, b und c könnten dann weniger wertvoll sein, als die Kenntnisse des Wortes d. Dann sagt jedoch die Anzahl der gekannten Wörter als solche nichts mehr über die Französischkenntnisse aus). Das Zählen der gewussten Wörter ist deshalb nicht eine valide Methode, um die Französischkenntnisse zu messen - sie wäre allerdings objektiv.

7. 1.PRICE=Sellingprice(\$hundreds) (metrisch skaliert, und zwar diskret, da nur Hunderter berücksichtigt werden)
 - 2.SQFT=Squarefeet of livingspace (stetig metrisch skaliert)
 - 3.AGE=Age of home (years) (diskret, metrisch skaliert)
 - 4.FEATS=Number out of 11 features (dishwasher, refrigerator, microwave, disposer, washer, intercom, skylight(s), compactor, dryer, handicapfit, cableTVaccess (diskret, metrisch skaliert, sofern man davon ausgeht, dass alle Maschinen oder Eigenschaften dasselbe Gewicht haben)
 - 5.NE=Located in north east sector of city(1) or not(0) (nominal skaliert; wenn „located in north east sector of city“ einen höheren Wert ergibt als nicht dort zu liegen, kann man auch von einer Ordinalskala sprechen).
 - 6.COR=Cornerlocation(1)or not(0) (nominal skaliert; siehe jedoch Bemerkung zu 5)
 - 7.TAX=Annualtaxes(\$) (stetig, metrisch skaliert).
8. Notenskalen stellen am ehesten Ordinalskalen dar. Für eine metrische Skala müsste man eine empirische Operation $\circ$ angeben können, so dass gilt: $f(x \circ y) = f(x) + f(y)$. Eine solche Operation scheint jedoch nicht angebar zu sein. Andererseits drücken Notenskalen nicht nur eine Nominalskala aus, da im Allgemeinen eine Ordnungsrelation T definierbar ist, die transitiv und asymmetrisch ist. Wir interpretieren T als „ist besser als“. und R als „ist gleich gut wie“. Damit werden die Bedingungen bezüglich der Skala erfüllt. Andererseits ist zu beachten, dass bei Noten oft Punkte zusammengezählt werden, ohne zu garantieren, dass diese Punkte wirklich dasselbe Gewicht haben. Entsprechend wären Noten nicht einmal ordinalskaliert, wie ein Beispiel zeigen kann: Wir nehmen an, für Aufgaben werden je ein Punkt verteilt, die bezüglich der effektiven Leistung das folgende Gewicht hätten:

Aufgabe	1	2	3	4	5	Summen
Punkt	1	1	1	1	1	5
Gewicht	0.75	0.5	1.5	1.75	1	5.5
Student 1 Punkte	1	1	0	0	1	3
Student 1 Gewichte	0.75	0.5	0	0	1	2.25
Student 2 Punkte	0	0	1	1	0	2
Student 2 Gewichte	0	0	1.5	1.75	0	3.25

Student 1 würde eine bessere Note erhalten, obwohl er bezüglich effektiver Leistung weniger erreicht hätte. Es geht um ein Validitätsproblem (s. o.). Es gibt Methoden, um das Problem zu lösen (z.B. Rasch-Modell, IRT-Methoden).

2.3 Lernziele

- Bei konkreten Variablen unterscheiden können, ob sie nominal, ordinal oder metrisch (diskret oder stetig) skaliert sind.
- Die Begriffe „Variable“, „Ausprägung einer Variable“, „Skalenwert“ korrekt verwenden können.
- Die Begriffe „Objektivität“ und „Validität“ bezüglich Messverfahren definieren können.
- Auf eine Grundmenge bezüglich einer spezifischen Variable eine angemessene Skala definieren können.

Kapitel 3

Univariate Häufigkeitsverteilungen und Verteilungsfunktionen

Wie in der Einleitung festgestellt, geht es in der beschreibenden Statistik darum, Datenmengen mit Verfahren so zu behandeln, dass sie aussagekräftig werden. Eine erste Möglichkeit, dies zu erreichen, besteht in Häufigkeitsverteilungen und Verteilungsfunktionen. Dabei werden wir absolute und relative Häufigkeitsverteilungen, absolute und relative kumulative Häufigkeitsverteilungen sowie Verteilungsfunktionen unterscheiden und diese Funktionen bezüglich der verschiedenen Skalenniveaus analysieren. In diesem Kapitel werden wir nur jeweils eine Variable (= univariat) betrachten und gemeinsame Häufigkeitsverteilungen (bivariate, multivariate Verteilungen) ausser acht lassen.

3.1 Absolute Häufigkeitsverteilungen

Bei einer Häufigkeitsverteilung ordnen wir den Skalenwerten der Ausprägungen einer Variablen die Häufigkeit des Auftretens von Objekten aus der Grundmenge mit dem entsprechenden Skalenwert zu. Etwas formaler können wir definieren ($Bild(f) = Bild$ der Abbildung f):

Definition 3.1.1. k ist eine absolute Häufigkeitsverteilung einer Grundmenge G bezüglich der f -skalierten Variable V genau dann, wenn

$$k : Bild(f) \rightarrow \mathbb{N}$$

$$k(x) := |\{z \mid f(z) = x \text{ und } z \in G\}|$$

Dabei ist $|\{z \mid f(z) = x \text{ und } z \in G\}|$ die Anzahl der Objekte z , so dass die Skala f diesen die Zahl x zuordnet. Eine Häufigkeitsverteilung ist somit eine Funktion von der Menge der den Ausprägungen einer Variablen zugeordneten Skalenwerten zur Anzahl der Objekte einer spezifischen Ausprägung (s. Abbildung [3.1.1](#)).

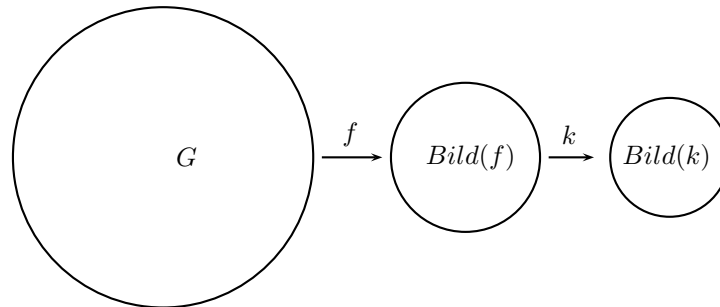


Abbildung 3.1.1: Häufigkeitsverteilung als Funktion auf das Bild der Skala f (G = Grundmenge; f = Skala auf G ; $Bild(f)$ = Bild von f ; k = Häufigkeitsverteilung; $Bild(k)$ = Bild von k).

Statt die Objekte pro Ausprägung können wir ebenso gut die Anzahl identischer Skalenwerte des Datenvektors in der Urliste der Daten zählen und den Skalenwerten diese Anzahlen zuordnen.

Beispiel 3.1.2. *Bezug von Überstunden (1 = Bezug durch Geld, 2 = Bezug durch Freizeit; s. Tabelle 3.1.1)*

Person	Bezug durch Überstunden (1 = Bezug durch Geld; 2 = Bezug durch Ferien)
1	1
2	2
3	2
4	2
5	1
6	1
7	2
8	1
9	1
10	2
11	2
12	1
13	1
14	2
15	2
16	1
17	1

Tabelle 3.1.1: Beispiel für die Berechnung der Häufigkeitsverteilung durch das Zählen von Skalenwerten

Die Menge der Skalenwerte ist $\{1, 2\}$. Der Skalenwert 1 kommt 9 mal im Datenvektor der Urliste der Daten vor, der Skalenwert 2 kommt 8 mal vor. Wir erhalten den folgenden Graphen der Häufigkeitsverteilung: $\{(1, 9), (2, 8)\}$.

Da Häufigkeitsverteilungen Funktionen $D \rightarrow \mathbb{N}$ mit $D \subset \mathbb{R}$ sind, können wir diese in Koordinatennetzen darstellen. Fürs Beispiel würden wir erhalten (s. Abbildung 3.1.2):

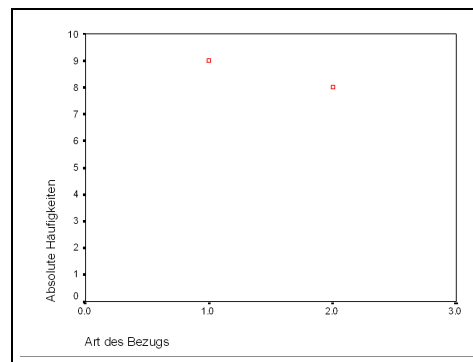


Abbildung 3.1.2: Graphische Darstellung der Häufigkeitsverteilung im Koordinatennetz des Beispiels 3.1.2)

Für Häufigkeitsverteilungen sind weitere graphische Darstellungen üblich (z.B. Balkendiagramme, Säulendiagramme, Kreisdiagramme, Histogramme). Bei den Balkendiagrammen ordnen wir jedem Skalenwert einer Variablen einen Balken zu, dessen Länge proportional zur Anzahl der Vorkommen einer bestimmten Ausprägung gewählt wird (in Excel werden Balkendiagramme „Säulendiagramme“ genannt, während „Balkendiagramm“ für vertikal liegende Stäbe verwendet wird). Fürs Beispiel würden wir erhalten (s. Abbildung 3.1.3):

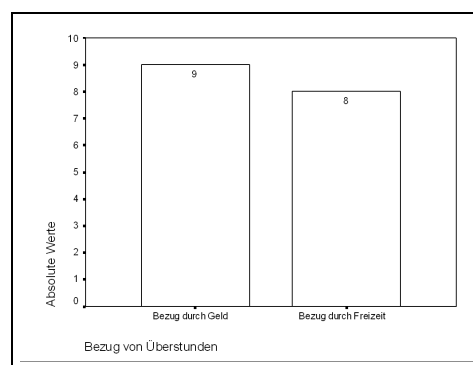


Abbildung 3.1.3: Graphische Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.1

Statt der Skalenwerte stehen im Balkendiagramm oft die Bezeichnungen für die Ausprägungen. Bei Kreisdiagrammen werden Kreisabschnitte so gewählt, dass die Proportionen zwischen den Häufigkeitsanteilen den Flächenproportionen der Kreisabschnitte entsprechen (s. Abbildung 3.1.4):

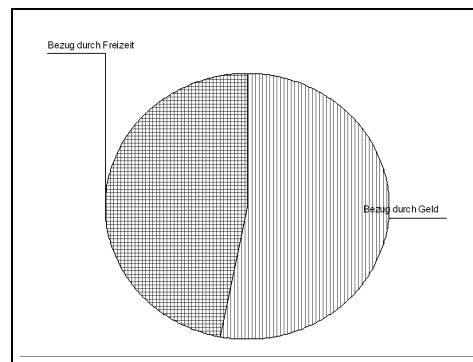


Abbildung 3.1.4: Häufigkeitsverteilung der Daten der Tabelle 3.1.1 mit Hilfe eines Kreisdiagramms

◇

Beispiel 3.1.3. Zwei verschiedene Häufigkeitsverteilungen erhalten wir, wenn wir zwei Variablen betrachten:

Person	Augenfarbe	Körpergrösse
1	1	1
2	2	1
3	1	1
4	3	2
5	3	2
6	1	1
7	2	2
8	2	1
9	3	1
10	1	2
11	3	1
12	3	2
13	1	2

Tabelle 3.1.2: Urliste von Daten zweier Variablen

Die Menge der Skalenwerte ist der ersten Variable (Augenfarbe) ist: $\{1, 2, 3\}$. Die Auszählung ergibt den folgenden Graphen der Häufigkeitsverteilung: $g_{k_1} = \{(1, 5), (2, 3), (3, 5)\}$

Variable „Körpergrösse“: Die Menge der Skalenwerte ist $\{1, 2\}$. Die Auszählung ergibt den folgenden Graphen der Häufigkeitsverteilung: $g_{k_2} = \{(1, 7), (2, 6)\}$ (siehe Abbildungen 3.1.5, 3.1.6, 3.1.7 und 3.1.8). ◇

Damit haben wir nominalskalierte Variablen betrachtet, wobei man die Variable „Körpergrösse“ auch als ordinalskalierte Variable mit nur zwei Ausprägungen betrachten könnte. Ordinalskalierte und metrisch skalierte Variablen werden im Prinzip gleich behandelt. Es sind jedoch zusätzlich zwei Punkte zu beachten:

(1) Während nominalskalierte Variablen im Allgemeinen wenige Ausprägungen haben (Ausnahme z.B. Variable „Blumenart“), können ordinalskalierte oder diskret metrisch skalierte Variablen viele Ausprägungen haben. Dies kann zur Folge haben, dass pro Skalenwert oft nur mehr ein Objekt oder sehr wenige Objekte gezählt werden. Daraus resultieren dann wenig aussagekräftige Häufigkeitsverteilungen. Bei stetig metrisch skalierten Variablen wird für jeden Skalenwert im Allgemeinen nur ein Objekt gezählt, was dazu führt, dass eine Häufigkeitsverteilung gewöhnlich gleich viele Balken enthält wie die Anzahl der Daten und damit keinen Informationswert aufweist. In

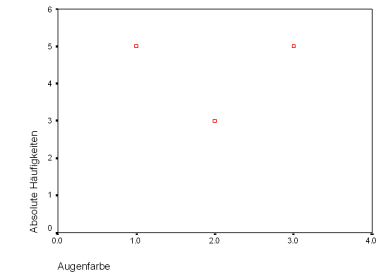


Abbildung 3.1.5: Graphische Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.2, Spalte 1, im Koordinatennetz.

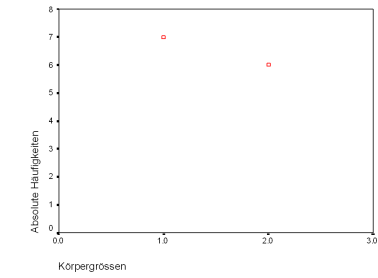


Abbildung 3.1.6: Graphische Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.2, Spalte 2, im Koordinatennetz.

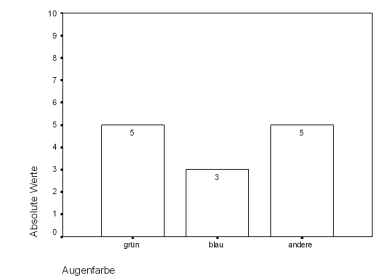


Abbildung 3.1.7: Graphische Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.2, Spalte 1, als Balkendiagramm.

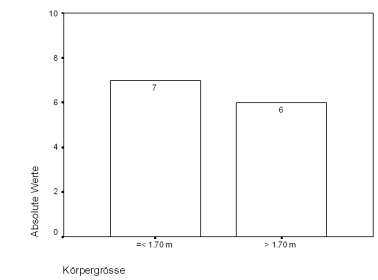


Abbildung 3.1.8: Graphische Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.2, Spalte 2, als Balkendiagramm.

solchen Fällen können wir zu sinnvollen Häufigkeitsverteilungen gelangen, indem wir die Daten zu Klassen zusammenfügen und dann bezüglich dieser Klassen Häufigkeitsauszählungen vornehmen. Bei metrisch skalierten Daten ist darauf zu achten, jeweils gleiche Klassenbreiten zu wählen. Die gewählte Klassenbreite ist an sich willkürlich, wobei das Ziel der Klassenbildung in interessanten und informativen Resultaten besteht. An den Rändern kann von der Regel der gleichen Klassenbreite abgewichen werden. Oft fasst man alle Restdaten an den Rändern zu einer einzigen Klasse zusammen. Bei ordinalskalierten Daten ist „Breite“ nicht definiert. Entsprechend ist die Wahl der „Klassenbreite“ willkürlich. Die Wahl der „Klassenbreite“ ist hier nur pragmatisch zu rechtfertigen: eine gute Klassifizierung führt zu interessanten Resultaten.

(2) Bei ordinalskalierten oder diskret metrisch skalierten Variablen ergibt sich für die x-Achse eine „natürliche“ Ordnung der Ausprägungen, während bei nominalskalierten Variablen die Anordnung der Balken ebenso willkürlich ist wie die Zuordnung von Skalenwerten zu den Ausprägungen.

Beispiel 3.1.4. Betriebe in einer Kleinstadt werden auf die Anzahl der Büroräume untersucht. Dabei ergibt sich der folgende Datensatz:

Es liegt eine diskret metrische Skala vor. Da jedoch nicht viele Ausprägungen vorliegen, kann auf eine vorgängige Klassifizierung verzichtet werden. Die Menge der Skalenwerte ist $\{1, 2, 3, 4, 5, 6\}$.

Wir erhalten für die Häufigkeitsauszählung den folgenden Graphen:

$g_k = \{(1, 3), (2, 3), (3, 4), (4, 5), (5, 3), (6, 1)\}$ Graphische Darstellungen (s. Abbildung 3.1.9 und 3.1.10):

Firma	Anzahl Büroräume
1	2
2	5
3	4
4	1
5	4
6	3
7	2
8	4
9	3
10	2
11	1
12	1
13	3
14	3
15	4
16	5
17	5
18	4
19	6

Tabelle 3.1.3: Urliste von Daten

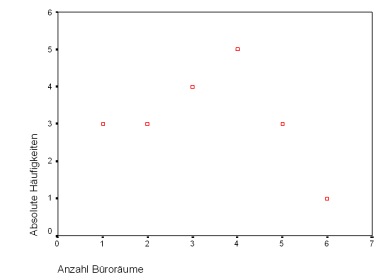


Abbildung 3.1.9: Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.3 im Koordinatennetz

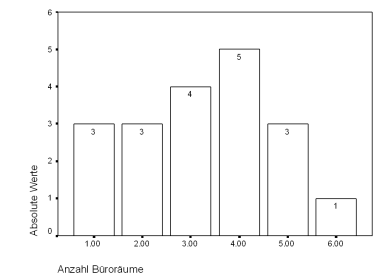


Abbildung 3.1.10: Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.3 mit Balkendiagramm

◇

Beispiel 3.1.5. In einer Klasse würden sich die folgenden Noten ergeben (Ordinalskala mit - im Verhältnis zur Zahl der Ausprägungen - wenigen Daten; mögliche Skalenwerte sind die Werte zwischen 1 und 6 auf eine Stelle nach dem Komma):

Erstellen wir ohne vorgängige Klassifizierung eine Häufigkeitsverteilung, so erhalten wir den Graphen:

$$g_k = \{(3.1, 1), (3.5, 1), (3.7, 1), (3.8, 1), (4.1, 1), (4.2, 1), (4.3, 1), (4.5, 1), (4.6, 2), (4.8, 1), (5.2, 1), (5.3, 1), (5.6, 1)\}$$

Dies ergibt folgendes, nicht sehr informatives Balkendiagramm (3.1.11):

Person	Noten
1	5.2
2	4.1
3	3.8
4	3.1
5	4.8
6	5.6
7	5.3
8	3.5
9	4.3
10	4.6
11	4.2
12	3.7
13	4.6
14	4.5

Tabelle 3.1.4: Beispiel für Daten, die ohne Klassifikation keine aussagekräftige Häufigkeitsverteilung ergeben

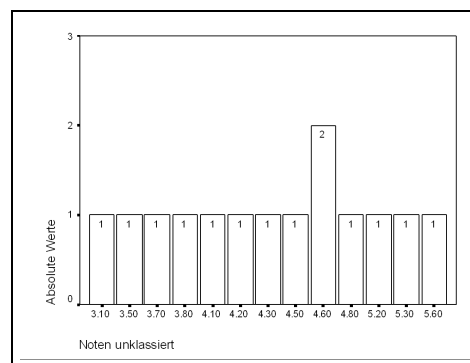


Abbildung 3.1.11: Daten der Tabelle 3.1.4 mit Hilfe eines Balkendiagramms

Entsprechend nehmen wir eine Klassifizierung vor. z.B.

$[3, 3.5]$, $]3.5, 4]$, $]4, 4.5]$, $]4.5, 5]$, $]5, 5.5]$, $]5.5, 6]$ und definieren die folgende Skala:

$f(x) = 3.25$ für $x \in [3, 3.5]$,

$f(x) = 3.75$ für $x \in]3.5, 4]$,

$f(x) = 4.25$ für $x \in]4, 4.5]$,

$f(x) = 4.75$ für $x \in]4.5, 5]$,

$f(x) = 5.25$ für $x \in]5, 5.5]$,

$f(x) = 5.75$ für $x \in]5.5, 6]$.

Wir erhalten nun den folgenden Graphen:

$\{(3.25, 1), (3.75, 3), (4.25, 3), (4.75, 4), (5.25, 2), (5.75, 1)\}$.

In Form eines Balkendiagrammes (s. Abbildung 3.1.12):

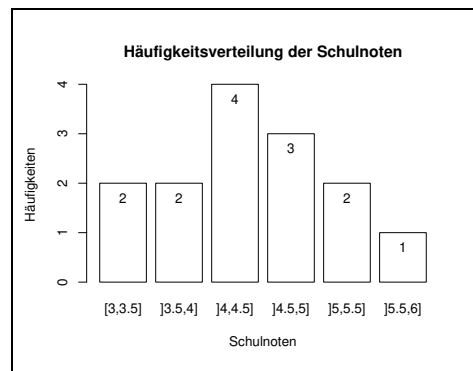


Abbildung 3.1.12: Graphische Darstellung der klassierten Daten der Tabelle 3.1.4 mit Hilfe eines Balkendiagramms

Wir haben nun ein informatives und anschauliches Bild über die Verteilung der Daten. ◇

Beispiel 3.1.6. Ordinalskalierte Variable mit - im Verhältnis zur Zahl der Daten - wenig Ausprägungen (s. Tabelle 3.1.5):

Person	Zufriedenheitsgrad
1	zufrieden
2	zufrieden
3	unzufrieden
4	unzufrieden
5	sehr zufrieden
6	sehr zufrieden
7	sehr zufrieden
8	sehr zufrieden
9	zufrieden
10	zufrieden

Tabelle 3.1.5: Urliste der Daten einer ordinalskalierten Variable

Da wenige Ausprägungen vorliegen, ist für die Erstellung einer informativen Häufigkeitsverteilung keine vorgängige Klassifizierung nötig. Eine solche würde vielmehr sogar jeglichen Informationsgehalt zerstören. Wir erhalten (mit 1 für „unzufrieden“, 2 für „zufrieden“ und 3 für „sehr zufrieden“):

Graph von $k = \{(1, 2), (2, 4), (3, 4)\}$

Dies gibt die graphische Darstellung mittels eines Balkendiagrammes (s. Abbildung 3.1.13):

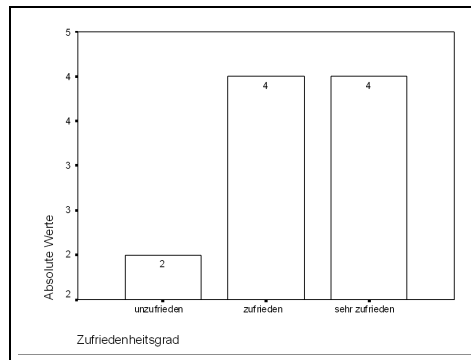


Abbildung 3.1.13: Graphische Darstellung der Häufigkeitsverteilung der Daten von Tabelle 3.1.5.

◇

Beispiel 3.1.7. für Skalenwerte, welche Körpergrößen ausdrücken:

Person	Grösse
1	176.25
2	179.59
3	163.31
4	189.95
5	161.55
6	169.23
7	188.37
8	183.16
9	178.41
10	190.11
11	183.14
12	171.65
13	179.31
14	175.57
15	180.03

Tabelle 3.1.6: Urliste von Daten einer metrisch skalierten Variable

Da bei solchen Variablen gewöhnlich pro Ausprägung höchstens ein Objekt auftaucht, wird eine Häufigkeitsverteilung gleichviele Elemente enthalten wie der Datensatz. Damit ergibt sich keine informative Häufigkeitsverteilung. Deshalb klassifizieren wir die Daten: Wir wählen z.B. folgende Intervalle für folgendes f :

$$f(x) = 162.5 \text{ für } x \in [160, 165],$$

$$f(x) = 167.5 \text{ für } x \in]165, 170],$$

$$f(x) = 172.5 \text{ für } x \in]170, 175],$$

$$f(x) = 177.5 \text{ für } x \in]175, 180],$$

$$f(x) = 182.5 \text{ für } x \in]180, 185],$$

$$f(x) = 187.5 \text{ für } x \in]185, 190]$$

$$f(x) = 192.5 \text{ für } x \in]190, 195]$$

In Grunde genommen erstellen wir eine neue Skala mit weniger Ausprägungen (Informationsreduktion). Wir sprechen in solche Fällen auch von „Umkodierung“.

Wir erhalten den folgenden Graphen von k :

$$\{(162.5, 2), (167.5, 1), (172.5, 1), (177.5, 5), (182.5, 3), (187.5, 2), (192.5, 1)\}$$

Dies ergibt dann das informative Balkendiagramm (s. Abbildung 3.1.14):

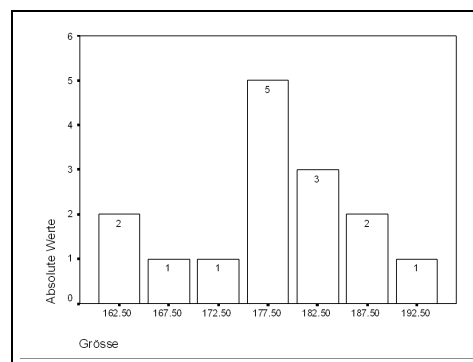


Abbildung 3.1.14: Graphische Darstellung der Häufigkeitsverteilung der klassifizierten Daten der Tabelle 3.1.6 mit Hilfe eines Balkendiagramms

Hier könnte auch ein Histogramm verwendet werden. Bei Histogrammen werden die Balken aneinander gefügt. Die Fläche der Balken ist proportional zu den (relativen) Häufigkeiten gewählt. Histogramme werden i.A. für metrisch skalierte Daten verwendet. Im folgenden Histogramm wurden andere Klassengrenzen gewählt als im Balkendiagramm (s. Abbildung 3.1.15):

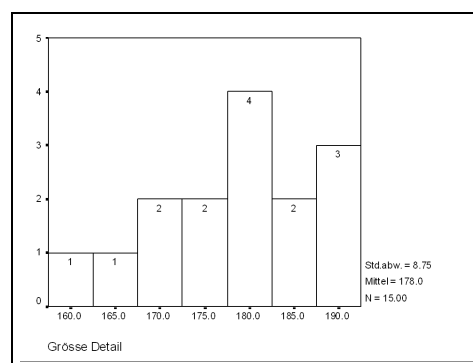


Abbildung 3.1.15: Graphische Darstellung der Häufigkeitsverteilung der klassifizierten Daten der Tabelle 3.1.6 mit Hilfe eines Histogramms.

Die beiden graphischen Darstellungen zeigen, wie unterschiedlich das graphische Ergebnis ausfallen kann, wenn die Intervallgrenzen verschieden gewählt werden. Solche Unterschiede werden bei grossen Datenmengen jedoch i.A. geringer. ◇

Bei sozialwissenschaftlichen (ökonomischen) Statistiken ergibt sich oft das Problem von nach unten oder oben offenen Klassengrenzen. Wollen wir z.B. die Einkommensverteilung der Steuerpflichtigen des Kantons Wallis graphisch darstellen, würden wir bei immer gleichbleibenden Intervallen am Schluss unzählige Intervalle erhalten, die oft leer sind. Um zu informativeren Darstellungen zu gelangen, würden wir zum Beispiel die Einkommen über 2 Millionen zusammenfassen.

Beispiel 3.1.8. In einer Gemeinde würden sich folgende steuerbare Einkommen ergeben:

Steuerbares Einkommen				
15414.2	44958.3	20504.3	9818.9	53899.3
33574.7	33479.3	24933.8	10491.3	1059815.1
41923.3	45956.8	18271.2	17927.5	523444.5
37068.6	38023.1	35104.2	35007.3	213860.7
50651.5	54790.3	45663.2	35134	142576.8
49524.6	35444	30391.8	25165.6	800433
28374.4	8155.2	35791	41378.4	
28259.1	64980.4	19116.2	41639.5	
44239.4	38558.9	50165.9	32043	
32790.8	24256.4	25497	34620.8	
38009.3	-5100.1	22652.7	29005.2	

Tabelle 3.1.7: Tabelle mit Daten einer metrisch skalierten Variable (Um Platz zu sparen nicht als Urliste von Daten mit einem Datum pro Objekt und Zeile geliefert). Steuerbares Jahreseinkommen in Franken.

Wir stellen fest, dass es über 70'000 nur mehr 5 Einkommen gibt, die sogar über eine Million streuen. Entsprechend fassen wir die steuerbaren Einkommen über 70'000 in einer Klasse zusammen und erhalten das folgende Balkendiagramm (s. Abbildung 3.1.16):

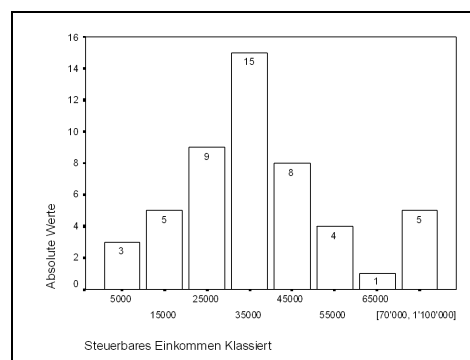


Abbildung 3.1.16: Graphische Darstellung der Häufigkeitsverteilung der Daten von Abbildung 3.1.7 mit Hilfe eines Balkendiagramms

◇

Eine anschauliche Variante von Häufigkeitsdiagrammen besteht in der „Stamm- und Blatt-Darstellung“ (Stem and Leaf): Wir nehmen an, wir hätten folgende Datensatz:

(1.3, 2.4, 1.5, 1.5, 0.1, 1.1, 1.8, 0.9, 4.7, 1.5). Wir schreiben die Ziffern 0 bis 9 untereinander hin und zu den passenden Stellen vor dem Komma die jeweils erste Stelle hinter dem Komma. Wir sehen unmittelbar, wie die Daten verteilt sind (s. Abbildung 3.1.17).

S t a m m	B l ä t t e r
0	1 9
1	3 5 5 1 8 5
2	4
3	
4	7
5	
6	
7	
8	
9	

Abbildung 3.1.17: Stamm- und Blattdarstellung von Daten

Liegen Daten vor, die nicht zwischen 0 und 10 liegen, kann man Zahlen jeweils in eine entsprechende Form bringen: $n.m \cdot 10^k$ oder $n.m \cdot 10^{-k}$ mit $n, m \in \mathbb{N}_9$ (= die ersten 9 natürlichen Zahlen mit der 0).

3.2 Relative Häufigkeitsverteilungen

Bei einer relativen Häufigkeitsverteilung ordnen wir den Skalenwerten der Ausprägungen einer Variablen die relative Häufigkeit des Auftretens des entsprechenden Skalenwertes zu (= Anteil des Auftretens des Merkmals an der Gesamtmenge). Etwas formaler können wir definieren: Sei k eine absolute Häufigkeitsverteilung mit $k(x) = |\{z \mid f(z) = x \text{ und } z \in G\}|$ und $n = |G|$ die Anzahl der Objekte der Grundmenge G und $\text{Bild}(f)$ das Bild der Skala f . Dann legen wir fest:

Definition 3.2.1. h ist eine relative Häufigkeitsverteilung der f -skalierten Variable V genau dann, wenn

$$h : \text{Bild}(f) \rightarrow [0, 1]$$

$$h(z) := \frac{k(z)}{n}$$

Oft werden relative Häufigkeitsverteilungen in Prozenten angegeben.

Beispiel 3.2.2. Für die Daten der Tabelle 3.1.2 erhalten wir (mit $n = 5 + 3 + 5 = 7 + 6 = 13$) :

Variable Augenfarbe: Graph von $h_1 = \{(1, \frac{5}{13}), (2, \frac{3}{13}), (3, \frac{5}{13})\}$

Variable „Körpergröße“: Graph von $h_2 = \{(1, \frac{7}{13}), (2, \frac{6}{13})\}$

Die graphischen Darstellungen unterscheiden sich von den absoluten Häufigkeitsverteilungen nur in den Beschriftungen der y-Achse (s. Abbildungen 3.2.18, 3.2.19, 3.2.20 und 3.2.21 für die Variable „Augenfarbe“).

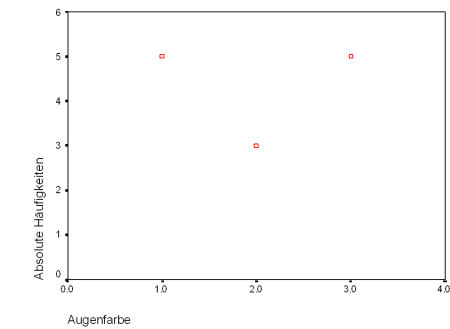


Abbildung 3.2.18: Absolute Häufigkeitsverteilung der Variable Augenfarbe der Tabelle 3.1.2 (Punkte im Koordinatennetz)

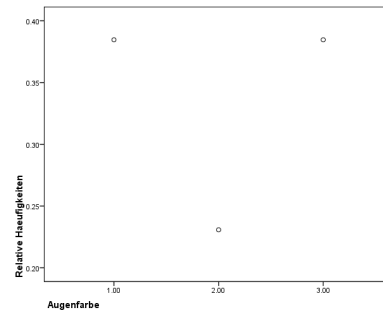


Abbildung 3.2.19: Relative Häufigkeitsverteilung der Variable Augenfarbe der Tabelle 3.1.2 (Punkte im Koordinatennetz)

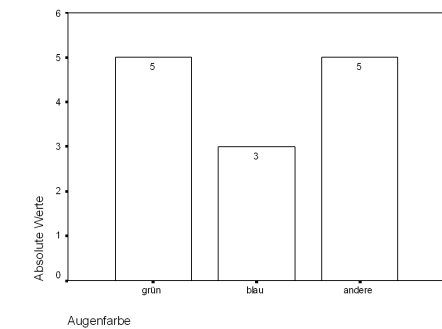


Abbildung 3.2.20: Absolute Häufigkeitsverteilung Variable Augenfarbe der Tabelle 3.1.2 (Balkendiagramm)

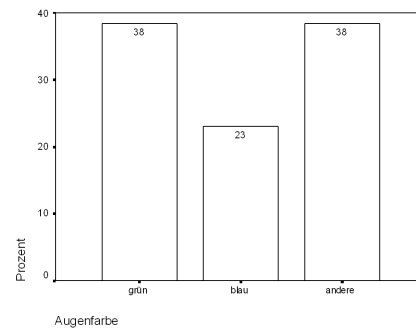


Abbildung 3.2.21: Relative Häufigkeitsverteilung der Variable Augenfarbe der Tabelle 3.1.2 (Balkendiagramm)

◇

Ebenso lassen sich Kreisdiagramme oder Histogramme erstellen, die sich höchstens in der Beschriftung von den graphischen Darstellungen der absoluten Häufigkeiten unterscheiden.

3.3 Absolute kumulative Häufigkeitsverteilung

Eine kumulative Häufigkeitsverteilung besteht darin, dass zu jeder Ausprägung einer Variable die Summe der vorangehenden Anzahlen (inklusive der eigenen Anzahl) von Skalenwerten zugeordnet werden. Entsprechend macht eine kumulative Häufigkeitsverteilung eigentlich nur dann Sinn, wenn die Skalenwerte geordnet sind (ordinale oder metrische Skalen).

Definition 3.3.1. k_k ist eine absolute, kumulative Häufigkeitsverteilung einer Grundmenge G bezüglich der f -skalierten Variable V (ordinal oder metrisch skaliert) genau dann, wenn

$$k_k : \text{Bild}(f) \rightarrow \mathbb{N}$$

$$k_k(x_i) := \sum_{x_j \leq x_i} k(x_j)$$

Beispiel 3.3.2. Bei einer Marktforschungstudie ergibt sich die folgende absolute Häufigkeitsverteilung (Zufriedenheit mit einem Produkt, 4 für „sehr zufrieden“, 3 für „zufrieden“, 2 für „als knapp genügend eingestuft“ und 1 für „als ungenügend eingestuft“.

Graph von $k = \{(1, 5), (2, 8), (3, 4), (4, 2)\}$. Dies ergibt den folgenden Graphen der kumulativen Häufigkeitsverteilung

$k_k : \{(1, 5), (2, 13), (3, 17), (4, 19)\}$

In einer Tabelle sieht das wie folgt aus (s. Tabelle 3.3.8):

Skalenwert	absolute Häufigkeiten	absolute kumulierte Häufigkeiten
1	5	5
2	8	13
3	4	17
4	2	19

Tabelle 3.3.8: Beispiel für die Berechnung der Häufigkeitsverteilung durch das Zählen von Skalenwerten

Eine graphische Darstellung kann in Koordinatennetzen oder mit Balkendiagrammen erfolgen (s. Abbildungen 3.3.22, 3.3.23, 3.3.24 und 3.3.25):

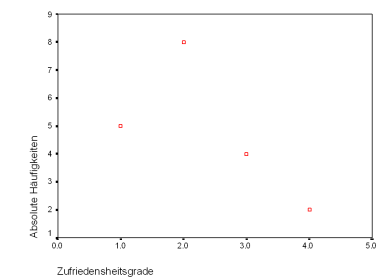


Abbildung 3.3.22: Absolute Häufigkeitsverteilung der Daten der Tabelle 3.3.8 (im Koordinatennetz)

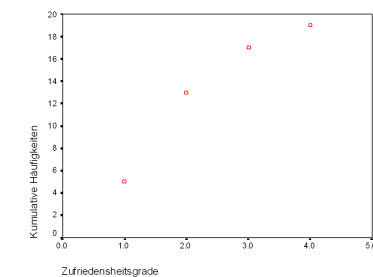


Abbildung 3.3.23: absolute kumulative Häufigkeitsverteilung der Daten der Tabelle 3.3.8 (im Koordinatennetz)

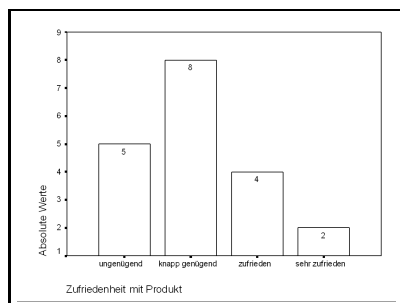


Abbildung 3.3.24: absolute Häufigkeitsverteilung der Daten der Tabelle 3.3.8 (Balkendiagramm)

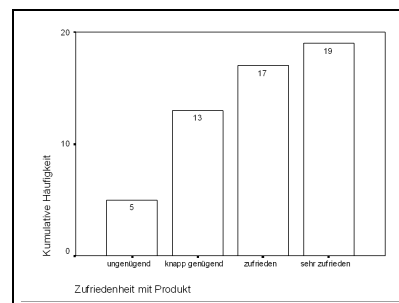


Abbildung 3.3.25: absolute kumulative Häufigkeitsverteilung der Daten der Tabelle 3.3.8 (Balkendiagramm)



3.4 Relative kumulative Häufigkeitsverteilung

Wir erhalten die kumulative Häufigkeitsverteilung aus der absoluten kumulativen Häufigkeitsverteilung, indem wir jeweils die zweite Komponente der Elemente der absoluten kumulativen Häufigkeitsverteilung durch n (= Gesamtanzahl von Daten) dividieren.

Definition 3.4.1. h_k ist eine relative, kumulative Häufigkeitsverteilung einer Grundmenge G bezüglich der f -skalierten Variable V genau dann, wenn

$$h_k : \text{Bild}(f) \rightarrow \mathbb{N} \quad (n = |G|)$$

$$h_k(x_i) := \frac{\sum_{x_j \leq x_i} k(x_j)}{n}$$

Mit den Daten von *Beispiel 3.3.2* wird mit $n = 19$:

Graph von $h_k = \{(1, \frac{5}{19}), (2, \frac{13}{19}), (3, \frac{17}{19}), (4, \frac{19}{19})\}$

Oft werden relative kumulative Häufigkeitsverteilungen auch in Prozenten angegeben.

Beispiel 3.4.2. Für *Beispiel 3.3.2* werden die relativen kumulativen und relativen nicht-kumulativen Häufigkeitsverteilungen im Koordinatennetz und als Balkendiagramm dargestellt (s. Abbildungen 3.4.26, 3.4.27, 3.4.28 und 3.4.29). 

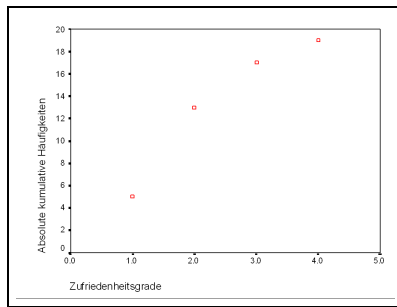


Abbildung 3.4.26: absolute kumulative Häufigkeitsverteilung der Daten des Beispiels 3.3.2 (im Koordinatennetz)

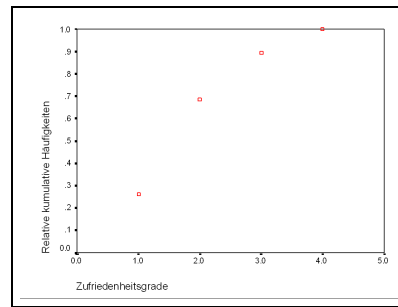


Abbildung 3.4.27: relative, kumulative Häufigkeitsverteilung der Daten des Beispiels 3.3.2 (im Koordinatennetz)

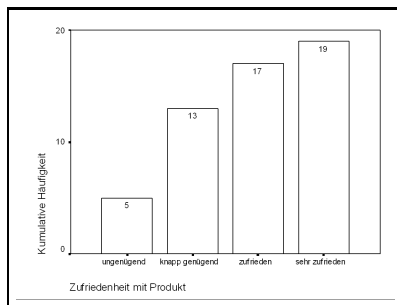


Abbildung 3.4.28: absolute kumulative Häufigkeitsverteilung der Daten des Beispiels 3.3.2 (Balkendiagramm)

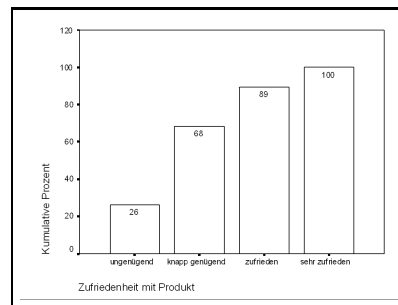


Abbildung 3.4.29: relative, kumulative Häufigkeitsverteilung der Daten des Beispiels 3.3.2 (Balkendiagramm)

Wir sehen am Beispiel, dass die kumulative Kurve links stärker ansteigt als rechts. Die Häufigkeitsunterschiede bei den tiefen Zufriedenheitsgraden sind damit grösser als bei den hohen Zufriedenheitsgraden.

3.5 Eindimensionales Streudiagramm und empirische Verteilungsfunktion

Für stetig, metrisch skalierte Daten ergibt sich ohne vorgängige Klassifizierung die folgende, recht informative Darstellung. Wir nennen sie künftig „eindimensionales Streudiagramm“. Dazu tragen wir die Skalenwerte auf einer zur x-Achse parallelen Geraden ab (s. Abbildung 3.5.30):

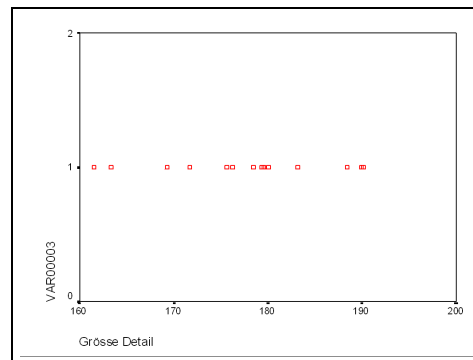


Abbildung 3.5.30: Eindimensionales Streudiagramm der Daten von Tabelle 3.1.6

Wir sehen, wie die Daten in etwa verteilt sind: wo sie sich häufen und wo sie weiter auseinander stehen (Die Beschriftung der y-Achse ist hier bedeutungslos).

Noch instruktiver ist für nicht klassifizierte metrisch skalierte Daten die folgende Methode, die auf der eben beschriebenen beruht (wir werden sie „(empirische) Verteilungsfunktion (der Daten)“ nennen): Die Daten werden geordnet. Wir können folgende Punkte in ein Koordinatennetz einzeichnen: $(x_i, \frac{i}{n})$. Dabei ist x_i das i -te Datum eines geordneten Datensatzes. $\frac{i}{n}$ ist die relative Häufigkeit von Daten x_j , die kleiner gleich x_i sind. Für die Körpergrössendaten erhalten wir (s. Abbildung 3.5.31):

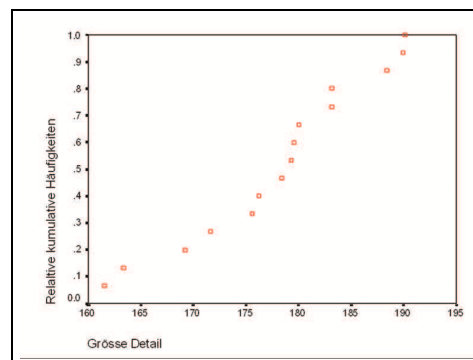


Abbildung 3.5.31: Graphische Darstellung der empirischen Verteilungsfunktion (Punkteform) der Daten von Tabelle 3.1.6

Formaler können wir die Verteilungsfunktion F wie folgt definieren (die Variable V sei f -skaliert; $Bild(f)$ sei das Bild der Skala f) :

Definition 3.5.1. F ist eine (empirische) Verteilungsfunktion genau dann, wenn

$$F : Bild(f) \rightarrow [0, 1]$$

$$F(x_i) : = \frac{|\{x_j : x_j \leq x_i \text{ und } x_j \in Bild(f)\}|}{|G|}$$

Bemerkung 3.5.2. Die empirische Verteilungsfunktion ist mit der relativen, kumulativen Häufigkeitsverteilung identisch. Entsprechend gilt:

$$F(x_i) = \frac{\sum_{x_j \leq x_i} k(x_j)}{|G|}.$$

◇

Empirische Verteilungsfunktionen werden oft und alternativ als Abbildungen

$$F : \mathbb{R} \rightarrow [0, 1]$$

$$F(x) : = \begin{cases} \frac{|\{x_j : x_j \leq x_i \text{ und } x_j \in \text{Bild}(f)\}|}{|G|} & \text{für } x \in]x_i, x_{i+1}] \text{ und } 1 \leq i \leq n-1; n = |G| \\ 1 & \text{für } x \geq x_n \\ 0 & \text{für } x < x_1 \end{cases}$$

definiert. Wir erhalten dann nicht eine Punktfunktion, sondern eine Treppenfunktion (s. Abbildung 3.5.32:

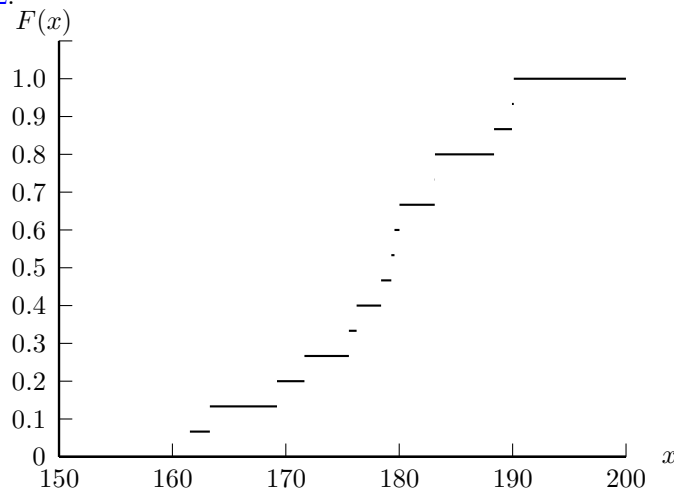


Abbildung 3.5.32: Empirische Verteilungsfunktion (Treppenform)

Beispiel 3.5.3. Die Lebenszeit von 10 Maschinen desselben Typs wird erhoben (s. Tabelle 3.5.9), was die folgenden Graphiken der empirischen Verteilungsfunktionen gibt (s. Abbildungen 3.5.33 und 3.5.34):

Maschine	Lebenszeit von Maschinen (in Monaten)
1	3
2	5
3	5.5
4	6
5	8
6	11
7	15
8	16
9	19
10	21

Tabelle 3.5.9: Beispieldaten für die Berechnung der empirischen Verteilungsfunktion

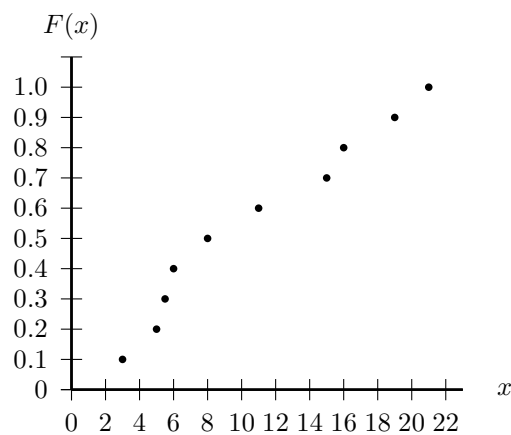


Abbildung 3.5.33: Empirische Verteilungsfunktion (Punkteform) der Daten von Tabelle 3.5.9

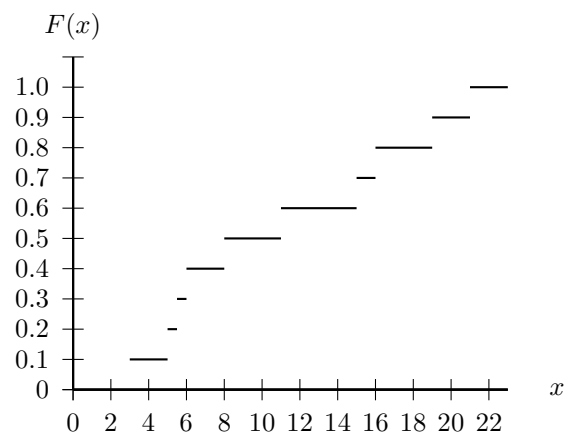


Abbildung 3.5.34: Empirische Verteilungsfunktion (Treppenform) der Daten von Tabelle 3.5.9

◇

Die Funktion ist jeweils fürs Intervall $]x_i, x_{i+1}]$ konstant. Wir werden i.A. die Punkteversion verwenden. Verteilungsfunktionen werden auch „kumulative Verteilungsfunktionen“ genannt (englisch: cumulative distribution function = cdf).

3.5.1 Übungen

Bemerkungen: die meisten Daten finden Sie unter <http://math.logik.ch> Statistik

1. Rechnen und zeichnen Sie die obigen Beispiele und Aufgaben selbständig durch (von Hand, mit Excel).
2. Zeichnen Sie eine Verteilungsfunktion (Punkteform und Treppenfunktion) der folgenden Daten von Hand (s. Tabelle 3.5.10):

Datennummern	Geordnete Daten	
1	0.06	
2	0.12	
3	0.24	
4	0.29	
5	0.3	
6	0.33	
7	0.36	

Tabelle 3.5.10: Daten in Tabellenform

- Erstellen Sie mit Excel eine Verteilungsfunktion (Punkteform) der Daten des Beispiels 3.1.8 (steuerbares Einkommen, Exceldatei „Häufigkeitsverteilungen“).

3.5.2 Lösungen

- siehe Abbildungen im Text
- Wir erhalten (Lösungen Häufigkeitsverteilungen.xls, s. Abbildungen 3.5.35 und 3.5.36):

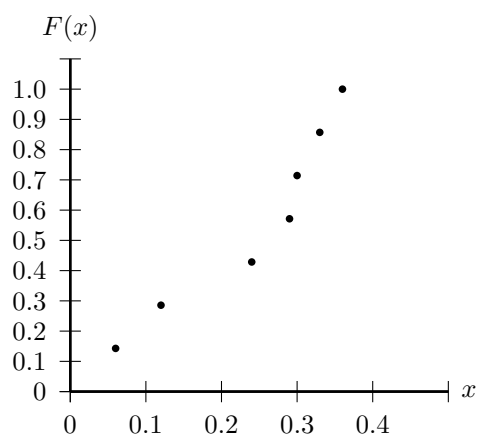


Abbildung 3.5.35: Graphische Darstellung der empirischen Verteilungsfunktion (Punkteform) der Daten der Tabelle 3.5.10

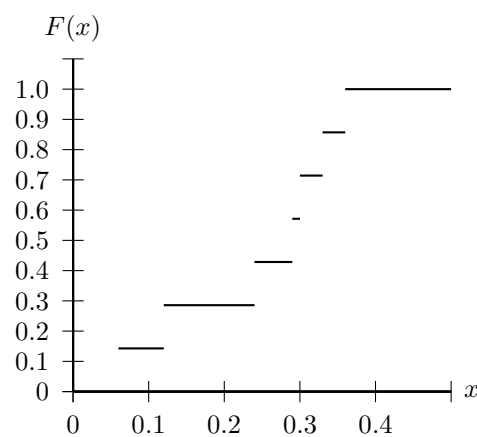


Abbildung 3.5.36: Graphische Darstellung der empirischen Verteilungsfunktion (Treppenform) der Daten der Tabelle 3.5.10

- Wir erhalten (s. Abbildungen 3.5.37):

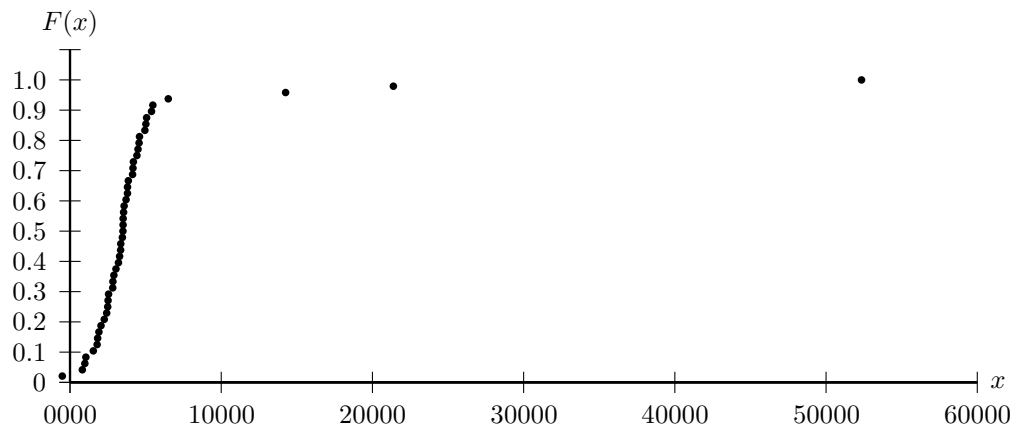


Abbildung 3.5.37: Graphische Darstellung der empirischen Verteilungsfunktion der Daten des Beispiels (Punkteform)

3.6 Lernziele

- Für Datensätze die zwei Arten von Häufigkeitsverteilungen und die zwei Arten von kumulativen Häufigkeitsverteilungen als mengentheoretischen Graphen der entsprechenden Häufigkeitsverteilungen hinschreiben können.
- Für Datensätze die zwei Arten von Häufigkeitsverteilungen und die zwei Arten von kumulativen Häufigkeitsverteilungen in einem Koordinatennetz und mit Balkendiagrammen zeichnen können (von Hand, mit Excel, mit einer Statistiksoftware).
- Für metrisch skalierte Daten Balkendiagramme auf Grund vorgängiger Klassifikation zeichnen können. (von Hand, mit Excel, mit einer Statistiksoftware). Klassifikationen von Daten mit Excel durchführen können.
- Für nicht-klassifizierte, metrisch skalierte Daten eindimensionale Streudiagramme zeichnen können (von Hand, mit Excel, mit einer Statistiksoftware)
- Für nicht klassifizierte, metrisch skalierte Daten eine Verteilungsfunktion zeichnen können (von Hand, mit Excel, mit einer Statistiksoftware).
- In einer Statistiksoftware Variablen bearbeiten können (Skalenniveau, Beschriftung der Variablen, Umkodieren von Variablen, Beschriftung der Ausprägungen).

Kapitel 4

Univariate statistische Kennwerte

Eine weitere Möglichkeit, unübersichtliche Datensätze so zu bearbeiten, dass informative Daten entstehen, besteht in der Berechnung von sogenannten statistischen Kennwerten (auch „statistische Masszahlen“ oder „statistische Kennzahlen“ genannt). Wir behandeln zuerst Kennwerte, die jeweils nur eine Variable betreffen (univariat).

Die Kennwerte sollen primär über folgende Sachverhalte Auskunft geben:

1. In welchem Bereich liegen die Daten.
2. Wo liegt der „Schwerpunkt“, die Mitte oder die Mehrzahl der beobachteten Werte?
3. Wie streuen die Daten?

Auf die erste Frage liefern das Maximum und das Minimum eines Datenvektors x eine Antwort. Bei metrisch skalierten Daten ist in diesem Zusammenhang die Angabe des Intervalls

$$[\min(x), \max(x)]$$

sinnvoll, wobei wir $\min(x)$ bezüglich Datenvektoren als das kleinste Element der Menge der Komponenten des Datenvektors definieren - analog für $\max(x)$, das Maximum eines Datenvektors, das wir als größtes Element der Menge der Komponenten eines Datenvektors definieren.

Auf die zweite Frage liefern die „Kennwerte der mittleren Lage“ die Antwort. Sie werden deshalb auch „Lagekennwerte“, „Lagekennzahlen“ oder „Masse der mittleren Lage“ genannt. Wir werden Definition und Bedeutung folgender Mittelwerte betrachten: Arithmetisches Mittel, Median, Modalwert.

Zur Beantwortung der dritten Frage dienen die Streuungskennwerte: Varianz, Standardabweichung, Spannweite, Quantile und Schiefe.

4.1 Kennwerte der mittleren Lage

Kennwerte der mittleren Lage dienen dazu, zwei Untersuchungseinheiten bezüglich der „durchschnittlichen“ Lage der Daten zu vergleichen. Wo ist die Butter im Durchschnitt teurer? Wo kosten die Arbeitskräfte im Durchschnitt weniger? Wo ist die Einstellung zu einem Produkt im Mittel positiver.

4.1.1 Das arithmetische Mittel (Durchschnitt)

Das arithmetische Mittel ist für metrisch skalierte Variablen geeignet. Das arithmetische Mittel ist nicht ein robuster Kennwert (Erläuterung folgt in Kürze).

Das arithmetische Mittel ist die Summe der Messwerte (= Skalenwerte) einer metrisch skalierten Variable dividiert durch die Anzahl der Messwerte. Für das arithmetische Mittel verwenden wir

die Abkürzung: $\bar{x}$. Statt „arithmetisches Mittel“ verwenden wir auch „Durchschnitt“ oder kurz „Mittelwert“.

Definition 4.1.1. $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ist das arithmetische Mittel des Datenvektors $(x_1, \dots, x_n)$

Beispiel 4.1.2. In den fünf Lebensmittelgeschäften einer Kleinstadt wird ein Pfund Butter für folgende Preise verkauft: Fr.7.30; 7.50; 7.60; 7.20; 7.80. Das arithmetische Mittel der Preise (= Durchschnittspreis) beläuft sich auf: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{5} (7.30 + 7.50 + 7.60 + 7.20 + 7.80) = 7.48$ $\diamond$

Die Einheiten des arithmetischen Mittels sind mit den Einheiten des Datensatzes identisch. Der Durchschnitt im Beispiel 4.1.2 beträgt somit Fr. 7.48.

Im Falle von m Klassen von Objekten mit den Häufigkeiten $k(x)$ pro Klasse mit dem Skalenwert x kann der Durchschnitt abkürzend wie folgt gebildet werden (gewichtetes arithmetisches Mittel, dabei wählen wir pro Klasse i einen Vertreter x_i):

$$\bar{x} = \frac{1}{\sum_{i=1}^m k(x_i)} \left(\sum_{i=1}^m k(x_i) \cdot x_i \right) = \frac{1}{n} \left(\sum_{i=1}^m k(x_i) \cdot x_i \right)$$

Es handelt sich nur um eine abkürzende Schreibweise. Dies beleuchten wir am besten mit einem Beispiel. Es gilt z.B. für $x = 5 + 5 + 5 + 7 + 7 + 8 + 8 + 8 + 8 + 8 + 9 + 9 + 9 + 9 + 9$:

$$\begin{aligned} \frac{1}{14} \sum_{i=1}^{14} x_i &= \frac{1}{14} (5 + 5 + 5 + 7 + 7 + 8 + 8 + 8 + 8 + 8 + 9 + 9 + 9 + 9 + 9) \\ &= \frac{1}{3 + 2 + 4 + 5} ((3 \cdot 5) + (2 \cdot 7) + (4 \cdot 8) + (5 \cdot 9)) \\ &= \frac{1}{\sum_{i=1}^4 k(x_i)} \left(\sum_{i=1}^4 k(x_i) \cdot x_i \right) \end{aligned}$$

Beispiel 4.1.3. In den fünf Lebensmittelgeschäften einer Kleinstadt wird ein Pfund Butter für folgende Preise verkauft: Fr.7.30; 7.50; 7.60; 7.20; 7.80. Folgende Mengen wurden in fünf Geschäften verkauft: 120; 80; 30; 500; 20 Pfund. Wieviel beträgt der durchschnittlich pro Pfund bezahlte Preis: Wir berechnen:

$$\bar{x} = \frac{1}{120+80+30+500+20} ((7.3 \cdot 120) + (7.5 \cdot 80) + (7.6 \cdot 30) + (7.2 \cdot 500) + (7.8 \cdot 20)) = 7.28$$

Der durchschnittliche bezahlte Preis beträgt somit: 7.28 Franken.

Wir betrachten das Beispiel 4.1.3 noch in Tabellenform (s. Tabelle 4.1.1):

Preis	Abgesetzte Menge	Produkt
7.3	120	876
7.5	80	600
7.6	30	228
7.2	500	3600
7.8	20	156
Summen	750	5460
Durchschnitt	5460:750=	7.28

Tabelle 4.1.1: Beispiel für die Berechnung des arithmetischen Mittels bei klassierten Daten

$\diamond$

Ein nur mehr näherungsweise arithmetisches Mittel erhalten wir bei klassierten Daten, wenn wird die Daten vorgängig umkodieren (s. Beispiel 4.1.4).

Beispiel 4.1.4. In einer Schulkasse wird die Körpergrösse der Schüler gemessen: 164, 165, 182, 164, 160, 175, 174, 169, 168, 163, 156, 161, 168, 175, 170, 165 cm. Wieviel misst durchschnittlich ein Schüler? Wir erhalten: 167.4375 (Rechnen Sie bitte nach!). Wir fassen die Einzelwerte zu Klassen zusammen und berechnen das arithmetische Mittel der entstandenen Häufigkeitsverteilung. Wir benutzen dabei die Tabelle 4.1.2.

Intervalle	Klassenmitte	H(xi)
[155, 160]	157.5	2
]160, 165]	162.5	6
]165, 170]	167.5	4
]170, 175]	172.5	3
]175, 180]	177.5	0
]180, 185]	182.5	1

Tabelle 4.1.2: Mittelwertwerte für klassierte Daten (Beispiel 4.1.4)

Unter der Annahme der gleichmässigen Verteilung der Messwerte innerhalb der Klassen ist die Verwendung der Klassenmitte als Merkmal x_i gerechtfertigt. Berechnen wir die durchschnittliche Grösse in dieser Häufigkeitsverteilung erhalten wir: 2; 6; 4; 3; 0; 1

$$\begin{aligned}\bar{x} &= \frac{1}{\sum_{i=1}^m k(x_i)} \left(\sum_{i=1}^m k(x_i) \cdot x_i \right) \\ &= \frac{(2 \cdot 157.5) + (6 \cdot 162.5) + (4 \cdot 167.5) + (3 \cdot 172.5) + (0 \cdot 177.5) + (1 \cdot 182.5)}{2 + 6 + 4 + 3 + 0 + 1} \\ &= 166.25\end{aligned}$$

Welchen Fehler macht man bei der Verwendung der Klassenmitten höchstens? Ein maximaler Fehler entsteht, wenn in allen Klassen die Werte extrem ungleich verteilt sind, d.h. wenn alle entweder an den jeweils oberen Intervallgrenzen liegen, oder an den jeweils unteren Intervallgrenzen. Im Beispiel beträgt der maximal mögliche Mittelwert $\bar{x}_o$: $\diamond$

$$\begin{aligned}& \frac{1}{2 + 6 + 4 + 3 + 0 + 1} ((2 \cdot 160) + (6 \cdot 165) + (4 \cdot 170) + (3 \cdot 175) + (0 \cdot 180) + (1 \cdot 185)) \\ &= 168.75\end{aligned}$$

Der minimal mögliche Mittelwert $\bar{x}_u$ ist:

$$\begin{aligned}& \frac{1}{2 + 6 + 4 + 3 + 0 + 1} ((2 \cdot 155) + (6 \cdot 160) + (4 \cdot 165) + (3 \cdot 170) + (0 \cdot 175) + (1 \cdot 180)) \\ &= 163.75\end{aligned}$$

Somit gilt für den tatsächlichen Mittelwert: $163.75 \leq \bar{x} \leq 168.75$

Der maximal mögliche Fehler ist damit: $\max\{|\bar{x}_k - \bar{x}_g|\}$, wobei $\bar{x}_k$ der errechnete Mittelwert bei Klassierung ist und $\bar{x}_g \in \{\bar{x}_o, \bar{x}_u\}$. Im Beispiel ist somit der maximale Fehler

$$\max\{|166.25 - 168.75|, |166.25 - 163.75|\} = 2.5$$

Ob ein Durchschnitt mit klassierten Daten angemessen ist oder nicht, hängt von der gewünschten Genauigkeit ab. Haben wir die Urliste der Daten nicht zur Verfügung, haben wir keine andere Möglichkeit, als die Näherung zu berechnen. Wir können dann wenigstens nach der obigen Methode den maximalen Fehler berechnen.

Das arithmetische Mittel kann auch beim Vorliegen der relativen Häufigkeiten pro Klasse berechnet werden. In diesem Fall erfolgt die Berechnung durch die Multiplikation der Klassenmitte und der relativen Häufigkeiten (m Klassen; wir wählen erneut x_i als Vertreter der Klasse i):

$$\bar{x} = \sum_{i=1}^m (h(x_i) \cdot x_i)$$

Beweis. $\bar{x} = \frac{1}{\sum_{i=1}^m k(x_i)} \sum_{i=1}^m (k(x_i) \cdot x_i) = \frac{1}{n} \sum_{i=1}^m (k(x_i) \cdot x_i) = \sum_{i=1}^m \left(\frac{k(x_i)}{n} \cdot x_i \right) = \sum_{i=1}^m (h(x_i) \cdot x_i)$ $\square$

Beispiel 4.1.5. Gegeben sei der folgende Graph einer relativen Häufigkeitsverteilung: $\{(5, 0.2), (15, 0.7), (25, 0.1)\}$. $\bar{x} = 5 \cdot 0.2 + 15 \cdot 0.7 + 25 \cdot 0.1 = 14$ $\diamond$

Das arithmetische Mittel kann durch einzelne Werte, die stark von den übrigen Daten abweichen, deutlich beeinflusst werden. Kennwerte mit einem solchen Verhalten nennen wir „nicht robust“. Daten, die stark von den übrigen abweichen, nennen wir „Ausreisser“. Kennwerte, die sich durch Ausreisser kaum oder nicht beeinflussen lassen, nennt man „robust“. Wir zeigen den Mangel an Robustheit des arithmetischen Mittels mit einem Beispiel 4.1.6.

Beispiel 4.1.6. Für die Daten 1, 2, 3, 4, 5, ist der Mittelwert: $\frac{1+2+3+4+5}{5} = 3$

Für die Daten 1, 2, 3, 4, 1000 ist der Mittelwert: $\frac{1+2+3+4+1000}{5} = 202$

Damit gibt der Mittelwert im Beispiel ein völlig falsches Bild über die mittlere Lage der Daten. Die meisten Daten liegen ja um 3 herum, nicht jedoch um 202. $\diamond$

Beispiel 4.1.7. Einkommensverteilung in zwei Bauerndörfern A und B. In der Gemeinde A wohnt ein Einkommensmillionär (s. Tabelle 4.1.3)

	Anzahl Steuerpflichtige		Intervallmitten	Produkte	
Intervalle	Dorf A	Dorf B		Dorf A	Dorf B
$[0, 10'000[$	2	0	5000	10000	0
$]10'000, 20'000]$	10	2	15000	150000	30000
$]20'000, 30'000]$	35	12	25000	875000	300000
$]30'000, 40'000]$	30	24	35000	1050000	840000
$]40'000, 50'000]$	14	42	45000	630000	1890000
$]50'000, 60'000]$	8	15	55000	440000	825000
$]60'000, 1'100'000]$	1	0	1100000	1100000	0
Summen	100	95	1280000	4255000	3885000
Mittelwerte				42550	40894.73684

Tabelle 4.1.3: Einkommensverteilung (Häufigkeitsverteilungen) zweier Dörfer: Berechnung des Mittelwertes mit klassierten Daten.

Während im Dorf A die Einkommen im Allgemeinen deutlich unter dem Einkommen des Dorfes B liegen, ergibt sich bei der Verwendung des arithmetischen Mittels ein Bild, dass diese Situation verfälscht. $\diamond$

Beispiel 4.1.8. Aus Nachfragemangel muss ein Treuhandbüro leider einen von zwei Angestellten entlassen. Die Person, die im Mittel mehr Umsatz pro Tag macht, soll im Betrieb bleiben (s. Tabelle 4.1.4).

	Person A	Person B
Tag 1	2.5	2.4
Tag 2	1.8	15
Tag 3	4	3.9
Tag 4	14	1.9

Tabelle 4.1.4: Daten des Beispiels 4.1.8. Leistungsvergleich mit Mittelwerten

Es gilt $\frac{2.5+1.8+4+14}{4} = 5.575$ für Person A und $\frac{2.4+15+3.9+1.9}{4} = 5.8$ für Person B. Damit hat die Person B einen höheren mittleren Umsatz. $\diamond$

Bemerkung 4.1.9. 1. Da das arithmetische Mittel nicht ein robuster Kennwert ist, müssen wir vor dessen Verwendung überprüfen, ob Ausreisser vorliegen. Nur wenn dies nicht der Fall ist, ist die Verwendung des arithmetischen Mittels sinnvoll.

2. In Häufigkeitsverteilungen mit Klassen, bei denen wir verstreute Daten an den Rändern mit auf einer Seite unendlichen Intervallen zusammengefasst haben, kann das arithmetische Mittel nicht berechnet werden (ausser wir hätten die Urdaten noch zur Verfügung. Alternative kann man auch ein gestutztes Mittel berechnen: die Daten in den Randintervallen werden weggelassen).

3. Wir können das arithmetische Mittel eines Datenvektors $x = (x_1, \dots, x_n)$ als Werte einer Abbildung

$$m : K \rightarrow \mathbb{R} \text{ mit } K = \{x : \text{es gibt ein } n \in \mathbb{N}^*, x \in \mathbb{R}^n\}$$

und

$$m(x) = \frac{1}{n} \sum_{i=1}^n x_i.$$

auffassen. x ist damit ein n -Tupel (=Datenvektor), dem die Abbildung m den jeweiligen Mittelwert $\bar{x}$ zuweist. Funktionen, die einem Datenvektor einen Kennwert zuordnen, werden „Kenngrössen“ genannt. Die Abbildung „arithmetisches Mittel“ ist eine lineare Abbildung.

4. Die Summe der Abweichungen der Daten vom Mittelwert ist 0, denn

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}) = \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{n} \sum_{i=1}^n \bar{x} = \bar{x} - \frac{1}{n} n \bar{x} = \bar{x} - \bar{x} = 0$$

5. Das arithmetische Mittel hat die bemerkenswerte positive Eigenschaft, dass es der Wert a ist, welcher die Summe der quadrierten Abstände minimiert. Es gilt nämlich mit Hilfe quadratischer Ergänzung (s. Herleitung der Formel für die Nullstellen von Polynomen zweiten

Grades):

$$\begin{aligned}
 \sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i^2 - 2ax_i + a^2) \\
 &= \sum_{i=1}^n x_i^2 - 2a \sum_{i=1}^n x_i + na^2 \\
 &= \sum_{i=1}^n x_i^2 - 2a \sum_{i=1}^n x_i + na^2 + \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \\
 &= \sum_{i=1}^n x_i^2 + n \left(-2\frac{a}{n} \sum_{i=1}^n x_i + a^2 + \frac{1}{n^2} \left(\sum_{i=1}^n x_i \right)^2 \right) - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \\
 &= \sum_{i=1}^n x_i^2 + n \left(a - \frac{1}{n} \sum_{i=1}^n x_i \right)^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2
 \end{aligned}$$

Wir möchten nun wissen, wie a zu wählen ist, damit der Ausdruck minimal wird. Der Ausdruck ist offensichtlich minimal, wenn

$$a - \frac{1}{n} \sum_{i=1}^n x_i = 0$$

ist. Damit gilt dann

$$a = \frac{1}{n} \sum_{i=1}^n x_i$$

(s. zur Illustration die Graphik 4.1.1)

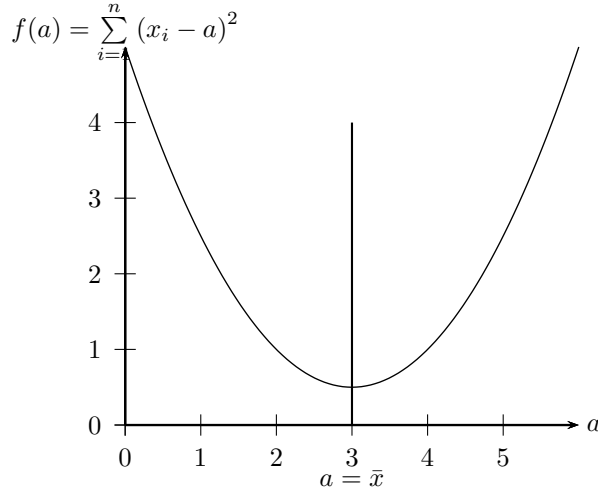


Abbildung 4.1.1: Graphische Darstellung für den Umstand, dass $f(a) = \sum_{i=1}^n (x_i - a)^2$ genau dann minimal wird, wenn $a = \bar{x}$

Intuitiv kann man sich die Bedeutung der Aussage wie folgt klarer machen. Wir spielen folgendes Spiel. Aus einem Datensatz $(x_1, \dots, x_n)$ wird nach Zufallsprinzip eine Zahl x_i gewählt. Das Resultat wird dem Spieler A nicht bekannt gegeben. Der Spieler A muss nun eine Zahl y wählen, wobei er erstens weiß, dass er $(y - x_i)^2$ als Strafe zahlen muss, und er zweitens den Datensatz $(x_1, \dots, x_n)$ kennt. Das Spiel wird 50 mal wiederholt. Welche Zahl y ist jeweils zu wählen, so dass man möglichst wenig Strafe zahlen muss? Antwort: Die Strafzahlung wird minimal, wenn man das arithmetische Mittel wählt.

6. Das arithmetische Mittel kann nur für metrisch skalierte Daten verwendet werden, da die Summenbildung für die Skalenwerte von ordinal- oder nominalskalierten Daten sinnlos ist. Uns fehlen damit immer noch Lagemasse für anders skalierte Daten. Zu beachten ist, dass Statistikprogramme bei codierten Variablen den Mittelwert selbst dann berechnen, wenn dies sinnlos ist. Es liegt am Nutzer, solche Ausgaben als sinnlos auszuschneiden.

◇

Auf Grund der erwähnten Nachteile des arithmetischen Mittels benötigen wir weitere Kennwerte der mittleren Lage.

4.1.2 Der Median

Der Median ist für metrisch und ordinalskalierte Daten geeignet. Der Median ist ein robuster Kennwert. Der Median ist derjenige Skalenwert, der eine der Grösse nach geordnete Reihe von Skalenwerten halbiert:

Definition 4.1.10. $\tilde{x} = \begin{cases} x_k & \text{für } k = (n+1) \cdot 0.5 \\ x_k + (x_{k+1} - x_k) \cdot 0.5 & \text{für } k < (n+1) \cdot 0.5 < k+1 \end{cases}$
 $x_k = k$ -tes Datum des geordneten Datensatzes x ($k \in \mathbb{N}^*$).

Zur Bestimmung des Medians ordnen wir die Skalenwerte der Grösse nach. Der Median ist der $\frac{n+1}{2} = (n+1) \cdot 0.5$ -te Skalenwert (n = Anzahl Daten). Bei geradem n gibt es diesen Wert nicht. Wir wählen das arithmetische Mittel der beiden nächstliegenden Skalenwerte als Median.

Beispiel 4.1.11. Die Zugehörigkeit der Beschäftigten eines Betriebes ergibt folgende Werte in Jahren: 2, 1, 2, 5, 3, 4, 9, 6, 7, 6, 8.

Diese Zahlen müssen zuerst der Grösse nach geordnet werden: 1, 2, 2, 3, 4, 5, 6, 6, 7, 8, 9

$n = 11$. Somit ist die Position des Medians: $\frac{11+1}{2} = 6$

Wir zählen den geordneten Datenvektor von links nach rechts 6 ab und erhalten: 5

Die „mittlere“ Betriebszugehörigkeit beträgt also 5 Jahre.

◇

Wie der Mittelwert weist der Median dieselben Einheiten wie die Daten auf.

Beispiel 4.1.12. Die Zugehörigkeit der Beschäftigten eines Betriebes ergibt folgende Werte in Jahren: 2, 1, 2, 5, 3, 4, 9, 6, 7, 6, 8, 10

Diese Reihe muss zuerst der Grösse nach geordnet werden:

1, 2, 2, 3, 4, 5, 6, 6, 7, 8, 9, 10

$n = 12$. Somit ist die Position des Medians: $\frac{12+1}{2} = 6.5$

Wir zählen den geordneten Datenvektor von links nach rechts 6 ab und erhalten keinen genauen Wert. Wir wählen das arithmetische Mittel der beiden Zahlen nächstliegenden Zahlen 5 und 6: $\frac{5+6}{2} = 5.5$

◇

Bei ordinalskalierten Daten macht die Berechnung des arithmetischen Mittels eigentlich keinen Sinn. Man kann die obige Anleitung in diesem Falle einfach als eine Methode betrachten, die zu einem jeweils eindeutigen Ergebnis kommt. Im Grunde genommen wäre jeder Wert zwischen 5 und 6 genau so geeignet, um den Median bei ordinalskalierten Daten zu repräsentieren.

Beispiel 4.1.13. 10 Personen wurden hinsichtlich ihres monatlichen Bruttoeinkommens untersucht. Man erhielt folgende Werte in Fr. (bereits geordnet):

1416, 2340, 2578, 2870, 3038, 3420, 3510, 3890, 4329, 12800

Position des Medians $\frac{10+1}{2} = 5.5$

$\tilde{x} = \frac{3038+3420}{2} = 3229$

◇

Beispiel 4.1.14. 9 Personen werden über die Beurteilung des Preises eines bestimmten Produktes befragt. Folgende Beurteilungen werden angeboten: günstig = 1, angemessen = 2, zu teuer = 3,

viel zu teuer = 4

Die geordneten Resultate hätten folgendes ergeben: 1, 2, 2, 3, 3, 3, 3, 4, 4

Position des Medians $\frac{9+1}{2} = 5 \implies \tilde{x} = 3$

Das Produkt wird im Mittel als zu teuer beurteilt. Die Firma muss sich überlegen, entweder den Preis zu senken oder am Image des Produktes zu arbeiten. $\diamond$

Um die Robustheit des Medians zu untersuchen, betrachten wir nochmals *Beispiel 4.1.6*:

Beispiel 4.1.15. Für die Daten 1, 2, 3, 4, 5, ist der Mittelwert:

$$\frac{1 + 2 + 3 + 4 + 5}{5} = 3$$

und der Median = 3.

Für die Daten 1, 2, 3, 4, 1000 ist der Mittelwert:

$$\frac{1 + 2 + 3 + 4 + 1000}{5} = 202$$

und der Median 3. Durch den Ausreisser wurde der Median überhaupt nicht beeinflusst.

Damit eignet sich der Median für Datensätze, für die der Mittelwert versagt, um die mittlere Lage von Daten anzugeben. Berechnet man beide Kennwerte, ohne vorgängige Überprüfung auf Ausreisser, weist eine grosse Differenz zwischen beiden Werten auf Ausreisser hin. $\diamond$

Wollen wir den Median für klassierte Daten berechnen, so gehen wir wie folgt vor: Wir berechnen $\frac{n+1}{2}$ und schauen, in welcher Klasse K das $\frac{n+1}{2}$ -te Datum zu liegen kommt. Es gibt nun verschiedene, übliche Verfahren:

(i) Wir wählen die Klassenmitte von K als Median.

(ii) Sei m die Zahl der Daten, die in tiefere Klassen zu liegen kommen und k die Anzahl der Daten der Klasse K , in die der Median zu liegen kommt und u_K die untere Grenze des Intervalls, in das der Median zu liegen kommt, und l_K die Intervalllänge des Intervalls K

Wir bestimmen den Median als:

$$u_K + \frac{\left(\frac{n+1}{2} - m\right)}{k} l_K,$$

d.h. wir legen den Median anteilmässig im Intervall K fest.

Beispiel 4.1.16. Betrachten wir noch einmal die Einkommensverteilungen aus *Beispiel 4.1.7* (s. *Tabelle 4.1.5*):

	Anzahl Steuerpflichtige		Intervallmitten	Produkte	
Intervalle	Dorf A	Dorf B		Dorf A	Dorf B
[0, 10'000]	2	0	5000	10000	0
]10'000, 20'000]	10	2	15000	150000	30000
]20'000, 30'000]	35	12	25000	875000	300000
]30'000, 40'000]	30	24	35000	1050000	840000
]40'000, 50'000]	14	42	45000	630000	1890000
]50'000, 60'000]	8	15	55000	440000	825000
1'100'000	1	0	1100000	1100000	0
Summen	100	95	1280000	4255000	3885000
Mittelwerte				42550	40894.73684

Tabelle 4.1.5: Daten von *Beispiel 4.1.7* Berechnung des Medians

Dorf A: Der Median kommt ins Intervall $[30'000, 40'000[$ zu liegen, da $2 + 10 + 35 = 47$ und $2 + 10 + 35 + 30 = 77$.

Methode 1: Wir wählen nun die Klassenmitte: Median = 35'000.

Methode 2: $u_K + \frac{\left(\frac{n+1}{2} - m\right)}{k} l_K = 30000 + \frac{\frac{100+1}{2} - 47}{30} 10000 = 31167$.

Dorf B: Der Median kommt ins Intervall $[40'000, 50'000]$ zu liegen, da $0 + 2 + 12 + 24 = 38.0$ und $0 + 2 + 12 + 24 + 42 = 80.0$

Methode 1: $\tilde{x} = 45'000$.

Methode 2: $\tilde{x}' = u_K + \frac{\left(\frac{n+1}{2} - m\right)}{k} l_K = 40000 + \frac{\frac{96}{2} - 38}{42} \cdot 10000 = 42381$.

Sind die Daten gleichmässig auf das Intervall verteilt, so ergibt die zweite Methode kleinere Abweichungen vom wirklichen Median. $\diamond$

Sind nur die relativen Häufigkeiten gegeben, so kann man mit der folgenden Formel rechnen:

$$u_K + \frac{0.5 - \frac{m}{n}}{\frac{k}{n}} l_K$$

wobei $\frac{m}{n}$ und $\frac{k}{n}$ relative Häufigkeiten sind - die Buchstaben sind wie oben zu interpretieren. Es gilt nämlich

$$u_K + \frac{\left(\frac{n+1}{2} - m\right)}{k} l_K = u_K + \frac{\frac{1}{n} \left(\frac{n+1}{2} - m\right)}{\frac{1}{n} k} l_K = \frac{\left(\frac{n+1}{n2} - \frac{m}{n}\right)}{\frac{k}{n}} l_K \approx \frac{\left(\frac{1}{2} - \frac{m}{n}\right)}{\frac{k}{n}} l_K.$$

denn $\frac{n+1}{n2} = \frac{n}{2n} + \frac{1}{n2} = \frac{1}{2} + \frac{1}{2n}$ und $\frac{1}{2n}$ wird bei genügend grossem n vernachlässigbar klein.

Bemerkung 4.1.17. Wir können den Median auffassen als den Wert der Medianabbildung $\text{med}: K \rightarrow \mathbb{R}$ mit $K = \{x : \text{es gibt ein } n \in \mathbb{N}^*, x \in \mathbb{R}^n\}$ und

$\text{med}(x) := \tilde{x}$

x ist damit ein n -Tupel, dem die Abbildung med den jeweiligen Median $\tilde{x}$ zuweist. $\diamond$

4.1.3 Der Modalwert

Zuletzt betrachten wir noch einen Kennwert, die für beliebig skalierte Daten verwendbar ist. Der Modalwert ist ein robuster Kennwert. Der Modalwert ist derjenige Skalenwert, der in einer Messreihe am häufigsten vorkommt (der Modalwert ist nur bestimmt, wenn es genau einen Skalenwert mit dem häufigsten Vorkommen gibt). Der Modalwert wird auch „Modus“ genannt.

Definition 4.1.18. $\text{Modalwert}(x) = x_i$ genau dann, wenn $k(x_i) = \max\{k(x_j) : x_j \in \text{Bild}(f)\}$ (max ist nur bestimmt, wenn es genau einen Wert gibt, der in der Menge maximal ist. f ist die Skala, k die Häufigkeitsverteilung der betrachteten Variable)

Beispiel 4.1.19. Aus einer Gruppe von 30 Personen an einer Tagung interessiert den Marketingchef eines Buchverlages die Nationalität dieser Personen. Er unterscheidet zwischen Deutschen (D), Schweizern (S), Österreichern (A) und anderen Nationalitäten (ND). Die Erhebung ergibt die Tabelle 4.1.6 (s. auch Abbildung 4.1.2):

Nationalität	Häufigkeit
A	6
D	8
S	10
ND	6

Tabelle 4.1.6: Häufigkeitstabelle für die Berechnung eines Kennwertes der mittleren Lage

Modalwert = S

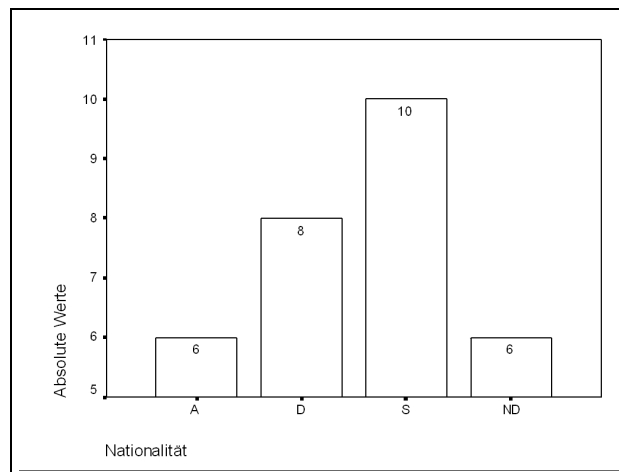


Abbildung 4.1.2:

◇

Beispiel 4.1.20. Vorliegen zweier Maxima (s. Tabelle 4.1.7) Es gibt zwei Ausprägungen, mit maximaler Häufigkeit. Entsprechend ist der Modalwert nicht bestimmt.

Betriebszugehörigkeit in Jahren	Anzahl Angestellte
3	1
4	2
5	7
6	5
7	6
8	7
9	4

Tabelle 4.1.7: Häufigkeitsverteilung mit zwei Maxima

◇

Bei ordinalen oder metrischen Daten ist es sinnvoll, von zweigipfligen oder mehrgipfligen (bimodalen, multimodalen) Verteilungen zu sprechen, wenn zwischen zwei (oder mehreren) Ausprägungen mit grossen Häufigkeiten Ausprägungen mit weniger grossen Häufigkeiten vorkommen (im Beispiel 4.1.20 könnte man somit von einer bimodalen Häufigkeitsverteilung sprechen).

Beispiel 4.1.21. Ein Würfel wird 1000 mal geworfen (s. Tabelle 4.1.8 und Abbildung 4.1.3):

1	165
2	175
3	159
4	167
5	155
6	179

Tabelle 4.1.8: Häufigkeitsverteilung eines 1000fachen Würfelwurfs

Der Modalwert ist 6.

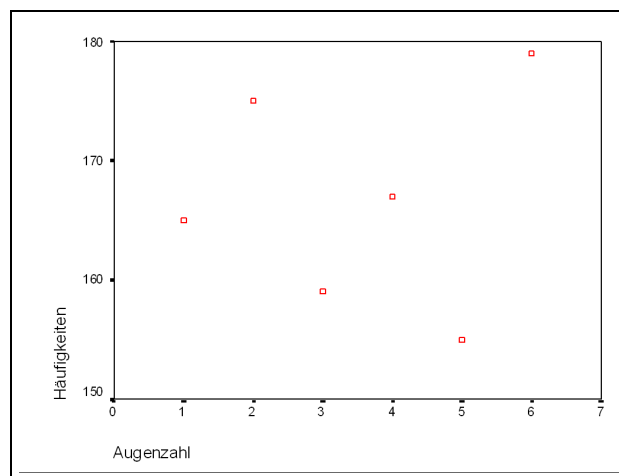


Abbildung 4.1.3: Graphische Darstellung der Häufigkeitsverteilung von Beispiel 4.1.21 im Koordinatennetz

◇

An den Rändern einer Verteilung werden oft grössere Intervalle verwendet als im Zentrum einer Verteilung. Werden endliche Intervalle an den Rändern verwendet, kann man die Häufigkeiten trotzdem vergleichen, wenn man diese an der Intervalllänge misst:

Wir berechnen die folgende Grösse: $D = \frac{\text{Häufigkeit}}{\text{Intervalllänge}}$

Der Modalwert befindet sich in der Klasse mit dem grössten D (nur für metrisch skalierte Daten sinnvoll).

Beispiel 4.1.22. *Arbeiter eines Unternehmens, gegliedert nach Leistungsklassen mit ungleich langen Intervallen (s. Tabelle 4.1.9): Der Modalwert liegt in der Mitte des Intervalls mit dem grössten D .*

Leistung in Stück	Intervallbreite	Anzahl Arbeiter	D
[270, 300]	30	48	1.6
]300, 310]	10	35	3.5
]310, 320]	10	28	2.8
]320, 330]	10	33	3.3
]330, 360[	30	22	0.73333333

Tabelle 4.1.9: Berechnung des Modalwert bei ungleich langen Intervallen.

Der Modalwert ist 305. Diese Methode ist nur anwendbar, wenn nicht unendlich grosse Intervalle an den Rändern verwendet werden. Kennt man das erste und das letzte Datum, kann man die Intervallgrenzen entsprechend festlegen. $\diamond$

Bemerkung 4.1.23. 1. Der Modalwert ist nur dann sinnvoll, wenn ein eindeutiges Maximum vorhanden ist.

2. Vorteil des Modalwert ist die einfache Bestimmbarkeit. Er hat dort seine Bedeutung, wo es auf schnelles Erkennen von zentralen Tendenzen ankommt.
3. Der Modalwert ist robust: die übrigen Werte beeinflussen ihn nicht, solange der Wert mit den meisten Ausprägungen nicht überflügelt wird.
4. Da wir den Modalwert nur zuordnen, wenn es ein eindeutiges Maximum gibt, kann man den Modalwert nicht als den Wert einer Abbildung $\text{Modalwert} : K \rightarrow \mathbb{R}$ mit $K = \{x : \text{es gibt ein } n \in \mathbb{N}^*, x \in \mathbb{R}^n\}$ betrachten (siehe entsprechende Bemerkungen zum arithmetischen Mittel und zum Median). Es liegt allerdings eine Funktion (die nicht Abbildung ist) vor: $\text{Modalwert}(x) = x_i$ genau dann, wenn es genau ein i gibt, so dass $k(x_i) > k(x_j)$ für alle $j \neq i$. Definitionsbereich ist das oben definierte K . Der Wertebereich ist $\mathbb{R}$.

 $\diamond$

4.1.4 Übungen

1. Aus einer Statistik ergab sich, dass 7 Personen sich im Jahre 1999 in den Glarner Alpen verirrt haben (s. Tabelle 4.1.10):

Person	Geschlecht	Alter (Jahre)	Religion
1	m	17	C
2	m	30	P
3	m	73	C
4	w	35	C
5	m	21	C
6	m	15	P
7	w	19	A

Tabelle 4.1.10: Tabelle der Urdaten mit drei Variablen: Bestimmung des geeigneten Kennwertes der mittleren Lage

Geschlecht: „m“ für „männlich“, „w“ für „weiblich“. Religion: „C“ für „Cerls“ und „P“ für „Perls“, „A“ für „andere“.

Bestimmen Sie die geeigneten Kennwerte mittlerer Lage für diese drei Variablen und berechnen Sie sie.

2. Eine Firma produziert Schrauben. Dabei ergeben sich die Längen in cm der Tabelle 4.1.11.

Intervalle	Relative Häufigkeiten
[0, 20]	0.15
]20, 40]	0.2
]40, 60]	0.3
]60, 80]	0.2
]80, 100]	0.1
]100, 200]	0.05

Tabelle 4.1.11: Relative Häufigkeitsverteilung klassierter Daten: Berechnung geeigneten Kennwertes

Berechnen Sie die geeigneten Kennwerte der mittleren Lage.

3. Berechnen Sie von Hand, mit Excel (ohne die Funktionenbefehle), mit Excel (mit den Funktionenbefehlen) und mit einer Statistiksoftware die jeweils passenden Kennwerte der mittleren Lage (Mittelwert, Median, Modalwert)
- a) Bei zehn verschiedenen Anbietern kostet ein Teelöffel folgenden Betrag:
5 6 3 4.5 4.7 5 6 2 8 7
- b) Bei zehn verschiedenen Anbietern kostet ein Teelöffel folgenden Betrag:
5 6 3 4.5 4.7 5 6 2 8 88
- c) (i) Bei einem Rating werden die acht Mitarbeiter wie folgt nach Ihrer Leistung eingeschätzt
3, 3.5, 4, 5.5, 5.8, 4.5, 6, 7
(ii) Ein anderer Personalchef kommt zur folgende Einschätzung:
2.8, 4, 4.2, 5.8, 7, 6.5, 6.5, 8
Welcher der beiden Chefs beurteilt im Schnitt strenger?
4. Schauen sich das Video <http://www.mathe-online.at/clips/mwstdabw/index.html> an.

4.1.5 Lösungen

1. *Geschlecht*: Modalwert: m
Alter: $\bar{x} = \frac{17+30+73+35+21+15+19}{7} = 30$; Allerdings wird der Mittelwert durch den Ausreisser 73 hinaufgedrückt, wodurch der Mittelwert ein verfälschtes Bild über die mittlere Lage der Daten gibt. Entsprechend müsste der Median verwendet werden: 15, 17, 19, 21, 30, 35, 73; $\frac{7+1}{2} = 4$. Der Median liegt beim 4. Datum: Dies ist 21. Der Modalwert kann ohne vorangehende Klassifizierung der Daten nicht berechnet werden. Man könnte z.B. folgende Klassen festlegen: [0,20],]20,40],]40, 60],]60, 80] und erhalten für die jeweiligen Klassenmittel die folgende Häufigkeitsverteilung: $\{(10, 3), (30, 3), (50, 0), (70, 1)\}$. Bei dieser Klassenbildung ist der Modalwert nicht bestimmt. Wählen wir jedoch 10er Intervalle, so würde sich ein Modalwert ergeben (nämlich für 15 in]10, 20].
Religion: Modalwert C
2. $\bar{x} = 10 \cdot 0.15 + 30 \cdot 0.2 + 50 \cdot 0.3 + 70 \cdot 0.2 + 90 \cdot 0.1 + 150 \cdot 0.05 = 53$
 Der Median kommt ebenfalls in die Klasse]40, 60] zu liegen. Nach der Methode des arithmetischen Mittels, wäre der Median 50, nach der Methode der Anteile ebenfalls 50, da 0.5 genau zwischen 0.35 und 0.65 liegt.
 Modalwert: Intervall [40, 60[oder Intervallmitte 50.

3. a) Es kommt kein Ausreisser vor, somit ist der Mittelwert und der Median geeignet.

$$\bar{x} = \frac{5+6+3+4.5+4.7+5+6+2+8+7}{10} = 5.12$$

$$\text{Median bei: } \frac{11}{2} = 5.5$$

geordnete Daten: 2 3 4.5 4.7 5 5 6 6 7 8

$$\tilde{x} = 5$$

Der Modalwert ist ohne vorgängige Klassifizierung nicht angebbbar. Bei folgender Intervallbildung: $[2,4[$, $[4,6[$, $[6,8[$, $[8,10]$ ergibt sich die folgende Häufigkeitsverteilung, wobei wir als Komponenten an erster Stelle die Klassenmitten verwenden: $\{(3, 2), (5, 4), (7, 3), (9, 1)\}$. Bei dieser Klassenbildung ist der Modalwert 5.

- b) Es liegt nun ein Ausreisser vor, der die Verwendung des arithmetischen Mittels fragwürdig macht. Wir überprüfen, ob die Abweichung vom Median bedeutend ist:

$$\frac{5+6+3+4.5+4.7+5+6+2+8+88}{10} = 13.22$$

Median: geordnete Daten: 2 3 4.5 4.7 5 5 6 6 7 88

$$\tilde{x} = 5$$

Der Mittelwert liegt im Gegensatz zum Median nicht in der ungefähren Mitte des Intervalls, wo die meisten Daten liegen. Entsprechend dürfte der Mittelwert hier nicht verwendet werden.

Der Modalwert ist ohne vorgängige Klassifizierung nicht angebbbar. Bei folgender Intervallbildung: $[2,4[$, $[4,6[$, $[6,8[$, $[8,90]$ ergibt sich die folgende Häufigkeitsverteilung, wobei wir als Komponenten an erster Stelle die Klassenmitten verwenden: $\{(3,2), (5,4), (7,3), (41,1)\}$. Wir müssten nun $D = \frac{k(x)}{\text{Intervallbreite}}$ berechnen, da die Klassenbreiten nicht immer gleich gross sind: $\frac{2}{2} = 1$; $\frac{4}{2} = 2$; $\frac{3}{2} = 1.5$; $\frac{1}{82} = 0.012195$. Der Modalwert liegt bei 5.

- c) (i) 3, 3.5, 4, 4.5, 5.5, 5.8, 6, 7,9

Es liegt offenbar eine Ordinalskala vor. Der Mittelwert kann nicht verwendet werden. Die Daten sind bereits geordnet. $\frac{10}{2} = 5$. Der Median ist 5.5.

Der Modalwert liesse sich nach vorgängiger Klassifikation bestimmen. z.B. $[3, 6]$, $[6, 7]$, $[7, 10]$. Dies ergibt die folgende Häufigkeitsverteilung: $\{(4.5, 7), (6.5, 1), (8.75, 1)\}$. Der Modalwert liegt bei 4.5. Zu beachten ist, dass bei Vorliegen einer Ordinalskala, der Ausdruck „Intervallbreite“ sinnlos ist und somit die Intervalle völlig willkürlich festgelegt werden. Entsprechend können Resultate auch manipuliert werden. Ebenso sinnlos ist der Ausdruck „Intervallmitte“.

- (ii) Wir ordnen die Daten:

2.8, 4, 4.2, 5.8, 6.5, 6.5, 7, 8. Der Median ist das 4.5. Datum. Wir erhalten: $5.8 + \frac{6.5-5.8}{2} = 6.15$. Der erste Chef ist strenger.

4.2 Streuungskennwerte

Die Verteilung von Daten kann sich nicht nur bezüglich der mittleren Lage, sondern auch bezüglich der Streuung unterscheiden. Die folgenden Daten seien aus zwei verschiedenen Populationen entnommen worden.

Beispiel 4.2.1. *Zwei Datensätze (klassierte Daten) mit identischem Mittelwert und verschiedenen Streuungen. Datensatz 1: Zeit in Monaten bis zur ersten Reparatur bei Gegenständen von der Produktionsstätte 1 fabriziert (s. Tabellen 4.2.12 und Abbildung 4.2.4); Datensatz 2: Zeit bis zur ersten Reparatur bei Gegenständen von der Produktionsstätte 2 fabriziert (s. Tabelle 4.2.13 und Abbildung 4.2.5)*

Intervall	Intervallmitte	Häufigkeit
[3, 4]	3.5	2
]4, 5]	4.5	2
]5, 6]	5.5	2
]6, 7]	6.5	2
]7, 8]	7.5	2
]8, 9]	8.5	2
]9, 10]	9.5	2
Summen		14

Tabelle 4.2.12: Datensatz 1 (Produktion 1):
Zeit bis erste Reparatur

Intervall	Intervallmitte	Häufigkeit
[3, 4]	3.5	0
]4, 5]	4.5	1
]5, 6]	5.5	3
]6, 7]	6.5	6
]7, 8]	7.5	3
]8, 9]	8.5	1
]9, 10]	9.5	0
Summen		14

Tabelle 4.2.13: Datensatz 2 (Produktion 2):
Zeit bis erste Reparatur

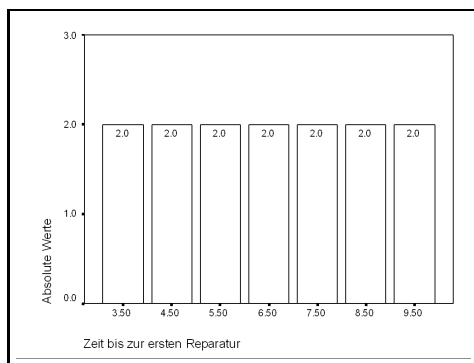


Abbildung 4.2.4: Datensatz 1: Zeit bis erste Reparatur ($\bar{x} = 6.5$), s. Tabelle 4.2.12 Spalte 3

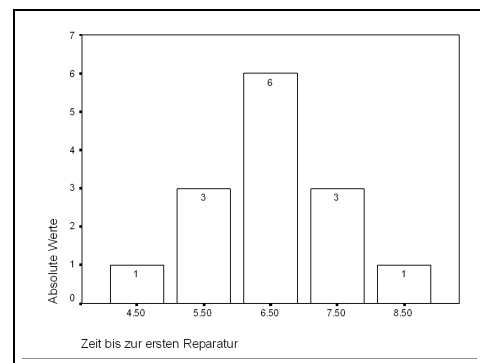


Abbildung 4.2.5: Datensatz 2: Zeit bis ersten Reparatur ($\bar{x} = 6.5$), s. Tabelle 4.2.13 Spalte 4

Obwohl die mittlere Lage (arithmetisches Mittel) für beide Verteilungen identisch ist, unterscheiden sie sich durch die Form der Verteilung erheblich. Beim Datensatz 2 liegen die Daten näher beim Mittelwert als beim Datensatz 1. $\diamond$

Beispiel 4.2.2. *Zwei verschiedene Stahllegierungen A und B sollen in einer Versuchsanordnung hinsichtlich ihrer Zugfestigkeit überprüft werden. Es werden von jeder Legierung 10 gleiche Stücke ausgewählt und es wird gemessen, bei welchem angehängten Gewicht ein solches Stück reisst. Die Untersuchung ergibt folgendes Resultat (s. Tabellen 4.2.14 und 4.2.15 sowie Abbildungen 4.2.6 und 4.2.7):* $\diamond$

Objekt	Angehängtes Gewicht in t
1	1.5
2	3.6
3	3.7
4	3.7
5	3.8
6	3.9
7	4
8	4.1
9	4.2
10	5.5

Tabelle 4.2.14: Datensatz A

Datensatz 2 Objekt	Angehängtes Gewicht in t
1	2
2	2.5
3	2.9
4	3.4
5	3.8
6	4.1
7	4.5
8	4.6
9	5
10	5.2

Tabelle 4.2.15: Datensatz B

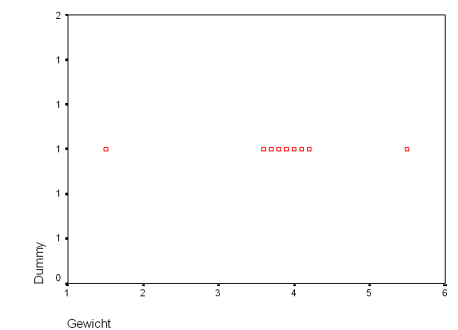


Abbildung 4.2.6: Graphische Darstellung der Daten der Tabelle 4.2.14 als eindimensionales Streudiagramm

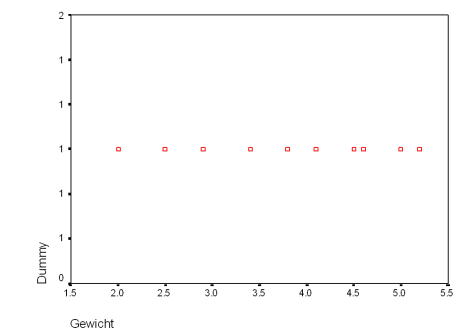


Abbildung 4.2.7: Graphische Darstellung der Daten der Tabelle 4.2.15 als eindimensionales Streudiagramm

Auch in diesem Beispiel ergeben sich identische Mittelwerte, jedoch recht verschiedene Streuungen, wie die eindimensionalen Streudiagramme zeigen. Sollte die Reißfestigkeit möglichst gleichmässig sein, so wäre die Qualität mit der kleineren Streuung vorzuziehen.

Es geht nun darum, für diese Unterschiede, die wir „Streuungsunterschiede“ nennen, Kennwerte zu entwickeln. Die Streuungs-Kennwerte sollen Auskunft geben, wie „eng“ die Messwerte eines metrisch skalierten Merkmals um die zentrale Lage beieinander liegen.

4.2.1 Die Spannweite

Metrische Daten, nicht robust. Sei x der Vektor, der die Messwerte von n Daten enthalte. $\min(x) := \min\{x_i : x_i \text{ ist eine Komponente von } x\}$. Analog für $\max(x)$.

Definition 4.2.3. Spannweite des Datenvektors $x = (x_1, \dots, x_n) := \max(x) - \min(x)$

Die Spannweite ist der Abstand zwischen dem kleinsten und dem grössten Messwert.

Beispiel 4.2.4. Wir bestimmen die beiden Spannweiten des Beispiels 4.2.2:

Für den Datensatz A: $\text{Spannweite} = 5.5 - 1.5 = 4$

Für den Datensatz B: $\text{Spannweite} = 5.2 - 2 = 3.2$

◇

Mit diesem einfachen Streumass kann bestimmt werden, wie gross das Intervall ist, in welchem die Messwerte liegen. Der Mangel an Robustheit ist offensichtlich: ein einziger Ausreisser beeinflusst die Spannweite stark. So ist die Spannweite bei des Datensatzes (1, 2, 3, 4); $4 - 1 = 3$ und

des Datensatzes $(1, 2, 3, 1000)$: 999. Die Spannweite ist nur für metrische Daten geeignet, da der Abstandsbegriff für die Nominalskala und die Ordinalskala nicht definiert ist. Bei der Ordinalskala können wir statt der Spannweite einfach den kleinsten und den grössten Wert angeben.

Bemerkung 4.2.5. *Der maximale und der minimale Wert eines Datensatzes können als Lagemasse betrachtet werden. Sie geben an, in welchem Intervall die Daten liegen.* $\diamond$

4.2.2 Mittlerer Abweichungsbetrag

Metrische Daten, nicht robust.

Nachteil der Spannweite ist, dass nur die beiden Extremwerte berücksichtigt werden. Stellen Sie sich vor, der Stahlabnehmer braucht in seinem Betrieb eine gleichmässige Qualität, wenige Extremwerte stören ihn weniger. Er würde sich wahrscheinlich für die Legierung A entscheiden, obwohl die Spannweite für diese Legierung grösser ist. Wir definieren deshalb ein Streumass, das alle Daten berücksichtigt. Es gibt verschiedene Kennwerte, die alle Daten berücksichtigen. Man kann z.B. das arithmetische Mittel des Betrags der Abweichungen vom arithmetischen Mittel berechnen:

Definition 4.2.6. Mittlerer Abweichungsbetrag $= \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$

(für den Datensatz $x = (x_1, \dots, x_n)$)

Der mittlere Abweichungsbetrag wird manchmal auch „mittlere Abweichung“ genannt. Das Betragszeichen müssen wir verwenden, da die mittleren Abweichungen ohne Betragszeichen - wie oben gezeigt (s. Seite 47) - mit 0 identisch ist.

Der mittlere Abweichungsbetrag vom Mittelwert des Datensatzes A des Beispiels 4.2.2 beträgt laut Definition: $\frac{1}{10}(|1.5 - 3.8| + |3.6 - 3.8| + 2 * |3.7 - 3.8| + |3.8 - 3.8| + |3.9 - 3.8| + |4 - 3.8| + |4.1 - 3.8| + |4.2 - 3.8| + |5.5 - 3.8|) = 0.54$

Der mittlere Abweichungsbetrag vom Mittelwert des Datensatzes B des Beispiels 4.2.2 beträgt laut Definition: $\frac{1}{10}(|2 - 3.8| + |2.5 - 3.8| + |2.9 - 3.8| + |3.4 - 3.8| + |3.8 - 3.8| + |4.1 - 3.8| + |4.5 - 3.8| + |4.6 - 3.8| + |5 - 3.8| + |5.2 - 3.8|) = 0.88$

Wir sehen: der Datensatz von Tabelle 4.2.15 streut mehr als der Datensatz von Tabelle 4.2.14.

Ein alternativer mittlerer Abweichungsbetrag könnte bei metrisch skalierten Daten auch mit dem Median als Kennwert der mittleren Lage berechnet werden. Der mittlere Abweichungsbetrag ist nur für metrische Daten geeignet, da Abstände in die Berechnung einfließen. Der mittlere Abweichungsbetrag ist nicht robust, da die Abstände der Daten vom jeweiligen Mittel verwendet werden. Ein einzelner Ausreisser hat damit ein grosses Gewicht. Der mittlere Abweichungsbetrag für $\{1, 2, 3, 4, 5\}$: $\frac{1}{5}(|1 - 3| + |2 - 3| + |3 - 3| + |4 - 3| + |5 - 3|) = 1.2$. Der mittlere Abweichungsbetrag für $\{1, 2, 3, 4, 1000\}$: $\frac{1}{5}(|1 - 202| + |2 - 202| + |3 - 202| + |4 - 202| + |1000 - 202|) = 319.2$

Beispiel 4.2.7. *Ein Unternehmer verkauft ein Produkt an Grosshändler (A), Fachhändler (B) und an Einzelhändler (C). Am Ende einer Berichtsperiode liegen folgende Ergebnisse vor (s. Tabelle 4.2.16):*

Abnehmer	Preis/St.	Anzahl
A	6	70000
B	8	30000
C	9	30000
Summen		130000

Tabelle 4.2.16: Beispieldaten für die Berechnung der durchschnittlichen Abweichung bei klassierten Daten

Wie gross ist der durchschnittliche Abweichungsbetrag der erzielten Stückpreise in Bezug auf den Durchschnittspreis?

$$\bar{x} = \frac{1}{70000+30000+30000}(6 \cdot 70000 + 8 \cdot 30000 + 9 \cdot 30000) = 7.153846154$$

Mittlerer Abweichungsbetrag =

$$\frac{1}{70000+30000+30000}((70000 \cdot |6 - 7.153846154|) + (30000 \cdot |8 - 7.153846154|) + (30000 \cdot |9 - 7.153846154|)) = 1.24260355$$

◇

Beispiel 4.2.8. Der Chef eines Unternehmens möchte sich über die täglich anfallenden Anfahrtswege seiner Mitarbeiter zum Arbeitsplatz informieren. Eine Umfrage ergibt folgende Daten (s. Tabelle 4.2.17):

km	Anzahl Beschäftigte
[0, 5]	7
]5, 10]	24
]10, 15]	35
]15, 20]	18
]20, 30]	16

Tabelle 4.2.17: Beispieldaten für Berechnung des durchschnittlichen Abweichungsbetrags (klassierte Daten)

Wie gross ist der mittlere Abweichungsbetrag vom durchschnittlichen Weg?

$$n = 7 + 24 + 35 + 18 + 16 = 100$$

$$\bar{x} = \frac{1}{100}(2.5 \cdot 7 + 7.5 \cdot 24 + 12.5 \cdot 35 + 17.5 \cdot 18 + 25 \cdot 16) = 13.5$$

Der mittlere Abweichungsbetrag =

$$\frac{1}{100}((7 \cdot |2.5 - 13.5|) + (24 \cdot |7.5 - 13.5|) + (35 \cdot |12.5 - 13.5|) + (18 \cdot |17.5 - 13.5|) + (16 \cdot |25 - 13.5|)) = 5.12$$

◇

4.2.3 Varianz und Standardabweichung

Metrisch skalierte Daten. Nicht robust.

Der mittlere Abweichungsbetrag ist ein sehr anschauliches Streumass und kann in der rein beschreibenden Betriebsstatistik durchaus Verwendung finden. Es ist sonst jedoch nicht gebräuchlich. Beim Entwickeln von Formeln für anspruchsvollere Probleme in der schliessenden Statistik sind die Absolut-Beträge nämlich recht umständlich. Deshalb benutzt man statt des durchschnittlichen Abweichungsbetrags die Varianz. Es handelt sich um die an der Datenhäufigkeit gemessene Summe der quadrierten Abstände der Daten vom Mittelwert.

Definition 4.2.9. Varianz $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
(für den Datenvektor $x = (x_1, \dots, x_n)$)

Durch das Quadrieren erhalten wir nur positive Summanden. Zudem werden grössere Abstände stärker gewichtet als kleinere Abstände. Aus Gründen, die wir im Rahmen der schliessenden Statistik anschauen werden, wird durch $\frac{1}{n-1}$ statt durch $\frac{1}{n}$ geteilt. Für grosse n unterscheiden sich die beiden Zahlen kaum mehr.

Die Varianz eignet sich nur für metrisch skalierte Variablen, da in deren Berechnung Abstände verwendet werden. Die Varianz ist nicht robust, da in ihr der nichtrobuste Mittelwert vorkommt und zudem quadrierte Abweichungen zusammengezählt werden. Eine einzige sehr starke Abweichung hat deshalb einen grossen Einfluss auf die Varianz:

Beispiel 4.2.10. Die Varianz von $(1, 2, 3, 4, 5)$:

$$s^2 = \frac{1}{4} ((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2) = \frac{5}{2} = 2.5$$

Die Varianz von $\{1, 2, 3, 4, 1000\}$:

$$s^2 = \frac{1}{4} ((1-202)^2 + (2-202)^2 + (3-202)^2 + (4-202)^2 + (1000-202)^2) = \frac{398\,005}{2}. \quad \diamond$$

Beispiel 4.2.11. Daten der Tabelle 4.2.14

$$s^2 = \frac{1}{10-1} ((1.5-3.8)^2 + (3.6-3.8)^2 + (3.7-3.8)^2 + (3.7-3.8)^2 + (3.8-3.8)^2 + (3.9-3.8)^2 + (4-3.8)^2 + (4.1-3.8)^2 + (4.2-3.8)^2 + (5.5-3.8)^2) = 0.948888889 \quad \diamond$$

Die Masseinheit der Varianz ist das Quadrat der Masseinheiten der Daten. Darum führen wir eine weitere Kennwert ein, die dieselbe Masseinheit wie die Daten hat und damit leichter interpretierbar ist: die Standardabweichung s . Dieser Kennwert ist die Wurzel aus der Varianz.

Definition 4.2.12. Standardabweichung $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$

(für den Datensatz $x = (x_1, \dots, x_n)$)

Damit ist die Standardabweichung des Datensatzes der Tabelle 4.2.14 (Beispiels 4.2.11): $s = \sqrt{0.948888889} = 0.97410928$

Die Standardabweichung kann in Streudiagramme oder Histogramme eingetragen werden, da sie dieselben Masseinheiten wie die Daten aufweist. Der Mittelwert der Daten der Tabelle 4.2.14 liegt bei 3.8. Wir können auf beiden Seiten des Mittelwertes die Standardabweichung 0.97410928 abtragen und erhalten (s. Abbildung 4.2.8):

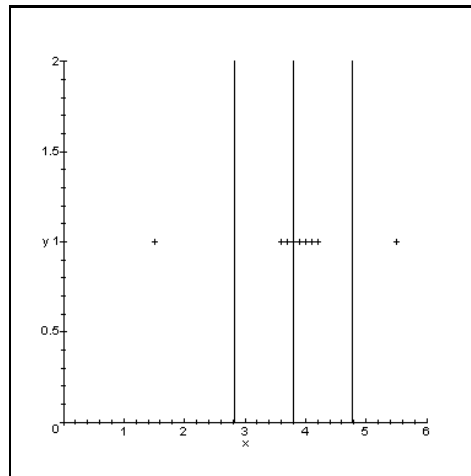


Abbildung 4.2.8: Eindimensionales Streudiagramm eines Datensatzes mit arithmetischem Mittel und Abtragung der Standardabweichung auf beiden Seiten des Mittelwertes

Im Beispiel 4.2.11 liegen damit nur 2 Daten ausserhalb des Intervalls $[\bar{x} - s, \bar{x} + s]$. Dies hängt allerdings mit der spezifischen Verteilung der Daten des Beispiels zusammen.

Betrachten wir noch die entsprechende Graphik der Tabelle 4.2.15 (s. Abbildung 4.2.9).

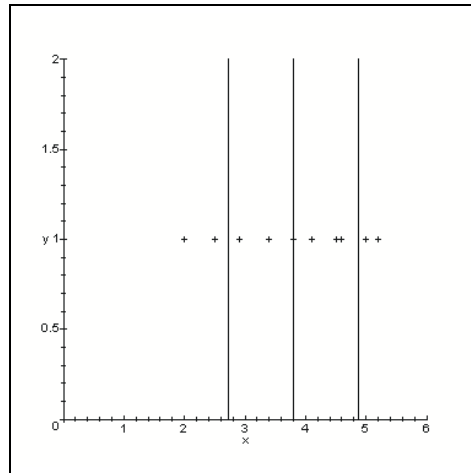


Abbildung 4.2.9: Eindimensionales Streudiagramm eines Datensatzes mit arithmetischem Mittel und Abtragung der Standardabweichung auf beiden Seiten des Mittelwertes

Die Daten sind recht gleichmässig über das gesamte Intervall, in dem Daten vorkommen, gestreut. Entsprechend fallen mehr Werte ausserhalb des Intervalls $[\bar{x} - s, \bar{x} + s]$ mit $\bar{x} = 3.8$ und $s = 1.081151649$

Bemerkung 4.2.13. Auch die Streuungskennwerte können jeweils als Werte von Kenngrössen (= Funktionen) betrachtet werden, die einem Datenvektor jeweils den entsprechenden Kennwert zuordnen. So ordnet z.B. die Kenngrösse „Varianz“ einem Datenvektor die konkrete Varianz s^2 der Daten des Vektors zu. Diese Sichtweise von Kennwerten ist für das Verständnis der später zu behandelnden schliessenden Statistik wichtig. $\diamond$

Manchmal gibt man die Streuung als Verhältnis der Standardabweichung zum Mittelwert an: Dieser Kennwert wird „Variationskoeffizient“ genannt (nur für Verhältnisskala geeignet, d.h. für Skalen mit natürlichem Nullpunkt wie Gewicht, Geld, Länge, nicht aber für Temperatur in Celsius oder in Fahrenheit).

Definition 4.2.14. (Variationskoeffizient): $v = \frac{s}{\bar{x}}$.
(für $x = (x_1, \dots, x_n)$)

Da die Standardabweichung und der Mittelwert dieselben Einheiten aufweisen, kürzen sich diese weg. Der Variationskoeffizient weist deshalb keine Einheiten auf - er ist „dimensionslos“. Er eignet sich deshalb, um das Streuungsverhalten von Daten mit unterschiedlichen Einheiten zu vergleichen (z.B. die Streuung von Aktienperformances in verschiedenen Währungen). Der Variationskoeffizient hat zudem die günstige Eigenschaft, auf proportionale Umrechnungen der Daten (= Multiplikation aller Daten mit einer Konstanten) nicht zu reagieren.

Beweis. Gegeben sei der Datensatz $(x_1, \dots, x_n)$ und der proportional transformierte Datensatz $\alpha \cdot (x_1, \dots, x_n) = (\alpha x_1, \dots, \alpha x_n) =: (x'_1, \dots, x'_n)$. (= bedeutet: der Ausdruck rechts wird durch den Ausdruck links definiert. := bedeutet: der Ausdruck links wird durch den Ausdruck rechts definiert).

Dann gilt $\bar{x}' = \alpha \bar{x}$, da $\alpha \bar{x} = \alpha \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{i=1}^n \alpha x_i = \frac{1}{n} \sum_{i=1}^n x'_i = \bar{x}'$.

Zudem gilt: $s' = \alpha s$, da $\alpha s = \alpha \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n \alpha^2 (x_i - \bar{x})^2} =$

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\alpha(x_i - \bar{x}))^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\alpha x_i - \alpha \bar{x})^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x'_i - \bar{x}')^2} = s'$$

Damit gilt dann: $\frac{s'}{\bar{x}'} = \frac{\alpha s}{\alpha \bar{x}} = \frac{s}{\bar{x}}$ □

Da der Variationskoeffizient gegenüber Währungsumrechnungen unempfindlich ist, können wir z.B. auch Einkommensverteilungen verschiedener Länder mit verschiedenen Währungen vergleichen, oder die Einkommensverteilung eines Landes über die Zeit hinweg, wenn sich eine allgemeine Anhebung des Preisniveaus ergeben hat (Inflation). Allerdings ist der Variationskoeffizient nicht robust. Allenfalls wird ein Wert mit gestutzten Daten berechnet - man lässt Ausreisser weg.

Beispiel 4.2.15. $x_i = 5, 6, 7, 8$ $\bar{x} = \frac{5+6+7+8}{4} = 6.5$;

$$s = \sqrt{\frac{(5-6.5)^2 + (6-6.5)^2 + (7-6.5)^2 + (8-6.5)^2}{3}} = 1.2910$$

$$v = \frac{1.291}{6.5} = 0.19862$$

Transformierte Daten: $x'_i = 5 \cdot 2 = 10$; $6 \cdot 2 = 12$; $7 \cdot 2 = 14$; $8 \cdot 2 = 16$

$$\bar{x}' = \frac{10+12+14+16}{4} = 13 = 2 \cdot 6.5$$

$$s' = \sqrt{\frac{(10-13)^2 + (12-13)^2 + (14-13)^2 + (16-13)^2}{3}} = 2.5820 = 2 \cdot 1.2910$$

$$v' = \frac{2.582}{13} = \frac{2 \cdot 1.291}{2 \cdot 6.5} = \frac{1.291}{6.5} = 0.19862$$
 ◇

Nachdem wir das arithmetische Mittel und die Standardabweichung eingeführt haben, können wir eine Standardisierung eines Datensatzes berechnen, so dass der Mittelwert des standardisierten Datensatzes 0 und die Standardabweichung 1 wird. Durch diese Operation geht Information verloren. Die Standardisierung ist ein Konzept, dass für die Theoriebildung von Bedeutung ist.

Definition 4.2.16. Wir nennen den Datensatz $z := (z_1, \dots, z_n)$ den standardisierten Datensatz von $x := (x_1, \dots, x_n)$ genau dann, wenn $z_i = \frac{x_i - \bar{x}}{s}$, wobei s die Standardabweichung von x ist. Wir nennen z_i ein standardisiertes Datum.

Beispiel 4.2.17. Gegeben sei der Datensatz $(4, 5, 6, 7)$

Dessen Mittelwert ist:

$$\bar{x} = \frac{4 + 5 + 6 + 7}{4} = 5.5$$

Dessen Standardabweichung ist:

$$s = \sqrt{\frac{(4-5.5)^2 + (5-5.5)^2 + (6-5.5)^2 + (7-5.5)^2}{4-1}} = 1.291$$

Wir transformieren die Werte:

$$\begin{aligned} z_1 &= \frac{4-5.5}{1.291} = -1.1619; & z_2 &= \frac{5-5.5}{1.291} = -.3873; \\ z_3 &= \frac{6-5.5}{1.291} = .3873; & z_4 &= \frac{7-5.5}{1.291} = 1.1619 \end{aligned}$$

und erhalten den standardisierten Datensatz $(-1.1619, -0.3873, 0.3873, 1.1619)$ ◇

Übung 4.2.18. Wählen Sie einen Datensatz und darin eine metrisch und stetig skalierte Variable. Berechnen Sie die Standardisierung der Daten.

Satz 4.2.19. Sei $\bar{z}$ der Mittelwert und s_z die Standardabweichung von standardisierten Daten. Dann gilt:

$$\begin{aligned} \bar{z} &= 0 \\ s_z &= 1 \end{aligned}$$

Beweis. Für den Mittelwert $\bar{z}$ gilt mit $z_i := \frac{x_i - \bar{x}}{s}$:

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i = \frac{1}{n} \sum_{i=1}^n \frac{x_i - \bar{x}}{s} = \frac{\frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{n} \sum_{i=1}^n \bar{x}}{s} =$$

$$\frac{\bar{x} - \frac{1}{n} n \bar{x}}{s} = \frac{\bar{x} - \bar{x}}{s} = \frac{0}{s} = 0$$

Für die Varianz: $s_z^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - 0)^2 =$

$$\frac{1}{n-1} \sum_{i=1}^n z_i^2 = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^2 = \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{s^2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}{s^2} = \frac{s^2}{s^2} = 1$$

Somit gilt für die Standardabweichung: $s_z = \sqrt{s_z^2} = \sqrt{1} = 1$ □

Übung 4.2.20. Bestätigen Sie die Aussage des Satzes 4.2.19, indem Sie den Mittelwert und die Standardabweichung für die standardisierten Daten des Beispiels 4.2.17 berechnen.

4.2.4 Lorenzkurve und Lorenzkonzentration

Eine weitere Möglichkeit, Streuung bei metrisch skalierten Daten darzustellen, bietet die Lorenzkurve (= Konzentrationskurve). Sie stellt graphisch dar, wo sich Werte massieren oder konzentrieren. Wir ordnen die Daten der Grösse nach. Für das i -te Datum wird auf der x-Achse $\frac{i}{n}$ aufgetragen (= relativer Anteil der Daten, die kleiner-gleich i sind, an der Gesamtpopulation). Dann berechnen wir die relativen, kumulativen Anteile

$$a\left(\frac{i}{n}\right) := \frac{\sum_{x_j \leq x_i} x_j}{\sum_{i=1}^n x_i}$$

an der betrachteten Gesamtgrösse. Diese tragen wir auf der y-Achse ab und verbinden die nebeneinanderliegenden Punkte

$$\left(\frac{i}{n}, a\left(\frac{i}{n}\right)\right) \text{ und } \left(\frac{i+1}{n}, a\left(\frac{i+1}{n}\right)\right)$$

(für $0 \leq i \leq n-1$) jeweils mit einer Strecke.

Beispiel 4.2.21. Betrachten wir ein Beispiel. 10 Personen hätten folgende Monatseinkommen (in Tausend) (s. Tabelle 4.2.18):

Person	Einkommen
1	8
2	4
3	3
4	6
5	5
6	7
7	1
8	2
9	3.5
10	4.5

Tabelle 4.2.18: Tabelle der Urdaten Einkommen (metrisch skaliert)

Wir berechnen die geordneten, kumulativen relativen Anteile am Gesamtvermögen sowie an der Anzahl Personen (s. Tabelle 4.2.19):

Personen	Einkommen	rel. Anteil am Gesamteinkommen	rel. kum. Anteile am Gesamteinkommen	rel. kum Anteile an Population
		0	0	0
7	1	0.022727273	0.022727273	0.1
8	2	0.045454545	0.068181818	0.2
3	3	0.068181818	0.136363636	0.3
9	3.5	0.079545455	0.215909091	0.4
2	4	0.090909091	0.306818182	0.5
10	4.5	0.102272727	0.409090909	0.6
5	5	0.113636364	0.522727273	0.7
4	6	0.136363636	0.659090909	0.8
6	7	0.159090909	0.818181818	0.9
1	8	0.181818182	1	1
Summe	44			Summe

Tabelle 4.2.19: Berechnungstabelle Lorenzkonzentration (Daten Tabelle 4.2.18)

Tragen wir die Punkte der letzten beiden Spalten in ein Koordinatensystem ein, erhalten wir, wenn wir die Punkte durch Strecken verbinden (s. Abbildung 4.2.10): $\diamond$

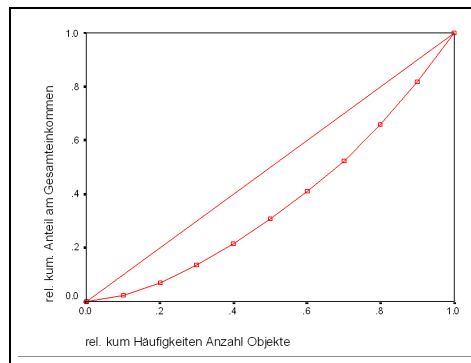


Abbildung 4.2.10: Lorenzkurve für die Daten der Tabelle 4.2.18

Die Kurve von *Abbildung 4.2.10* nennen wir „Lorenz- oder Konzentrationskurve“. Je mehr die Kurve von der $y = x$ -Geraden abweicht, desto ungleicher ist das Einkommen verteilt. Als Kennwert der Ungleichheit könnte man die Fläche zwischen der $x = y$ -Geraden und der Lorenzkurve betrachten. Nützlicher ist es jedoch, das Doppelte dieser Fläche als Kennwert zu verwenden, da sich dieser Wert bei maximaler Abweichung 1 nähert. Fällt die Lorenzkurve mit der $x = y$ -Geraden zusammen, so wird der Kennwert 0. Wir nennen diesen Kennwert „Lorenzkonzentration“.

Die Berechnung der Fläche zwischen der $x = y$ -Geraden und der Lorenzkurve erfolgt wie folgt: zwischen zwei Punkten x_1 und x_2 auf der x -Achse ist die Fläche zwischen der Lorenzkurve und der $x = y$ -Geraden ein Trapez. Dieses ist die Differenz zweier Trapeze, die durch die x -Achse und die Punkte $(x_1, f(x_1))$ und $(x_2, f(x_2))$ einerseits sowie $(x_1, a(x_1))$ und $(x_2, a(x_2))$ andererseits gegeben ist - mit sehr günstigen Eigenschaften (s. *Abbildung 4.2.11*).

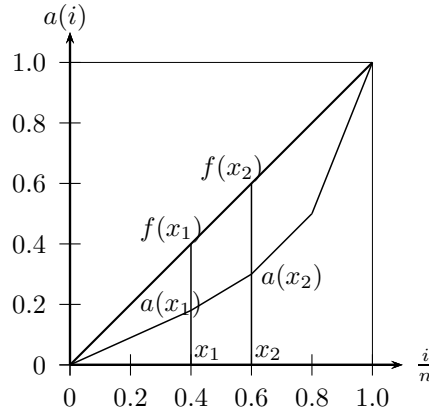


Abbildung 4.2.11: Berechnung der Fläche zwischen der x-y-Achse und der Lorenzkurve

Bemerkung 4.2.22. Obwohl dies wohl vorausgesetzt werden könnte, zeigen wir, wie die Fläche eines Trapezes zwischen der x-Achse und den durch zwei Punkten $(x_1, f(x_1))$ und $(x_2, f(x_2))$ gegebenen oberen Eckpunkte allgemein berechnet werden kann - wobei f eine beliebige reelle Funktion ist. Wir können das Trapez in zwei Flächen aufteilen (s. Abbildung 4.2.12):

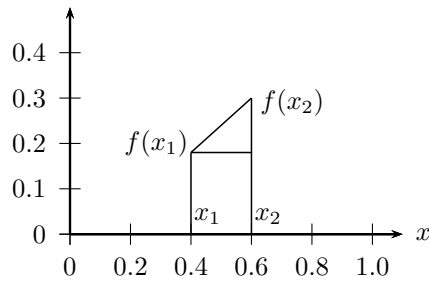


Abbildung 4.2.12: Berechnung eines Trapezes mit Ecken auf der x-Achse

Die Fläche setzt sich aus einem Rechteck

$$f(x_1) \cdot (x_2 - x_1)$$

und einem Dreieck (Höhe mal Basis geteilt durch 2)

$$\frac{(f(x_2) - f(x_1))}{2} \cdot (x_2 - x_1)$$

zusammen. Dies ergibt als Gesamtfläche des Trapezes :

$$\begin{aligned} & f(x_1)(x_2 - x_1) + \frac{(f(x_2) - f(x_1))}{2}(x_2 - x_1) \\ = & (x_2 - x_1) \left(\frac{2f(x_1) + f(x_2) - f(x_1)}{2} \right) \\ = & (x_2 - x_1) \frac{f(x_1) + f(x_2)}{2}. \end{aligned}$$

◇

Wir wenden die Berechnungsformel für Trapeze auf die beiden Trapeze, die durch die Eckpunkte $(x_1, 0), (x_1, x_1), (x_2, x_2), (x_2, 0)$ und durch $(x_1, 0), (x_1, a(x_1)), (x_2, a(x_2)), (x_2, 0)$ gegeben sind (beachten Sie, dass die oberen Eckpunkte des grösseren Trapezes auf der $y = x$ -Achse liegen; siehe

Graphik 4.2.11). Die Fläche des grösseren Trapezes ist:

$$(x_2 - x_1) \frac{x_1 + x_2}{2}$$

Fläche des kleineren Trapezes ist:

$$(x_2 - x_1) \frac{a(x_1) + a(x_2)}{2}$$

Die gesuchte Fläche im Intervall x_2 und x_1 zwischen der $x = y$ -Achse und der Lorenzkurve ist die Differenz:

$$(x_2 - x_1) \frac{x_1 + x_2}{2} - (x_2 - x_1) \frac{a(x_1) + a(x_2)}{2} = (x_2 - x_1) \frac{x_1 + x_2 - a(x_1) - a(x_2)}{2}$$

Sei $\frac{i}{n}$ der kumulierte relative Anteil an der Grundgesamtheit und $a(\frac{i}{n})$ der kumulierte relative Anteil an der untersuchten Grösse. Dann erhalten wir eingesetzt in diese Formel:

$$\left(\frac{i+1}{n} - \frac{i}{n} \right) \frac{\frac{i}{n} + \frac{i+1}{n} - a(\frac{i}{n}) - a(\frac{i+1}{n})}{2}. \quad (4.2.1)$$

Diese Formel gilt auch für die Spezialfälle der ersten und der letzten Fläche, bei denen es sich um zu Dreiecken degenerierte Trapeze handelt: es gilt nämlich für die erste Fläche

$$\begin{aligned} \left(\frac{0+1}{n} - 0 \right) \frac{\frac{0}{n} + \frac{0+1}{n} - a(\frac{0}{n}) - a(\frac{0+1}{n})}{2} &= \frac{1}{n} \frac{\frac{1}{n} - a(\frac{1}{n})}{2} \\ &= \frac{1}{n} \frac{1}{2} - \frac{1}{n} \frac{a(\frac{1}{n})}{2}. \end{aligned}$$

Dies ist die Differenz des grossen Dreiecks vom kleineren Dreieck und ergibt das gewünschte Dreieck, wie Sie sich überlegen können. Das rechte Dreieck ist die Differenz von zwei Trapezen, womit sich zu den anderen Fällen nichts neues ergibt. Entsprechend müssen wir diese Randfälle nicht gesondert betrachten.

Wir vereinfachen:

$$\begin{aligned} \frac{i+1}{n} - \frac{i}{n} &= \frac{1}{n} \\ \frac{i}{n} + \frac{i+1}{n} &= \frac{2i+1}{n} \end{aligned}$$

und summieren über $0 \leq i \leq n-1$ und klammern $\frac{1}{n}$ aus (Begründung der Summationsgrenzen: das erste Trapez, das zum Dreieck degeneriert ist, hat die linke Ecke bei 0. Rechts dürfen wir bei $\frac{i+1}{n}$ nicht über $\frac{n}{n}$ hinausfahren. Dies erreichen wir mit der Grenze $n-1$, da $\frac{n-1+1}{n} = \frac{n}{n} = 1$):

$$\frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{\frac{2i+1}{n} - a(\frac{i}{n}) - a(\frac{i+1}{n})}{2} \right)$$

Die Lorenzkonzentration ist das Doppelte.

Definition 4.2.23. Lorenzkonzentration (nicht klassierte Daten): $\frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{2i+1}{n} - a(\frac{i}{n}) - a(\frac{i+1}{n}) \right)$

mit $\frac{i}{n}$ der kumulierte relative Anteil an der Grundgesamtheit der Untersuchungsobjekte und mit $a(\frac{i}{n})$ der kumulierte relative Anteil an der untersuchten Grösse

Für unser Beispiel 4.2.21 ist die Lorenzkonzentration damit (s. für Berechnung die Tabelle 4.2.20):

Personen	Einkommen	rel. Anteil am Gesamteinkommen	rel. kum. Anteile am Gesamteinkommen	rel. kum. Anteile an Population	Berechnung laut Formel
		0	0	0	
7	1	0.022727273	0.022727273	0.1	0.007727273
8	2	0.045454545	0.068181818	0.2	0.020909091
3	3	0.068181818	0.136363636	0.3	0.029545455
9	3.5	0.079545455	0.215909091	0.4	0.034772727
2	4	0.090909091	0.306818182	0.5	0.037727273
10	4.5	0.102272727	0.409090909	0.6	0.038409091
5	5	0.113636364	0.522727273	0.7	0.036818182
4	6	0.136363636	0.659090909	0.8	0.031818182
6	7	0.159090909	0.818181818	0.9	0.022727273
1	8	0.181818182	1	1	0.008181818
Summe	44			Summe	0.268181818

Tabelle 4.2.20: Berechnung der Lorenzkonzentration mit einer Tabellenkalkulation

Damit umfasst die Lorenzkonzentration im Beispiel 4.2.21 in etwa ein Viertel der Fläche des Quadrates $[0, 1] \times [0, 1]$.

Ohne Beweis halten wir fest: Die Lorenzkonzentration nimmt einen Wert zwischen 0 und $\frac{n-1}{n}$ an (n Anzahl Daten oder bei klassierten Daten Anzahl Klassen). Um einen Wert zwischen 0 und 1 zu erhalten, kann man die *normierte Lorenzkonzentration* definieren:

$$\begin{aligned} & \frac{n}{n-1} \frac{1}{n} \sum_{i=0}^{n-1} \left[\left(\frac{1+2i}{n} - a\left(\frac{i}{n}\right) - a\left(\frac{i+1}{n}\right) \right) \right] \\ &= \frac{1}{n-1} \sum_{i=0}^{n-1} \left[\left(\frac{1+2i}{n} - a\left(\frac{i}{n}\right) - a\left(\frac{i+1}{n}\right) \right) \right] \end{aligned}$$

Definition 4.2.24. Normierte Lorenzkonzentration: $\frac{1}{n-1} \sum_{i=0}^{n-1} \left[\left(\frac{1+2i}{n} - a\left(\frac{i}{n}\right) - a\left(\frac{i+1}{n}\right) \right) \right]$

mit $\frac{i}{n}$ der kumulierte relative Anteil an der Grundgesamtheit der Untersuchungsobjekte und mit $a\left(\frac{i}{n}\right)$ der kumulierte relative Anteil an der untersuchten Grösse

Dieser Ausdruck nimmt bei maximaler Ungleichverteilung den Wert 1 an.

Am obigen Beispiel 4.2.21 erhalten wir: $\frac{10}{9} \cdot 0.268181818 = 0.29798$

Auch bei klassierten Daten ist es möglich, eine Lorenzkurve zu zeichnen. Wir verbinden die nebeneinanderliegenden Punkte

$$\left(\sum_{x_j \leq x_i} h(x_j), \frac{\sum_{x_j \leq x_i} x_j \cdot k(x_j)}{\sum_{i=1}^m x_i \cdot k(x_i)} \right)$$

und

$$\left(\sum_{x_j \leq x_{i+1}} h(x_j), \frac{\sum_{x_j \leq x_{i+1}} x_j \cdot k(x_j)}{\sum_{i=1}^m x_i \cdot k(x_i)} \right)$$

mit Strecken (m = Anzahl der Klassen; $1 \leq j \leq i \leq m$; $h(x_i)$ relative Häufigkeit des Auftretens von x_i ; $k(x_i)$ absolute Häufigkeit des Auftretens von x_i):

Beispiel 4.2.25. mit klassierten Daten (s. Tabelle 4.2.21).

km	Anzahl Beschäftigte	
[0, 5]	7	
]5, 10]	24	
]10, 15]	35	
]15, 20]	18	
]20, 30]	16	

Tabelle 4.2.21: Häufigkeitsverteilung: Berechnung Lorenzkonzentration (klassierte Daten von *Beispiel 4.2.25*)

Die entsprechenden kumulierten Anteile sind dann:

km	Anzahl Beschäftigte	Kilometer pro Klasse	Anteile km	kumulative Anteile km	relative Anteile Personen	kumulative Anteile Personen
[0, 5]	7	17.5	0.003849175	0.003849175	0.07	0.07
]5, 10]	24	180	0.135742341	0.139591516	0.24	0.31
]10, 15]	35	437.5	0.481146897	0.620738413	0.35	0.66
]15, 20]	18	315	0.178161822	0.798900236	0.18	0.84
]20, 30]	16	400	0.201099764	1	0.16	1
Summen	100	1350	1		1	

Tabelle 4.2.22: Berechnung der Lorenzkonzentration mit Hilfe einer Tabellenkalkulation (klassierte Daten von *4.2.25*)

Dies ergibt die Graphik (s. *Abbildung 4.2.13*):

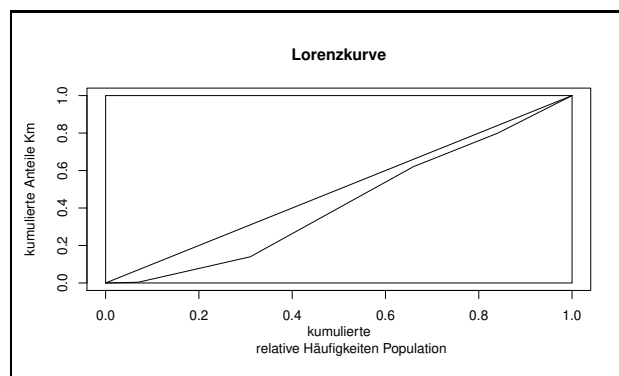


Abbildung 4.2.13: Lorenzkurve bei klassierten Daten (Daten Beispiel *4.2.25*)

◇

Die Berechnung der Lorenzkonzentration für diesen Fall können wir aus

$$\left(\frac{i+1}{n} - \frac{i}{n}\right) \frac{\frac{i}{n} + \frac{i+1}{n} - a\left(\frac{i}{n}\right) - a\left(\frac{i+1}{n}\right)}{2}$$

herleiten (s. (4.2.1)): sei $h_i^+ = \sum_{1 \leq j \leq i} h(x_j)$ und $h(x_j)$ die relativen Häufigkeiten pro Klasse j . Dann entspricht $\frac{i}{n}$ in diesem Zusammenhang der Ausdruck h_i^+ (relativer kumulierter Anteil der Objekte

an der Gesamtanzahl von Objekten) und $a\left(\frac{i}{n}\right)$ entspricht

$$a\left(h_i^+\right) := \frac{\sum_{1 \leq j \leq i}^m x_j \cdot k(x_j)}{\sum_{i=1}^m x_i \cdot k(x_i)}$$

(Anteil an der untersuchten Grösse des Anteils h_i^+ , m die Anzahl der Klassen). Es gilt damit

$$\left(\frac{i+1}{n} - \frac{i}{n}\right) = h_{i+1}^+ - h_i^+ = h(x_{i+1})$$

und damit insgesamt:

$$\begin{aligned} & \left(\frac{i+1}{n} - \frac{i}{n}\right) \frac{\frac{i}{n} + \frac{i+1}{n} - a\left(\frac{i}{n}\right) - a\left(\frac{i+1}{n}\right)}{2} \\ &= h(x_{i+1}) \frac{h_i^+ + h_{i+1}^+ - a(h_i^+) - a(h_{i+1}^+)}{2} \end{aligned}$$

Wir summieren und multiplizieren mit 2 um die nichtnormierte Lorenzkonzentration bei Klassenbildung zu erhalten:

$$\sum_{i=0}^{m-1} h(x_{i+1}) (h_i^+ + h_{i+1}^+ - a(h_i^+) - a(h_{i+1}^+))$$

($h(x_i)$ = relative Häufigkeit der Klasse i ; $h_i^+ = \sum_{1 \leq j \leq i} h(x_j)$; $a(h_i^+) =$ kumulierter Anteil an untersuchter Grösse bezüglich h_i^+ ; $m =$ Anzahl Klassen)

Fürs Beispiel erhalten wir (s. Tabelle 4.2.23):

Anteile Anteile Kilometer	kumulative kum. Anteile Kilometer	relative Anteile Personen	kumulative Anteile Personen	Flächen laut Formel
0.003849175	0.003849175	0.07	0.07	0.004630558
0.135742341	0.139591516	0.24	0.31	0.056774234
0.481146897	0.620738413	0.35	0.66	0.073384525
0.178161822	0.798900236	0.18	0.84	0.014465043
0.201099764	1	0.16	1	0.006575962
Summen		1		0.155830322

Tabelle 4.2.23: Berechnung der Lorenzkonzentration mit Hilfe einer Tabellenkalkulation des Beispiels 4.2.25.

Somit macht die Lorenzkonzentration rund einen Sechstel (0.1558303) der Fläche des Quadrates aus. Die normierte Lorenzkonzentration bei Klassenbildung ist dann:

$$\frac{m}{m-1} \sum_{i=0}^{m-1} h(x_{i+1}) (h_i^+ + h_{i+1}^+ - a(h_i^+) - a(h_{i+1}^+))$$

wobei m die Anzahl der Klassen ist. Wir erhalten damit fürs Beispiel: $m = 5$:

$$\frac{5}{5-1} 0.1558303 = 0.19479$$

4.2.5 Quantile und Quartilsabstand

Bisher haben wir nur Streumasse kennengelernt, die für metrisch skalierte Variablen geeignet sind. Allerdings sind die eingeführten Kennwert nicht robust. Bei Ausreissern sollten sie deshalb nicht verwendet werden. Deshalb führen wir noch robuste Kennwerte ein, die zudem teilweise auch für ordinalskalierte Datensätze Anwendung finden. Wir führen die sogenannten Quantile ein, die in der Literatur unterschiedlich definiert werden, wobei für die Praxis diese Unterschiede kaum eine Bedeutung haben. Wir führen die von Excel verwendete Definition ein, die auch im Statistikprogramm **R** die Standardeinstellung ist. Circa $i \cdot 100\%$ der Daten befinden sich unterhalb des i -Quantils, und circa $100\% - (i \cdot 100\%)$ der Daten befinden sich oberhalb des i -Quantils.

Definition 4.2.26. Sei x ein geordneter Datenvektor. Dann definieren wir:

$$i\text{-Quantil}(x) := \begin{cases} x_j & \text{für } j = (n-1)i + 1 \\ x_j + r(x_{j+1} - x_j) & \text{für } j < (n-1)i + 1 < j + 1 \end{cases}$$

mit:

$x_j = j$ -tes Datum des geordneten Datensatzes x .

$r = (n-1)i + 1 - j$ (= Zahl hinter dem Komma der Zahl $(n-1)i + 1$).

Die Definition ist wie folgt zu verstehen: Ist $(n-1)i + 1$ nicht eine ganze Zahl, so wird die nächstkleinere ganze Zahl j gewählt. Zum Datum x_j wird proportional zu r ein Teil der folgenden Intervalllänge $x_{j+1} - x_j$ dazugerechnet.

Quantile sind robuste Kennwerte. Quantile sind mit dem Median verwandt. Der Median ist das 0.5-Quantil (oder 50%-Quantil), denn $(n-1)0.5 + 1 = (n+1)0.5$.

Beispiel 4.2.27. Gegeben seien die folgenden Daten:

2, 3, 3.2, 5, 5.2, 5.8, 6, 7.8, 9, 10, 11 $\Rightarrow n = 11$

Die Lage des 0.5-Quantils (= Median) der Daten ist: $(11+1) \cdot 0.5 = 6^{\text{te}}$ Datum. Somit ist der Median: 5.8

Die Lage des 0.25-Quantils der Daten ist: $(11-1) \cdot 0.25 + 1 = 3.5$. Somit ist das 0.25-Quantil: $3.2 + (5 - 3.2)0.5 = 4.1$

Die Lage des 0.27-Quantils der Daten ist: $(11-1) \cdot 0.27 + 1 = 3.7$.

Somit ist das 0.27-Quantil: $3.2 + (5 - 3.2)0.7 = 4.46$

Die Lage des 0.59-Quantils der Daten ist: $(11-1) \cdot 0.59 + 1 = 6.9$

Damit ist das 0.59-Quantil: $5.8 + (6 - 5.8)0.9 = 5.98$ ◇

Spezifische Quantile haben einen eigenen Namen. Den Median haben wir bereits kennengelernt. Das 0.25-Quantil, das 0.5-Quantil und das 0.75-Quantil werden „Quartile“ genannt, da sie die Datenmenge in vier anzahlmässig gleichgrosse Datenmengen aufteilen. Das 0.25-Quantil wird „erstes Quartil“ genannt. Das 0.75-Quantil wird „drittes Quartil“ genannt. Entsprechend ist dann der Median das zweite Quartil. Perzentile sind Quantile, die in Hundertsteln angegeben werden - eventuell als Prozente ausgedrückt. Gebräuchlich ist noch der Ausdruck „Dezentile“ für 10er Perzentile.

Für metrisch skalierte Variablen ist der Quartilsabstand ein sinnvoller und robuster Kennwert der Streuung:

Definition 4.2.28. Der Quartilsabstand = 0.75-Quantil - 0.25-Quantil
(= 3. Quartil - 1. Quartil). (für $x = (x_1, \dots, x_n)$)

Im obigen *Beispiel 4.2.27*: Die Lage des 0.75-Quantils: $(11-1) \cdot 0.75 + 1 = 8.5$. Das 3. Quartil ist damit $7.8 + (9 - 7.8)0.5 = 8.4$.

Wir zählen das 1. vom 3. Quartil ab und erhalten: $8.4 - 4.1 = 4.3$.

Da bei ordinalskalierten Daten der Abstandsbegriff keinen Sinn macht, ist der Quartilsabstand für solche Daten sinnlos. Wir können jedoch die Quartile und den Median angeben und haben damit ein recht gutes Bild von der Verteilung der Daten gewonnen.

Beispiel 4.2.29. Qualitäts- und Leistungskontrolle: In der Fertigungsabteilung 12 sind 10 Arbeiter damit beschäftigt, Metallteile zu formen, was besondere Geschicklichkeit voraussetzt. Der

anfallende Ausschuss ist daher beträchtlich. Um sich eine gewisse Übersicht zu verschaffen und die Leistungen bewerten zu können, wurden die fehlerhaften Stücke während rund eines halben Jahres in einer Kontrolle erfasst:

Arbeiter	Ausschuss
1	1
2	5
3	2
4	9
5	3
6	2
7	3
8	6
9	8
10	6

Tabelle 4.2.24: Tabelle der Urdaten einer metrisch skalierten Variable

Zu bestimmen sind die Quartile und der Quartilsabstand. Wir ordnen die Daten (s. Tabelle 4.2.25):

Arbeiter	Ausschuss
1	1
3	2
6	2
5	3
7	3
2	5
8	6
10	6
9	8
4	9

Tabelle 4.2.25: Geordnete Daten der 4.2.24

Lage des 1. Quartils: $(10 - 1) \cdot 0.25 + 1 = 3.25$
 $\Rightarrow$ 1. Quartil = $2 + (3 - 2) \cdot 0.25 = 2.25$
Median: Lage: $(10 + 1) \cdot 0.5 = 5.5 \Rightarrow$ Median: $3 + (5 - 3) \cdot 0.5 = 4$
3. Quartil: Lage: $(10 - 1) \cdot 0.75 + 1 = 7.75 \Rightarrow$ Quartil: $6 + (6 - 6) \cdot 0.75 = 6$.
Quartilsabstand: $6 - 2.25 = 3.75$

◇

4.2.6 Streuungskennwerte für nominalskalierte Daten

Definition 4.2.30. (Simpsonindex): $D = \frac{m}{m-1} \left(1 - \sum_{i=1}^m (h(x_i))^2 \right)$ wobei $h(x_i)$ die relativen Häufigkeiten des Skalenwertes x_i und m die Anzahl Ausprägungen sind.

D gibt an, ob die Daten gleichmässig über die Kategorien verteilt sind oder ob sie sich in wenigen oder gar nur einer Kategorie konzentrieren.

Satz 4.2.31. $0 \leq D \leq 1$ (D definiert wie unter Definition 4.2.30)

Beweis. Fallen alle Werte in eine Kategorie j , so ist $h(x_j) = 1$ und $h(x_{i \neq j}) = 0$. Damit gilt dann:

$$\sum_{i=1}^m (h(x_i))^2 = 0^2 + \dots + 0^2 + 1^2 + 0^2 + \dots + 0^2 = 1. \text{ In diesem Fall gilt } D = \frac{m}{m-1} (1 - 1) = 0$$

Bei gleichmässiger Verteilung gilt: $h(x_i) = h(x_j) = \frac{1}{m}$ für alle i und j .

Damit gilt dann: $\sum_{i=1}^m (h(x_i))^2 = m (h(x_i))^2 = m \left(\frac{1}{m}\right)^2 = \frac{1}{m}$ und:

$$D = \frac{m}{m-1} \left(1 - \sum_{i=1}^m (h(x_i))^2\right) = \frac{m}{m-1} \left(1 - \frac{1}{m}\right) = \frac{m}{m-1} \frac{m-1}{m} = 1.$$

□

Bei maximaler Streuung ist damit der Wert 1 und bei minimaler Streuung ergibt sich 0.

Bemerkung 4.2.32. $\frac{m}{m-1}$ in D ist ein Normierungsfaktor, der garantiert, dass $0 \leq D \leq 1$. Für

$$1 - \sum_{i=1}^m (h(x_i))^2$$

gilt nämlich:

$$0 \leq 1 - \sum_{i=1}^m (h(x_i))^2 \leq \frac{m-1}{m},$$

wie Sie selber nachrechnen können (mit den obigen Rechengängen). Multiplizieren wir diesen Ausdruck mit $\frac{m}{m-1}$, erhalten wir

$$0 \leq \frac{m}{m-1} \left(1 - \sum_{i=1}^m (h(x_i))^2\right) \leq 1.$$

Wir werden in der Folge

$$1 - \sum_{i=1}^m (h(x_i))^2$$

„nichtnormierten Simpsonindex“ nennen. ◇

Beispiel 4.2.33. Für die Häufigkeitsverteilung $\{(1, 9), (2, 8)\}$ erhalten wir den folgenden Simpsonindex:

$$\begin{aligned} 1 - \sum_{i=1}^m (h(x_i))^2 &= 1 - \left(\left(\frac{9}{17}\right)^2 + \left(\frac{8}{17}\right)^2 \right) \\ &= 0.498\,27 \\ D &= \frac{2}{1} \cdot 0.498\,27 = 0.996\,54 \end{aligned}$$

◇

Beispiel 4.2.34. Für die Häufigkeitsverteilung $\{(1, 5), (2, 3), (3, 5)\}$:

$$\begin{aligned} 1 - \sum_{i=1}^m (h(x_i))^2 &= 1 - \left(\left(\frac{5}{13}\right)^2 + \left(\frac{3}{13}\right)^2 + \left(\frac{5}{13}\right)^2 \right) \\ &= 0.650\,89 \\ D &= \frac{3}{2} \cdot 0.650\,89 = 0.976\,34 \end{aligned}$$

◇

Beispiel 4.2.35. Für die Häufigkeitsverteilung $\{(1, 2), (2, 8), (3, 3), (4, 16)\}$:
 $2 + 8 + 3 + 16 = 29$

$$\begin{aligned} 1 - \sum_{i=1}^m (h(x_i))^2 &= 1 - \left(\left(\frac{2}{29} \right)^2 + \left(\frac{8}{29} \right)^2 + \left(\frac{3}{29} \right)^2 + \left(\frac{16}{29} \right)^2 \right) \\ &= 0.60404 \\ D &= \frac{4}{3} \cdot 0.60404 = 0.80539 \end{aligned}$$

◇

4.2.7 Schiefe

Verteilungen brauchen nicht symmetrisch um den Mittelwert herumzuliegen, sondern können links- oder rechtsschief sein, wobei diese Unterscheidungen nur für metrisch skalierte Daten sinnvoll sind.

Beispiel 4.2.36. Beispiele für schiefe Verteilungen (klassierte Daten) (s. Abbildungen 4.2.14 und 4.2.15)

◇

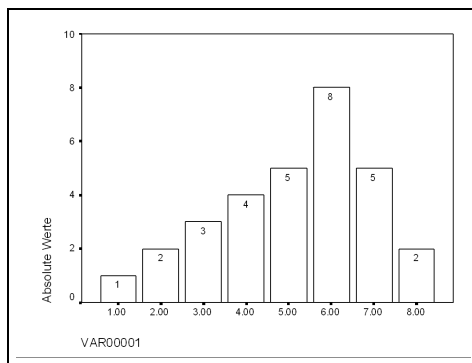


Abbildung 4.2.14: linksschiefe Verteilung (Klassierte Daten)

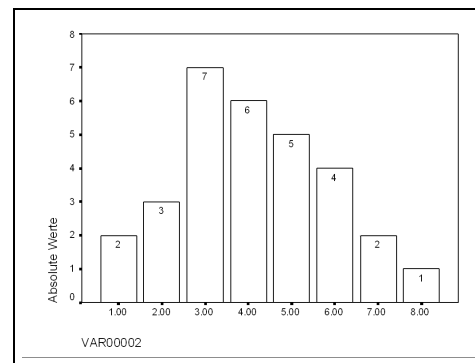


Abbildung 4.2.15: rechtsschiefe Verteilung (Klassierte Daten)

Beispiel 4.2.37. Für nicht-klassierte Daten sieht ein links- und ein rechtsschiefes Streudiagramm wie folgt aus (s. Abbildungen 4.2.16 und 4.2.17):

◇

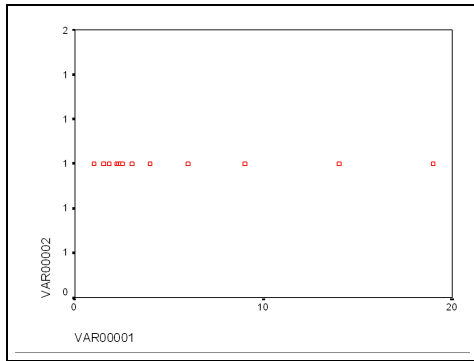


Abbildung 4.2.16: rechtsschiefe Verteilung (nicht klassierte Daten)

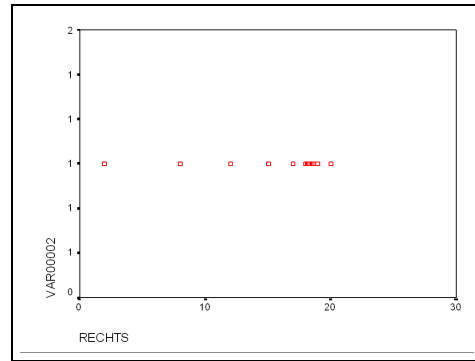


Abbildung 4.2.17: linksschiefe Verteilung

Statt „rechtsschief“ wird auch „linkssteil“ und statt „linksschief“ „rechtssteil“ verwendet.

Wir verwenden den folgende Kennwert nur bei eingipfligen Verteilungen. Für metrischskalierte Daten kann folgender Kennwert für die Schiefe angegeben werden:

Definition 4.2.38. Schiefekennwert: $\frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$
(für $x = (x_1, \dots, x_n)$, $s \neq 0$; für $s = 0$, Schiefe = 0)

$\frac{x_i - \bar{x}}{s}$ stellt die oben behandelte Standardisierung der Daten dar (s. Seite 63). Durch das Potenzieren mit 3 ergeben sich positive und negative Summanden. Der Faktor

$$\frac{n}{(n-1)(n-2)}$$

lässt sich im Rahmen der schliessenden Statistik begründen. Zu beachten ist, dass für sehr grosse n gilt:

$$\frac{n}{(n-1)(n-2)} \approx \frac{1}{n}.$$

Das Schiefemass ist positiv bei rechtsschiefen Verteilungen (da die positiven Abweichungen vom Mittelwert dann gross sind) und negativ bei linksschiefen Verteilungen (da die negativen Abweichungen vom Mittelwert dann gross sind). Bei strikt symmetrischen Daten ist das Schiefemass 0 (Eine symmetrische Verteilung ist in etwa spiegelsymmetrisch zu einer Achse, die durch den Mittelwert gezogen wird - parallel zur y -Achse).

Beispiel 4.2.39. Daten von Beispiel 4.2.37:

1, 1.50, 1.80, 2.20, 2.30, 2.33, 2.50, 3, 4, 6, 9, 14, 19

$$\frac{13}{(13-1)(13-2)} \sum_{i=1}^{13} \left(\frac{x_i - 5.279230769}{5.515019585} \right)^3 = 1.765495623 \text{ (also rechtsschief)}$$

20, 19, 18.5, 18, 17, 15, 12, 8, 2, 18.7, 18.3, 18.2, 18.1

$$\frac{13}{(13-1)(13-2)} \sum_{i=1}^{13} \left(\frac{x_i - 15.6}{5.26782687} \right)^3 = -1.854855905 \text{ (also linksschief)}$$

◇

Bei ordinalskalierten Daten ist ein Schiefemass sinnlos, da die Abstände zwischen den Daten nicht definiert sind und damit nicht sinnvoll interpretierbar sind. Entsprechend kann man nicht von grossen oder kleinen Abweichungen vom Mittelwert sprechen.

4.2.8 Übungen

1. Berechnen Sie mit Excel (ohne die Funktionenbefehle), mit Excel (mit den Funktionenbefehlen) und mit einer Statistiksoftware die jeweils passenden Streuungskennwerte (Spannweite, mittlerer Abweichungsbetrag, Varianz, Standardabweichung, Quartilsabstand, Quartile,

Simpson-Index, Schiefe)

a) Bei 9 verschiedenen Anbietern kostet ein Teelöffel folgenden Betrag:

5 6 3 4.5 4.7 6 2 8 7

b) Bei zehn verschiedenen Anbietern kostet ein Teelöffel folgenden Betrag;

5 6 3 4.5 4.7 5 6 2 8 88

c) Bei einem Rating werden die acht Mitarbeiter wie folgt nach Ihrer Leistung eingeschätzt
3, 3.5, 4, 5.5, 5.5, 4.5, 6, 7

d) Beim Untersuchen von Sportverhalten ergibt sich folgende Häufigkeitsverteilung: $\{(1,3), (2,5), (3,2)\}$ mit 1 für „fährt nur Schlittschuh“, 2 für „fährt nur Ski“, 3 für „fährt sowohl Schlittschuh wie Ski“.

2. Zeichnen Sie für die geeigneten Daten der Übung 4.2.8.1 die Lorenzkurve und berechnen Sie die entsprechende Kennzahl mit R.
3. Berechnen Sie für die Schiefe für die folgenden Daten:
 - a) 1, 1.50, 1.80, 2.20, 2.30, 2.33, 2.50, 3, 4, 6, 9, 14, 19
 - b) 20, 19, 18.5, 18, 17, 15, 12, 8, 2, 18.7, 18.3, 18.2, 18.1
4. Wir haben folgende Daten zur Verfügung: 9.21, 9.52, 9.62, 9.63, 9.61, 9.71, 9.83. Transformieren Sie diese Werte mit der Transformation von Seite 63 und zeigen sie, dass der Mittelwert der standardisierten Werte = 0 und die Varianz der Werte = 1 (um wirklich diese Werte zu erhalten, müssen möglichst viele Stellen berechnet und verwendet werden).

4.2.9 Lösungen

1. a) Die Daten sind metrisch skaliert. Es liegen keine Ausreisser vor. Wir können somit die folgenden Kennwerte verwenden: Spannweite, mittlere Abweichung, Varianz, Standardabweichung, Quartilsabstand, Quartile, Schiefe

Berechnung von Hand:

Spannweite: $\min\{5, 6, 3, 4.5, 4.7, 6, 2, 8, 7\} = 2$; $\max\{5, 6, 3, 4.5, 4.7, 6, 2, 8, 7\} = 8$; Somit ist die Spannweite: $8 - 2 = 6$

Mittlerer Abweichungsbetrag: Mittelwert $\bar{x} = 5.133$. Wir berechnen:

$(|5 - 5.133| + 2 \cdot |6 - 5.133| + |3 - 5.133| + |4.5 - 5.133| + |4.7 - 5.133| + |6 - 5.133| + |2 - 5.133| + |8 - 5.133| + |7 - 5.133|) = 1.437037$.

Varianz: $s^2 = \frac{1}{8}((5 - 5.133)^2 + 2 \cdot (6 - 5.133)^2 + (3 - 5.133)^2 + (4.5 - 5.133)^2 + (4.7 - 5.133)^2 + (2 - 5.133)^2 + (8 - 5.133)^2 + (7 - 5.133)^2) = 3.5225$

Standardabweichung: $s = \sqrt{3.5225} = 1.8768$

Quartile: geordnete Daten: 2, 3, 4.5, 4.7, 5, 6, 6, 7, 8

1. Quartil: Lage: $(9 - 1)0.25 + 1 = 3 \Rightarrow 4.5$

3. Quartil: Lage: $(9 - 1)0.75 + 1 = 7 \Rightarrow 6$

Median: Lage: $(9 - 1)0.5 + 1 = 5 \Rightarrow 5$

Quartilsabstand: $6 - 4.5 = 1.5$

Schiefe: $\frac{n}{(n-1)(n-2)} = \frac{9}{8 \cdot 7} = 0.16071$

Standardabweichung: 1.876832438

$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^3 = \left(\frac{2-5.133333}{1.876832438}\right)^3 + \left(\frac{3-5.133333}{1.876832438}\right)^3 + \left(\frac{4.5-5.133333}{1.876832438}\right)^3 + \left(\frac{4.7-5.133333}{1.876832438}\right)^3 + \left(\frac{5-5.133333}{1.876832438}\right)^3 + 2\left(\frac{6-5.133333}{1.876832438}\right)^3 + \left(\frac{7-5.133333}{1.876832438}\right)^3 + \left(\frac{8-5.133333}{1.876832438}\right)^3 = -1.4287$

$\frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^3 = 0.16071 \cdot -1.4287 = -0.22961$

Für Excel siehe Übungen Streuung Lösungen.xls

b) Bei zehn verschiedenen Anbietern kostet ein Teelöffel folgenden Betrag;

5 6 3 4.5 4.7 5 6 2 8 88

Die Daten sind metrisch skaliert, weisen jedoch einen Ausreisser aus. Entsprechend berechnen wir:

Spannweite, Quartilsabstand, Quartile, eventuell die Schiefe

Spannweite: $88 - 2 = 86$ (Die Spannweite ist zwar nicht ein robustes Mass. Sie bleibt jedoch interessant. Von Nutzen könnte auch eine jeweils eine ohne den äussersten Wert berechnete Spannweite sein. Im Beispiel erhalten wir $8 - 6 = 2$. Die extreme Reduktion der Spannweite durch das Fallenlassen der äussersten Daten zeigt Ausreisser an.

Quartile: geordnete Daten: 2 3 4.5 4.7 5 5 6 6 8 88

Lage der Quartile: $(10 - 1) 0.25 + 1 = 3.25$

$(10 - 1) 0.5 + 1 = 5.5$

$(10 - 1) 0.75 + 1 = 7.75$

1. Quartil: $4.5 + (4.7 - 4.5) 0.25 = 4.55$

2. Quartil = Median: $5 + (5 - 5) 0.5 = 5$

3. Quartil: 6

Quartilsabstand: $6 - 4.55 = 1.45$

Die Schiefe ist nicht robust. Andererseits wollen wir mit ihr ja Asymmetrie zeigen. Entsprechend kann eine Berechnung durchaus Sinn machen. Um zu überprüfen, ob die von 0 verschiedene Schiefe nur durch ein einziges (oder relativ sehr wenige Daten zustande kommt), können wir jeweils die äussersten Werte streichen und die Schiefe nochmals berechnen. Durch dieses Verfahren ergibt sich ein informatives Bild über die Schiefestruktur der Daten.

Obwohl die Standardabweichung sonst für diese Daten kein sinnvolles Mass ist, verwenden wir sie für die Schiefe $s = 26.32628$

$$\begin{aligned} \frac{10}{9 \cdot 8} \sum_{i=1}^{10} \left(\frac{x_i - \bar{x}}{s} \right)^3 &= \left(\frac{2-13.22}{26.32628} \right)^3 + \left(\frac{3-13.22}{26.32628} \right)^3 + \left(\frac{4.5-13.22}{26.32628} \right)^3 + \left(\frac{4.7-13.22}{26.32628} \right)^3 + \\ &2 \left(\frac{5-13.22}{26.32628} \right)^3 + 2 \left(\frac{6-13.22}{26.32628} \right)^3 + \left(\frac{7-13.22}{26.32628} \right)^3 + \left(\frac{88-13.22}{26.32628} \right)^3 \\ &= 3.139238 \end{aligned}$$

Lassen wir die äussersten Werte (min und max) weg, erhalten wir die folgende Schiefe: 0.5317037. Offenbar ist der obige Schiefewert (mindestens) einem Ausreisser zuzuschreiben. Um herauszufinden, welche Extremwerte etwas zur positiven Schiefe im ursprünglichen Datensatz beitragen, kann man die Schiefe für die Datensätze berechnen, die jeweils nur um eines der Extrema gekürzt wurden. Wir erhalten: 0.009916653 (incl. 2, excl. 88) und 2.986095 (incl. 88, excl. 2). Es wird klar, dass die positive Schiefe der ursprünglichen Daten fast nur dem Ausreisser 88 zuzuschreiben ist.

Für Excellösung siehe: Übungen Streuung Lösungen.xls

c) Die Daten sind ordinalskaliert. Es können einfach Quartile berechnet werden (der Quartilsabstand oder andere Streumasse machen keinen Sinn). Die Lage des 1. Quartils des geordneten Datensatzes (3, 3.5, 4, 4.5, 5.5, 5.5, 6, 7) ist $(8 - 1) 0.25 + 1 = 2.75$. Damit erhalten wir das 1. Quartil als $3.5 + (4 - 3.5) 0.75 = 3.875$. Die Lage des 2. Quartils (= Median) ist $(8 - 1) 0.5 + 1 = (8 + 1) 0.5 = 4.5$. Damit erhalten wir den Median $4.5 + (5.5 - 4.5) 0.5 = 5$. Die Lage des 3. Quartils ist: $(8 - 1) 0.75 + 1 = 6.25$. Damit erhalten als 3. Quartil: $5.5 + (6 - 5.5) 0.25 = 5.625$.

d) Es kann der Simpsonindex berechnet werden:

$$\frac{k}{k-1} \left(1 - \sum_{i=1}^k (h(x_i))^2 \right) = \frac{3}{2} \left(1 - \left(\left(\frac{3}{10} \right)^2 + \left(\frac{5}{10} \right)^2 + \left(\frac{2}{10} \right)^2 \right) \right) = \frac{93}{100} = 0.93$$

2. Die Lorenzkurve kann für die ersten beiden Datensätze sinnvoll berechnet werden. Sie drückt aus, wie stark die Angebote streuen (Lorenzkonzentration für Datensatz 1: 0.2180736; für

Datensatz 2: 0.6895277).

3. a) 1.765495623
b) -1.854855905
4. Wir berechnen $\bar{x}$ und s :

$$\bar{x} = \frac{9.21+9.52+9.62+9.63+9.61+9.71+9.83}{7} = 9.59; s = 0.193304595565. \text{ (s. Tabelle 4.2.26)}$$

$$\begin{aligned}
 z_1 &= \frac{9.21-9.59}{0.193304595565} = -1.9658 & z_4 &= \frac{9.63-9.59}{0.193304595565} = .20693 \\
 z_2 &= \frac{9.52-9.59}{0.193304595565} = -.36212 & z_5 &= \frac{9.61-9.59}{0.193304595565} = .10346 \\
 z_3 &= \frac{9.62-9.59}{0.193304595565} = .1552 & z_6 &= \frac{9.71-9.59}{0.193304595565} = .62078 \\
 z_7 &= \frac{9.83-9.59}{0.193304595565} = 1.2416 & \bar{z} &= 0; \bar{s}_z = 1
 \end{aligned}$$

Tabelle 4.2.26: Standardisierung von Daten gemäss Definition von Seite 63

4.3 Zeitreihen und Indizes

4.3.1 Zeitreihen

Oft möchte man die Entwicklung von Grössen über die Zeit hinweg untersuchen. So möchte man wissen, wie sich das Klima entwickelt (Wärme), wie sich die Umsätze einer Firma über die letzten 20 Jahre entwickelt haben, wie sich die Steuerbelastung der Konsumenten verändert hat. Haben wir Daten $y_1, \dots, y_n$ für eine Reihe von Zeitpunkten $t_1, \dots, t_n$, so sprechen wir von „Zeitreihen“. Diese können wir auch als Funktionen von $\mathbb{R}_+$ nach $\mathbb{R}$ auffassen mit dem Graphen $\{(t_j, y_j) : t_j \text{ ist ein Zeitpunkt und } y_j \text{ ist der Messwert bezüglich der Variable } Y \text{ der Grösse } x\}$. Entsprechend können wir Zeitreihen im Koordinatensystem darstellen. Gewöhnlich werden auf der x-Achse gleiche Zeitabstände gewählt und die Werte aufeinanderfolgender Zeitpunkte mit Strecken verbunden. Die Strecken dürfen nicht so gedeutet werden, dass zwischen den Werten der Zeitreihe die Werte der entsprechenden Funktion geliefert werden.

Beispiel 4.3.1. Die Beschäftigungsquote der Schweiz hat sich in den 90er Jahren wie folgt entwickelt (s. Tabelle 4.3.27):

Erwerbs- quoten	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Total	54,0	53,5	53,5	53,0	52,8	53,0	53,3	53,4	53,2	53,0
Frauen	43,2	43,0	43,2	43,0	42,8	43,5	44,0	44,4	44,4	44,6
Männer	65,3	64,5	64,3	63,5	63,3	63,0	63,0	62,8	62,4	61,8
Ausländer	64,9	64,1	62,3	61,1	61,2	60,6	60,3	60,7	58,6	57,4
Schweizer	51,8	51,3	51,6	51,2	50,9	51,2	51,6	51,6	51,9	52,0

Tabelle 4.3.27: Erwerbstätige und Erwerbslose in % der Bevölkerung. Als Erwerbstätige gelten hier Personen mit einer bezahlten Arbeit von mind. 6 Std./Woche. Ab 1991 fliessen in die Erwerbsquoten nicht mehr die Anzahl der Arbeitslosen gemäss seco, sondern die Erwerbslosen nach internationalen Normen ein (Quelle: Bundesamt für Statistik, Homepage: <http://www.bfs.admin.ch/bfs/portal/de/index.html>).

Diese Daten können wir auch graphisch darstellen (s. Abbildung 4.3.18):

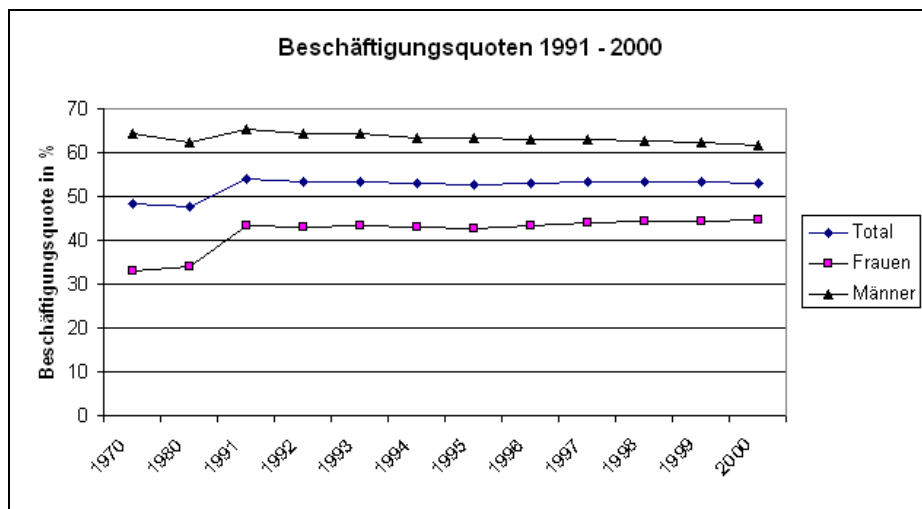


Abbildung 4.3.18: Erwerbstätige und Erwerbslose in % der Bevölkerung (Quelle: Bundesamt für Statistik, Homepage: <http://www.bfs.admin.ch/bfs/portal/de/index.html>)

Beispiel 4.3.2. Die Unternehmung ProFormaggio hat von 1990 bis 2005 folgende Umsätze gemacht (s. Tabelle 4.3.28 und Abbildung 4.3.19):

Jahr	Umsatz in Franken
1990	994754.4
1991	995135.5
1992	996601.0
1993	997262.3
1994	998354.6
1995	998660.7
1996	994497.9
1997	999835.0
1998	996383.9
1999	998987.0
2000	1000410.6
2001	1000746.5
2002	999946.3
2003	999887.6
2004	1000203.4
2005	1001633.8

Tabelle 4.3.28: Umsätze der Unternehmung ProFormaggio von 1990 bis 2005

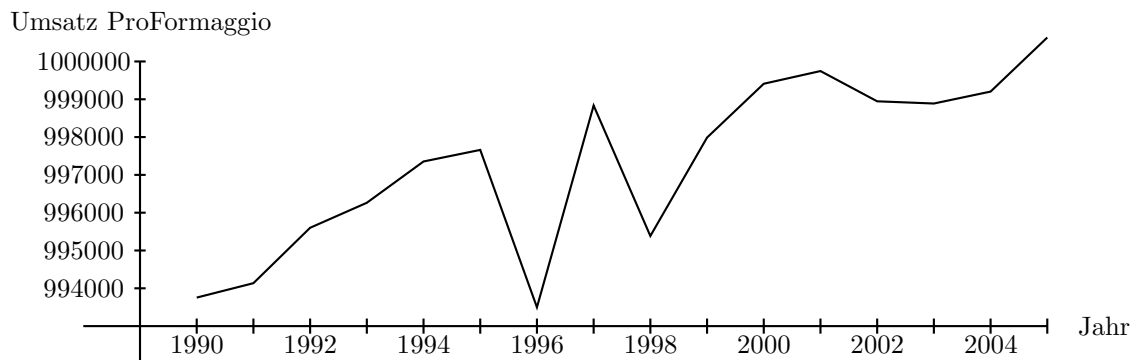


Abbildung 4.3.19: Graphische Darstellung der Daten von Tabelle 4.3.28

◇

Bei der Interpretation der graphischen Darstellungen von Zeitreihen ist Vorsicht geboten. Je nach Skalierung der y-Achse erscheinen Schwankungen als gross oder als klein. Die obigen Daten (Tabelle 4.3.28) könnte man auch wie folgt darstellen (4.3.20):

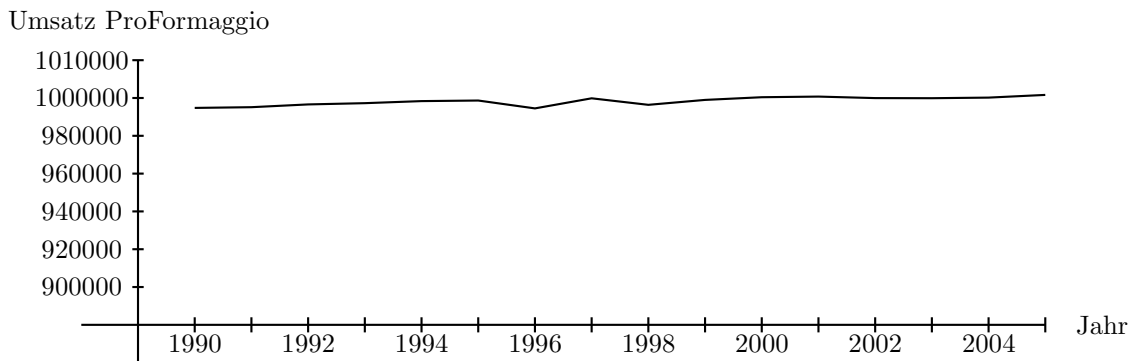


Abbildung 4.3.20: Graphische Darstellung der Daten von Tabelle 4.3.28. Beispiel für die Auswirkungen der Skalierung der y-Achse.

Würden wir die y-Achse von 0 weg angegeben, würde der Umsatz nur mehr als gerader Strich erscheinen. Entsprechend muss man bei der Interpretation von Zeitreihen immer auf die y-Skala achten.

Oft ist die Annahme vertretbar, dass Zeitreihen aus drei Komponenten bestehen:

1. Saisonale Schwankungen: wellenförmige Komponente (z.B. Schwankungen der Arbeitslosigkeit, die sich jedes Jahr saisonal ähnlich wiederholt, mittlere Monatstemperaturen im Wallis).

Beispiel 4.3.3. Für saisonale Schwankungen ohne Trendkomponente (s. Abbildung 4.3.21):

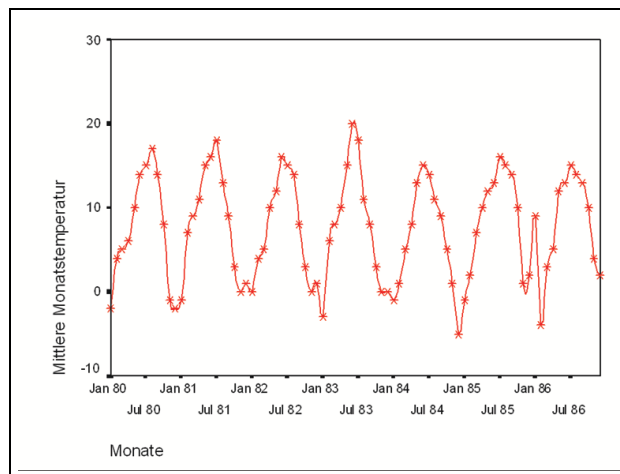


Abbildung 4.3.21: Beispiel für eine saisonale Komponente in einer Zeitreihe (Darstellung mit SPSS: Graphik, Sequenzen)

◇

2. Trend: Diese Komponente drückt die langfristige Entwicklung aus. Im obigen Beispiel stellt sich z.B. die Frage ob in diesen Daten ein langfristiger Trend zu durchschnittlich höheren Jahrestemperaturen festzustellen ist. Bezüglich der Wirtschaft könnte uns z.B. die Frage interessieren, wie der Wachstumstrend unabhängig von konjunkturellen oder saisonalen Einfluss aussieht. Während im obigen Beispiel vermutlich kein Trend herauszulesen ist, ist bei der folgenden Graphik ein deutlicher Trend sichtbar:

Beispiel 4.3.4. Für saisonale Schwankungen mit Trendkomponente (s. Abbildung 4.3.22)

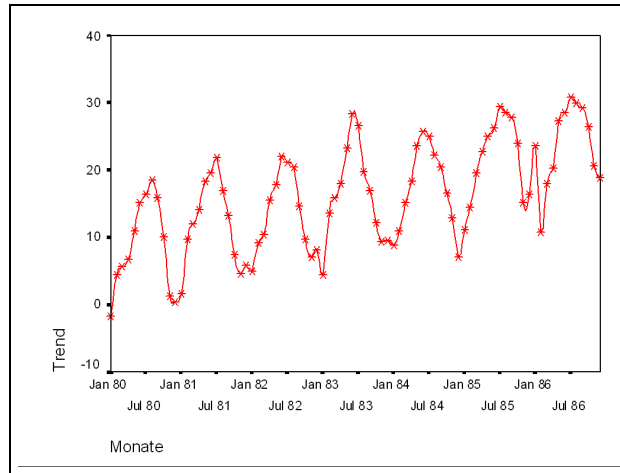


Abbildung 4.3.22: Beispiel für eine Zeitreihe mit saisonaler Komponente und mit einem Trend

◇

3. Zufällige Schwankungen: Es handelt sich um die Abweichungen der Daten vom Trend und den saisonalen Schwankungen.

Bei der Analyse von Zeitreihen geht es darum, aus den Beobachtungswerten die verschiedenen Komponenten herauszufiltrieren, um eventuell Voraussagen machen zu können. Wir verfolgen diese Ideen hier nicht weiter. Die Zeitreihenanalyse ist ein anspruchsvoller Zweig der Statistik, der auch in der Finanzmathematik eine wichtige Rolle spielt (Anlagestrategien).

4.3.2 Indexzahlen

Indexzahlen dienen der Darstellung der zeitlichen Entwicklung einer oder mehrerer Größen. Statt wie vorhin die Daten unmittelbar zu analysieren, rechnen wir diese in Prozentzahlen um.

Bei der Berechnung der Indexzahlen wird aus einer Zahlenreihe eine Basiszahl ausgewählt, und auf diese werden dann die übrigen Zahlen bezogen. Meistens setzt man die Basiszahl = 100.

Man unterscheidet zwischen einfachem Index und zusammengesetztem Index.

Der einfache Index

Mit einem einfachen Index wird die zeitliche Entwicklung einer Grösse dargestellt.

Man unterscheidet zwischen

- Index mit fixer Basis - die Grösse x des Jahres t wird in Prozenten bezüglich der Basis b ausgedrückt. Damit erhalten wir den Index i_t für das Jahr t , indem wir rechnen:

$$i_t = \frac{100x_t}{b}$$

- Index mit beweglicher Basis, d.h. die Zahl des Vorjahres wird jeweils als Basis (=100) für den Index gewählt. Auf diese Weise entsteht ein Kettenindex. Formal: die Grösse x_{t+1} wird in Prozenten bezüglich der Grösse x_t des Vorjahres ausgedrückt. Damit erhalten wir den Index i_{t+1} für das Jahr $t+1$:

$$i_{t+1} = \frac{100x_{t+1}}{x_t}$$

Beispiel 4.3.5. *Bruttosozialprodukt in Perschka (Phantasieland, s. Tabelle 4.3.29)*

Jahr	in Mio. Fr.	Index 1980 = 100	Index 1997 =100	Prozentuale Veränderung gegenüber Vorjahr	Kettenindex
1980	93930	100	35.27622338		
1990	177345	188.8054935	66.60344763		
1991	193975	206.5101671	72.84898787	9.377202628	109.3772026
1992	205170	218.4286171	77.05336688	5.771362289	105.7713623
1993	213950	227.7760034	80.35077177	4.279378077	104.2793781
1994	226060	240.668583	84.89878695	5.660200982	105.660201
1995	241355	256.9519855	90.6429564	6.765902858	106.7659029
1996	254925	271.3989141	95.73928719	5.622423401	105.6224234
1997	266270	283.4770574	100	4.450328528	104.4503285

Tabelle 4.3.29: Entwicklung des Bruttosozialproduktes eines (fiktiven) Landes

◇

Übung 4.3.6. Suchen Sie auf der Home-Page des Bundesamtes für Statistik (<http://www.statistik.admin.ch>) die Zahlen des Bruttosozialproduktes der Schweiz der letzten 10 Jahre und erstellen Sie mit Hilfe von Excel eine Tabelle, die der obigen Tabelle entspricht.

Der zusammengesetzte Index

Beim zusammengesetzten Index werden mehrere Grössen zu einer Indexzahl zusammengefasst. Im allgemeinen verwendet man den gewogenen Summenindex.

Bekannte zusammengesetzte Indices sind: Landesindex der Konsumentenpreise, Aktienindex. Wir unterscheiden folgenden Arten von zusammengesetzten Indices (dabei ist p_{t_i} = der aktuelle Preis des Gutes i ; m_{t_i} die aktuelle Menge des Gutes i ; p_{0_i} der Basispreis des Gutes i und m_{0_i} die Basismenge des Gutes i).

Definition 4.3.7. Wertindex: $W = \frac{\sum_{i=1}^n p_{t_i} m_{t_i}}{\sum_{i=1}^n p_{0_i} m_{0_i}}$

Bezüglich eines einzelnen Gutes beschreibt der Wertindex das Verhältnis des Produktes von aktueller Menge und aktuellem Preis zum Produkt des Basispreises und der Basismenge.

Definition 4.3.8. Preisindex nach Paasche: $P_p = \frac{\sum_{i=1}^n p_{t_i} m_{t_i}}{\sum_{i=1}^n p_{0_i} m_{t_i}}$

Definition 4.3.9. Preisindex nach Laspeyres: $P_L = \frac{\sum_{i=1}^n p_{t_i} m_{0_i}}{\sum_{i=1}^n p_{0_i} m_{0_i}}$

Der Preisindex nach Laspeyres hat den Vorteil, dass seine Berechnung in der Praxis einfacher ist. Für den Preisindex nach Paasche müssen die Mengen und die Preise für jede Periode erhoben werden, während für den Preisindex nach Laspeyres nur die Preise für jede Periode erhoben werden müssen. Bei der Berechnung des Indexes der Konsumentenpreise wären die entsprechenden Erhebungen recht aufwendig (zeitlich und finanziell). Zudem sind beim Laspeyres-Index die Zahlen einer Indexreihe mit konstanter Basis miteinander vergleichbar.

Der Nachteil des Laspeyres-Preisindexes ist demgegenüber, dass die Aufgabenstruktur der Basisperiode veraltet, so dass nach einer bestimmten Zeit der Warenkorb neu festgesetzt werden muss. Der Preisindex nach Paasche passt sich demgegenüber laufend der neuen Verbrauchssituation an. In den meisten Fällen wird der Preisindex nach Laspeyres benutzt und zu Kontrollzwecken

der nach Paasche. Ergeben sich grössere Abweichungen, muss der Warenkorb aktualisiert werden (Indexrevision).

Definition 4.3.10. Mengenindex nach Paasche: $M_p = \frac{\sum_{i=1}^n p_{t_i} m_{t_i}}{\sum_{i=1}^n p_{t_i} m_{0_i}}$

Definition 4.3.11. Mengenindex nach Laspeyres: $M_L = \frac{\sum_{i=1}^n p_{0_i} m_{t_i}}{\sum_{i=1}^n p_{0_i} m_{0_i}}$

Bezüglich der Mengenindices können ähnliche Bemerkungen gemacht werden wie bezüglich der Preisindices.

Oft werden die obigen Indices in Prozenten angegeben. Sie sind dann mit 100 zu multiplizieren.

Beispiel 4.3.12. In der Tabelle 4.3.30 werden der Preis und die Menge von drei Gütern angegeben. Es wird vorgeführt, wie man die verschiedenen zusammengesetzten Indices berechnet.

	Gut A		Gut B		Gut C	
Jahr	Preis (Fr./St.)	Menge (Stück)	Preis (Fr./m)	Menge (m) (m)	Preis (Kr./kg)	Menge (kg)
0	31	18	11	340	8.7	520
1	48	20	15	350	9	530
2	55	30	20	290	9	510

Tabelle 4.3.30: Umgesetzte Mengen und Preise von 3 Gütern

$$\begin{aligned}
 P_{1_p} &= \frac{48 \cdot 20 + 15 \cdot 350 + 9 \cdot 530}{31 \cdot 20 + 11 \cdot 350 + 8.7 \cdot 530} = 1.2091 & P_{1_L} &= \frac{48 \cdot 18 + 15 \cdot 340 + 9 \cdot 520}{31 \cdot 18 + 11 \cdot 340 + 8.7 \cdot 520} = 1.2065 \\
 M_{1_p} &= \frac{48 \cdot 20 + 15 \cdot 350 + 9 \cdot 530}{48 \cdot 18 + 15 \cdot 340 + 9 \cdot 520} = 1.0316 & M_{1_L} &= \frac{31 \cdot 20 + 11 \cdot 350 + 8.7 \cdot 530}{31 \cdot 18 + 11 \cdot 340 + 8.7 \cdot 520} = 1.0294 \\
 P_{2_L} &= \frac{55 \cdot 18 + 20 \cdot 340 + 9 \cdot 520}{31 \cdot 18 + 11 \cdot 340 + 8.7 \cdot 520} = 1.4135 \quad (= 141.35\%) \\
 P_{2_P} &= \frac{55 \cdot 30 + 20 \cdot 290 + 9 \cdot 510}{31 \cdot 30 + 11 \cdot 290 + 8.7 \cdot 510} = 1.407 \quad (\text{in Prozent: } 140.7) \\
 M_{2_L} &= \frac{31 \cdot 30 + 11 \cdot 290 + 8.7 \cdot 510}{31 \cdot 18 + 11 \cdot 340 + 8.7 \cdot 520} = 0.96996 \quad (= 96.996\%) \\
 M_{2_P} &= \frac{55 \cdot 30 + 20 \cdot 290 + 9 \cdot 510}{55 \cdot 18 + 20 \cdot 340 + 9 \cdot 520} = 0.96552 \quad (= 96.552\%) \\
 W_1 &= \frac{48 \cdot 20 + 15 \cdot 350 + 9 \cdot 530}{31 \cdot 18 + 11 \cdot 340 + 8.7 \cdot 520} = 1.2446 \quad (\text{in Prozent: } 124.46) \\
 W_2 &= \frac{55 \cdot 30 + 20 \cdot 290 + 9 \cdot 510}{31 \cdot 18 + 11 \cdot 340 + 8.7 \cdot 520} = 1.3648 \quad (\text{in Prozente: } 136.48)
 \end{aligned}$$

◇

Umbasierung

Zwei Indexreihen können nur dann miteinander verglichen werden, wenn sie das gleiche Basisjahr haben. Ist das nicht der Fall, so wird eine der beiden auf die Basis der anderen Reihe umgerechnet (umbasiert).

Beispiel 4.3.13. Wir wollen den Grosshandels- und Konsumentenpreisindex eines Landes vergleichen. Während der Konsumentenpreisindex auf den Dezember 1992 indexiert ist, ist der Grosshandelspreisindex auf 1973 indexiert. Um die beiden Indices zu vergleichen, müssen sie auf das selbe Datum basiert werden (s. Tabelle 4.3.31):

Grosshandels und Konsumentenpreisindex eines Phantasielandes

Jahr	Konsumentenpreisindex		Grosshandelsindex	
	Dez. 1992 = 100	1990 = 100	1973 = 100	1990 = 100
1990	87.1	100	155.9	100
1991	92.8	106.5442021	165	105.837075
1992	98	112.5143513	169.2	108.5311097
1993	100.9	115.8438576	170	109.0442591
1994	103.9	119.2881745	175.5	112.5721616
1995	107.4	123.3065442	179.5	115.1379089
1996	108.2	124.2250287	172.4	110.5837075
1997	109.8	126.0619977	169	108.4028223
1998	111.8	128.358209	175	112.2514432
1999	113.7	130.5396096	180	115.4586273
2000	115.8	132.9506315	178	114.1757537

Tabelle 4.3.31: Grosshandels und Konsumentenpreisindex eines Phantasielandes

Wir berechnen den umbasierten Index durch $\frac{100 \cdot x_t}{87.1}$ mit $x_t = 92.8, 98, 100.9, 103.9$, etc. für den Konsumentenpreisindex und durch $\frac{100 \cdot x_t}{155.9}$ mit $x_t = 165, 169.2$, etc. für den Grosshandelspreisindex
 ◇

Wird ein Index revidiert (mit neuer Basis), kann es notwendig sein, den alten und den neuen Index zu verketteten. Dies geschieht, indem der eine auf die Basis des anderen umbasiert wird.

Landesindex der Konsumentenpreise

Der Landesindex der Konsumentenpreise hat die Aufgabe, die zeitliche Veränderung der Detailhandelspreise von Waren und Dienstleistungen, die von Haushaltungen von Unselbständigerwerbenden konsumiert werden, zu messen.

Der seit 1922 berechnete Index wurde mehrmals vollständig revidiert. 1922 umfasste er lediglich die Bedarfsgruppen „Nahrungsmittel“, „Brenn- und Leuchtstoffe“ und „Bekleidung“. In den folgenden Revisionen wurden weitere Bedarfsgruppen hinzugefügt. 1966 wurde die Beschränkung auf die Güter für die elementaren Lebensbedürfnisse fallengelassen. Es werden nun alle im Privathaushalt relevanten Grössen berücksichtigt. Um die Vergleichbarkeit von Daten zu gewährleisten, erfolgen neuerdings Anpassungen an den Preisindex, wie er von den EU-Ländern verwendet wird. Seit 2000 wird der Index alle 5 Jahre überprüft und damit dem Revisionsrhythmus der EU angeglichen.

Der Landesindex beruht auf den drei Säulen Warenkorb, Preiserhebung und Berechnung. Der Warenkorb enthält möglichst die Gesamtheit der Güter und Dienste, die tatsächlich konsumiert werden. Der Warenkorb von 1977 umfasste z.B. 290 gewichtete Artikelgruppen (z.B. verpackte Vollmilch) mit über 1000 Erhebungspositionen (z.B. Vollmilch, uperisiert, 1/4 Liter). Die Artikelgruppen werden zu gewichteten Bedarfsgruppen zusammengefasst. Die Gewichtung erfolgt durch eine empirische Erhebung: Hausfrauen und -männer tragen den von ihnen eingekauften Artikel in ihr Haushaltsbüchlein ein, bei Nahrungsmitteln mit Mengenangabe. Die Preiserhebung erfolgt beim Verkäufer. Die Berechnung ist im Prinzip ein Preisindex nach Laspeyres. Die Gewichte werden also nur im Basisjahr erhoben. Um das Veralten des Warenkorbes zu verzögern (Waren können verschwinden, neue Artikel erreichen einen ansehnlichen Marktanteil), werden Artikel, die an Bedeutung verlieren, durch neue ersetzt (unter Beibehaltung der Gruppengewichtung). Die Erhebungsmethoden des Landesindex der Konsumentenpreise sind nicht so neutral wie es scheinen möchte. So setzt sich der Warenkorb von Personen aus anderen Schichten jeweils anders zusammen. Damit können Waren des alltäglichen Gebrauchs teurer werden und Flugpreise billiger. Für Personen mit kleinem Einkommen ergibt sich eine Teuerung, für Personen mit hohem Einkommen kann es jedoch zu einer Reduktion der Gesamtausgaben kommen. Die berechnete, mittlere Teuerung wirkt sich auf Lohnverhandlungen und die Zinspolitik der Nationalbank aus. Entsprechend

ist die Berechnungsmethode des Landesindex der Konsumentenpreise eine politisch „geladene“ Angelegenheit.

Übung 4.3.14. Suchen Sie auf der Home-Page des Bundesamtes für Statistik Angaben zur Berechnung des Landesindex der Konsumentenpreise.

Aktienindices

Aktienindices werden berechnet auf Grund der Börsenkurse und manchmal der Kurse einiger ausgewählter vorbörslicher Titel. Neben dem Gesamtindex werden Hauptindizes (Inhaber-, Namenaktien und Partizipationsscheine), Teilindizes (Dienstleistungen und Industrie) sowie Branchenindices berechnet.

Dabei wird oft das Verfahren von Laspeyres verwendet. Die Kurse der einzelnen Titel werden mit der Anzahl der ausgegebenen Aktien gewichtet (Basiskapitalisierung). Nach Kapitalveränderungen (Kapitalerhöhungen bzw. -herabsetzungen und Aktiensplits) werden die Gewichte neu bestimmt und der neue Index mit dem alten verkettet.

Übung 4.3.15. Wie heisst der heutige Aktienindex der Schweiz? Suchen Sie auf dem Internet Angaben zur Berechnungsweise dieses Index.

4.3.3 Übungen

1. Der Aussenhandel von Perschka entwickle sich wie folgt (s. Tabelle 4.3.32):

	Ausfuhren	Einfuhren
1990	40	30
1991	44	32
1992	43	35
1993	41	32
1994	43	35
1995	45	40
1996	48	38
1997	50	42
1998	49	43
1999	52	45

Tabelle 4.3.32: Aussenhandel von Perschka

Berechnen Sie einen Basisindex, indem sie 1990 auf 100 setzen, indem sie 1995 auf 100 setzen. Berechnen Sie einen Kettenindex.

2. Berechnen Sie den Wertindex, den Preisindex nach Paasche und Laspeyres, den Mengenindex nach Paasche und Laspeyres für die folgenden Daten (s. Tabelle 4.3.33):

	Gut 1		Gut 2		Gut 3	
Jahr	Mengen	Preise	Mengen	Preise	Mengen	Preise
1995	50	20	100	5	3	150000
1996	30	22	110	5.5	4	160000
1997	60	22	105	6	3	160000
1998	55	23	110	6	5	165000
1999	62	23	115	6.5	4	168000
2000	65	24	130	7	4	169000

Tabelle 4.3.33: Beispieldaten für Übung 4.3.3.2

3. Gegeben sei die folgende Indexreihe (s. Tabelle 4.3.34):

Jahr	Indexwert
1980	100
1981	103
1982	105
1983	104
1984	106
1985	107
1986	111
1987	110
1988	112
1989	113
1990	115
1991	114
1992	116
1993	118
1994	119
1995	120

Tabelle 4.3.34:

1990 wird auf 100 gesetzt. Rechnen Sie die alten Indices auf diese Basis um.

4. Vollziehen Sie die Graphiken des Unterkapitels 4.3 mit Excel nach.

4.3.4 Lösungen

1. Man erhält die folgende Tabelle 4.3.35:

	A	E	Index der Ausfuhren 1990 = 100	Index der Einfuhren 1990 = 100	Index der Ausfuhren 1995 = 100	Index der Einfuhren 1995 = 100	Kettenindex
1990	40	30	100	100	88.88888889	75	
1991	44	32	110	106.6666667	97.77777778	80	110
1992	43	35	107.5	116.6666667	95.55555556	87.5	97.72727273
1993	41	32	102.5	106.6666667	91.11111111	80	95.34883721
1994	43	35	107.5	116.6666667	95.55555556	87.5	104.8780488
1995	45	40	112.5	133.3333333	100	100	104.6511628
1996	48	38	120	126.6666667	106.6666667	95	106.6666667
1997	50	42	125	140	111.1111111	105	104.1666667
1998	49	43	122.5	143.3333333	108.8888889	107.5	98
1999	52	45	130	150	115.5555556	112.5	106.122449

Tabelle 4.3.35: Ergebnisse der Berechnungen

2. Man erhält die folgende Tabelle [4.3.36](#):

Jahr	Wertindex	Preis: Paasche	Preis: Laspeyres	Menge: Paasche	Menge: Laspeyres
1995	1	1	1	1	1
1996	1.420299003	1.066730433	1.066777409	1.33139209	1.33145072
1997	1.06744186	1.066910178	1.066888151	1.000518995	1.000498339
1998	1.831506091	1.100146345	1.100221484	1.664670357	1.664784053
1999	1.493186047	1.120233793	1.120265781	1.332885528	1.332923588
2000	1.502702104	1.127120193	1.127131783	1.333208882	1.333222591

Tabelle 4.3.36: Ergebnisse der Berechnungen

3. Man erhält die folgende Tabelle [4.3.37](#):

4. siehe Exceltabelle

Jahr	Indexwert	neuer Indexwert
1980	100	86.95652174
1981	103	89.56521739
1982	105	91.30434783
1983	104	90.43478261
1984	106	92.17391304
1985	107	93.04347826
1986	111	96.52173913
1987	110	95.65217391
1988	112	97.39130435
1989	113	98.26086957
1990	115	100
1991	114	99.13043478
1992	116	100.8695652
1993	118	102.6086957
1994	119	103.4782609
1995	120	104.3478261

Tabelle 4.3.37: Ergebnisse der Berechnungen

4.4 Lernziele

- Für Datensätze in Anhängigkeit von der Skalierung die geeigneten Kennwerte der mittleren Lage wählen und berechnen können.
- Für Datensätze in Anhängigkeit von der Skalierung die geeigneten Kennwerte der Streuung (mittlerer Abweichungsbetrag, Varianz, Standardabweichung, Quartilsabstand, Simpson-Index) auswählen und berechnen können.
- Die Begriffe des arithmetischen Mittels, der Quantile (Median), der Varianz und der Standardabweichung definieren können.
- Wirkung der Standardisierung von Daten kennen und sie für konkrete Datensätze von Hand, mit Excel berechnen können.
- Die Lorenzkurve für klassierte und nichtklassierte Daten zeichnen können.
- Die Grundidee der Lorenzkonzentration kennen (Lorenzkonzentration = doppelte Fläche zwischen Lorenzkurve und $x = y$ -Gerade).
- Simpsonindex berechnen und interpretieren können.
- Zwischen ein- und mehrgipfligen Verteilungen sowie zwischen rechts- und linksschiefen Verteilungen am Beispiel unterscheiden können. Wissen, wann Daten symmetrisch verteilt sind.
- Den Schiefekennwert für geeignete Daten berechnen können
- Den einfachen Index mit fixer Basis und mit beweglicher Basis berechnen können (von Hand und mit Excel)
- Den Wertindex, sowie die Preis- und Mengenindices nach Paasche und Laspeyres berechnen können (von Hand und mit Excel).
- Umbasierungen berechnen können (von Hand und mit Excel).
- Zeitreihen zeichnen und interpretieren können.

Kapitel 5

Bivariate beschreibende Statistik

In der bivariaten, beschreibenden Statistik werden zwei Variablen jeweils auf einmal behandelt. Es geht darum zu beschreiben, wie die beiden Variablen zusammenhängen. So möchte man z.B. Wissen, ob Frauen mehr rauchen als Männer (Variablen Geschlecht und Rauchverhalten), oder ob es der Fall ist, dass Raucher weniger lange leben als Nichtraucher (Variablen Rauchverhalten, Lebensdauer).

5.1 Häufigkeiten und graphische Darstellungen

Wir behandeln zuerst Variablen mit wenig Ausprägungen (nominal, ordinal oder metrisch; wenige Ausprägungen erhalten wir auch durch Klassifizierung bei einer Variable mit vielen Ausprägungen). Wie in der univariaten Statistik können wir Häufigkeitsauszählung vornehmen. Dabei zählen wir jeweils die Anzahl der Objekte, denen der Skalenwert x_i und der Skalenwert y_j der Variablen X und Y zugeordnet sind. Wir erhalten sogenannte bivariate Häufigkeitsverteilungen (= die gemeinsame Häufigkeitsverteilung der Variablen X und Y). Wir untersuchen dies an einem bereits mehrmals verwendeten Beispiel (s. Tabelle 5.1.1):

StudentInnen	Augenfarbe	Körpergrösse
1	1	1
2	2	1
3	1	1
4	3	2
5	3	2
6	1	1
7	2	2
8	2	1
9	3	1
10	1	2
11	3	1
12	3	2
13	1	2

Tabelle 5.1.1: Beispieldaten zweier Variablen

Ein „Objekt“ kann die Augenfarbe 1 und die Körpergrösse 1 haben, die Augenfarbe 2 und die Körpergrösse 1, die Augenfarbe 3 und die Körpergrösse 1, die Augenfarbe 1 und die Körpergrösse 2, die Augenfarbe 2 und die Körpergrösse 2, die Augenfarbe 3 und die Körpergrösse 2. Dies Varianten können wir mit Hilfe geordneter Paare darstellen: (1,1), (2,1), (3,1), (1,2), (2,2), (3,2). Jedem dieser geordneten Paare ordnen wir die Häufigkeit des Auftretens der entsprechenden Variante zu:

(1, 1) kommt 3 mal vor.
 (2, 1) kommt 2 mal vor.
 (3, 1) kommt 2 mal vor.
 (1, 2) kommt 2 mal vor
 (2, 2) kommt 1 mal vor
 (3, 2) kommt 3 mal vor.

Da wir jedem geordneten Paar genau eine Zahl zuordnen, liegt eine Funktion vor, als die wir die bivariate Häufigkeitsverteilung formal definieren werden. Der Graph der obigen bivariaten Häufigkeitsverteilung ist:

$$k_{X,Y} = \{((1, 1), 3), ((2, 1), 2), ((3, 1), 2), ((1, 2), 2), ((2, 2), 1), ((3, 2), 3)\}$$

Definition 5.1.1. Sei f die Skala der Variable X auf die Grundmenge G und g die Skala der Variable Y auf die Grundmenge G , und $Bild(f)$ das Bild von f und $Bild(g)$ das Bild von g . Die gemeinsame Häufigkeitsverteilung k der Variablen X und Y ist die Abbildung

$$k_{X,Y} : Bild(f) \times Bild(g) \rightarrow \mathbb{N}$$

$$k_{ij} := k_{X,Y}((x_i, y_j)) := |\{u : f(u) = x_i \text{ und } g(u) = y_j \text{ und } u \in G\}|$$

Gemeinsame Häufigkeitsverteilungen werden oft mit Hilfe sogenannter Kreuztabellen dargestellt (s. Abbildung 5.1.2):

Augenfarbe * Körpergröße Kreuztabelle				
Anzahl		Körpergröße		Gesamt
		< 1.7m	>= 1.7m	
Augenfarbe	grün	3	2	5
	blau	2	1	3
	andere	2	3	5
Gesamt		7	6	13

Tabelle 5.1.2: Kreuztabelle der Beispieldaten Tabelle 5.1.1

Neben „Gesamt“ steht in der untersten Zeile die univariate Häufigkeitsverteilung der Variable „Körpergröße“. Es handelt sich um die jeweiligen Summen einer Spalte. In diesem Zusammenhang nennen wir diese Zeile „Randverteilung“. Dasselbe gilt für die letzte Spalte unter „Gesamt“. Es werden jeweils die Zeilen zusammengezählt und wir erhalten die univariate Häufigkeitsverteilung der Variable „Augenfarbe“, die auch hier „Randverteilung“ genannt wird. Die Summen der Randverteilungen müssen identisch sein ($= n = |G|$). Bei Kreuztabellen verwenden wir künftig folgende Abkürzungen: k_{ij} für die absolute Häufigkeit der Zelle in der i -ten Zeile und der j -ten Spalte. k_{+j} für die Spaltensumme der j -ten Spalte und k_{i+} für die Zeilensumme der i -ten Zeile. Im obigen Beispiel würde damit etwa gelten:

$$k_{11} = 3; k_{12} = 2; k_{21} = 2; k_{22} = 1; k_{31} = 2; k_{32} = 3;$$

$$k_{+1} = 7; k_{+2} = 6; k_{1+} = 5; k_{2+} = 3; k_{3+} = 5;$$

Bivariate Häufigkeitsverteilungen können auch graphisch dargestellt werden. Dabei gibt es recht unterschiedliche Möglichkeiten (s. u.a. Abbildungen 5.1.1 und 5.1.2):

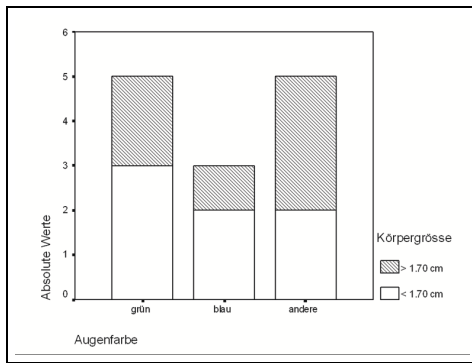


Abbildung 5.1.1: Balkendiagramm, Über-einanderstellen der Anzahlen; Beispieldaten Tabelle 5.1.1

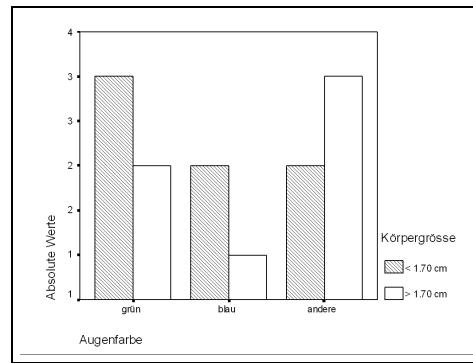


Abbildung 5.1.2: Balkendiagramm, Neben-einanderstellen der Anzahlen; Beispieldaten Tabelle 5.1.1

Weitere Häufigkeitsverteilungen

Wir erhalten eine relative gemeinsame Häufigkeitsverteilung (Abkürzung für die Häufigkeit einer Zelle h_{ij}), indem wir die Funktionswerte der absoluten, gemeinsamen Häufigkeitsverteilung durch n teilen. Damit gilt: $h_{ij} = \frac{k_{ij}}{n}$. Wir verwenden für die Randsummen die zu den absoluten Häufigkeitsverteilungen analogen Abkürzungen: h_{i+} für die Zeilensumme (i -te Zeile) und h_{+j} für die Spaltensumme (j -te Spalte).

Kumulierte gemeinsame Häufigkeitsverteilungen sind ebenfalls bildbar. Wir müssen für jede Zeile und für jede Spalte eine kumulierte Häufigkeitsverteilung erstellen (in jeder Zelle steht die Summen der Zellen, die in der Häufigkeitsverteilung links und ob ihr liegen). Kumulierte gemeinsame Häufigkeitsverteilungen sind allerdings nicht besonders instruktiv.

5.2 Streudiagramme

Ist mindestens eine der Variablen metrisch skaliert (ohne Klassifizierung), so können auch für bivariate Untersuchungen Streudiagramme erstellt werden. Je nach dem ob auch die andere Variable metrisch skaliert ist (ohne Klassifizierung) oder ob die andere Variable klassifizierte Daten aufweist (nominal, ordinal oder metrisch), ergeben sich unterschiedliche Streudiagramme.

Sind beide Variablen metrisch skaliert, ohne dass eine Klassifizierung vorliegt, so kann man jedes geordnete Paar (x_i, y_i) des Objektes i in einem Koordinatennetz darstellen. (z.B. Variable X = Körpergröße in cm und Variable Y = Gewicht in Gramm).

Beispiel 5.2.1. Beispiel für Daten, für die Streudiagramme geeignet sind (s. Tabelle 5.2.3 und Abbildung 5.2.3):

Objekt	Körpergrösse	Gewicht
1	187	80
2	100	25
3	198	86
4	130	35
5	165	69
6	178	75
7	181	77
8	178	85
9	50	4
10	175	70
11	174	72
Durchschnitt	5460:750=	7.28

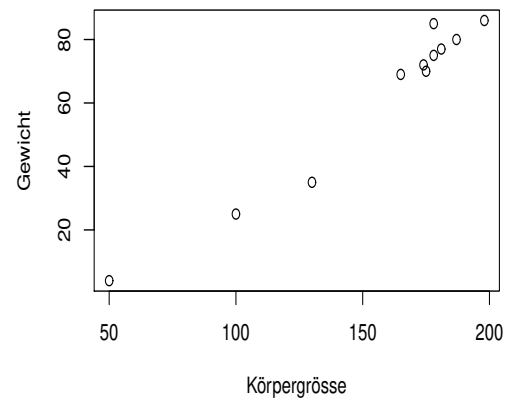


Tabelle 5.2.3: Körpergrösse und Gewicht: metrisch skalierte Daten, für Streudiagramm geeignet

Abbildung 5.2.3: Zweidimensionales Streudiagramm der Beispieldaten der Tabelle 5.2.3

◇

Übung 5.2.2. Zeichnen Sie den graphischen Zusammenhang folgender Daten von Hand und mit Excel (s. Tabelle 5.2.4):

Objekt	Gewicht	Alter
1	61	41
2	62	31
3	76	50
4	70	34
5	94	34
6	85	43
7	61	41
8	92	43
9	52	57
10	76	31
11	82	57
12	94	38

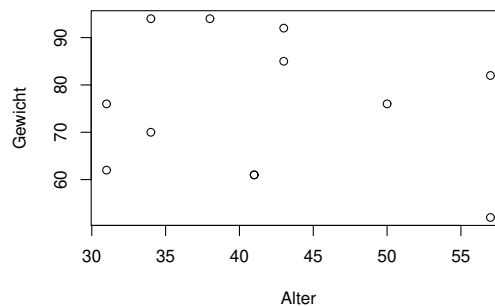


Tabelle 5.2.4: Beispieldaten zweier metrisch skalierte Datensätze

Abbildung 5.2.4: Streudiagramm der Beispieldaten der Tabelle 5.2.4

Zwischen zwei metrisch skalierten Datensätzen sind u.a. folgende Beziehungen möglich:
 (1) tendenzmässig linearer Zusammenhang: je grösser x_i , desto grösser ist y_i ; je kleiner x_i , desto kleiner y_i (proportionaler linearer Zusammenhang) (siehe Beispiel 5.2.1).

- (2) *tendenzmässig linearer Zusammenhang*: je grösser x_i , desto kleiner ist y_i ; je kleiner x_i , desto grösser y_i (*umgekehrt proportionaler linearer Zusammenhang*) (s. Beispiel 5.2.3).
 (3) *es gibt keine derartige Beziehung* (siehe Übung 5.2.2).

Beispiel 5.2.3. für tendenzmässig linearen, umgekehrt proportional Zusammenhang (s. Tabelle 5.2.5 und Abbildung 5.2.5):

Objekt	Alter	Prämien für Autounfallversicherung
1	60	350
2	30	540
3	20	470
4	35	490
5	45	450
6	44	450
7	28	550
8	64	360
9	18	600
10	54	400

Tabelle 5.2.5: Beispieldaten zweier metrisch skalierten Variablen

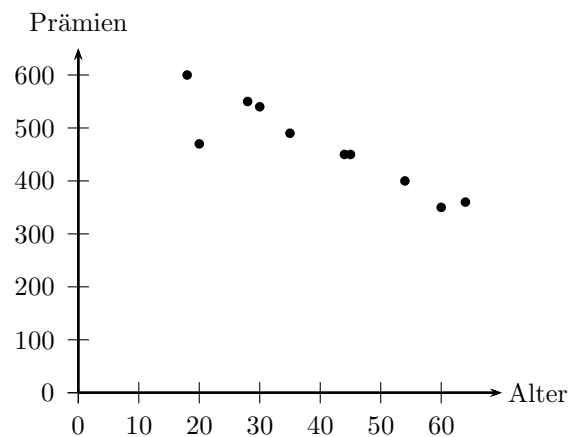


Abbildung 5.2.5: Zweidimensionales Streudiagramm der Tabelle 5.2.5

◇

Weist eine der Variablen klassifizierte Daten auf, siehe z.B. Tabelle 5.2.6,

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
50000	42000	41000	50100
40000	38000	39000	51000
45000	50100	36000	38000
43000	52000	43500	49000
60200	48000	38000	45000
55000	37000	37500	55000
47000	25000	2800	42000
58000	37500		48000
	41000		53000
			48000

Tabelle 5.2.6: Beispieldaten einer nominal und einer metrisch skalierten Variable

so können wir das folgende Streudiagramm erstellen (s. Abbildung 5.2.6).

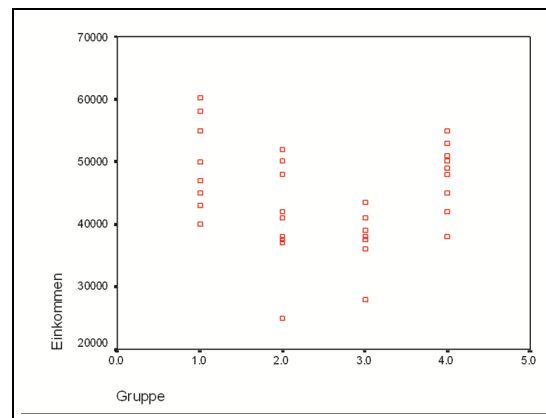


Abbildung 5.2.6: Streudiagramm der Beispieldaten Tabelle 5.2.6

Nützlich sind auch die Boxplots (s. Abbildung 5.2.7):

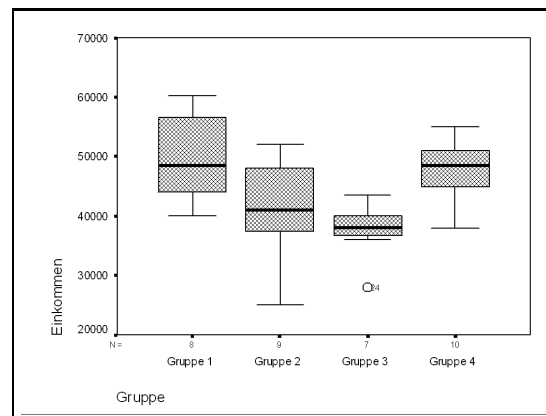


Abbildung 5.2.7: Beispieldaten Tabelle 5.2.6

Der schwarze, horizontale Strich in der Box drückt den Median aus. Die obere und untere Begrenzung der Box drücken das erste und das dritte Quartil aus. Die waagrechte Strich der T's gibt die Randwerte an, ausser bei „Ausreisser“ und „Extremwerten“. Als „Ausreisser“ werden Werte betrachtet, die zwischen 1.5- und 3-facher Boxlänge vom oberen oder unteren Rand der Box liegen. Sie werden als Kreis dargestellt. Als „Extremwerte“ werden Werte betrachtet, die ausserhalb der dreifachen Boxlänge liegen (von den Boxgrenzen an). Sie werden mit einem Stern markiert. An Box-Plots sieht man zudem sofort, ob die Daten symmetrisch verteilt sind oder nicht. Die Daten sind symmetrisch verteilt, wenn die Medianstrich in der Mitte der Box liegt. Wir können zudem einen ersten Augenschein nehmen, ob sich die Streuungen der verschiedenen Gruppen wesentlich unterscheiden. Die Definitionen der Ausreisser und der Extremwerte können von Programm zu Programm unterschiedlich sein oder können je nach Programm selber festgelegt werden.

5.3 Kennwerte

Wie bei den univariaten Kennwerten, kann man bivariate Kennwerte als Werte von Funktionen betrachten, die Datenvektoren jeweils eine Zahl zuordnen. In diesem Falle wird Teilmengen aus $\mathbb{R} \times \mathbb{R}$ mit n Elementen für ein $n \in \mathbb{N} \setminus \{0, 1\}$ eine reelle Zahl zugeordnet. Teilmengen von $\mathbb{R} \times \mathbb{R}$ mit n Elementen kann man als zweispaltige Matrix mit n Zeilen auffassen.

5.3.1 Quoten-Verhältnis (Odds-Ratio)

Für 2×2 -Kreuztabellen ist das sogenannten Quoten-Verhältnis (auch Chancenverhältnis, Odds-Verhältnis, Kreuzproduktverhältnis oder Chancenquotient genannt; englisch: Odds-Ratio) gebräuchlich. Die Gewinnquote (englisch: Odds) wird bei Spielern als das Verhältnis der Gewinnwahrscheinlichkeit zur Verlustwahrscheinlichkeit definiert. Beträgt die Gewinnwahrscheinlichkeit eines Pferdes 0.2, so ist die Gewinnquote $\frac{0.2}{0.8} = 0.25 = \frac{1}{4}$ (d.h. 1 zu 4. Es ist viermal wahrscheinlicher, dass es nicht gewinnt, als dass es gewinnt). Ist die Gewinnwahrscheinlichkeit eines Pferdes 0.8, ist die Gewinnquote $\frac{0.8}{0.2} = 4$. Es ist viermal wahrscheinlicher, dass es gewinnt, als dass es nicht gewinnt. Gewinnquoten können aus absoluten Häufigkeiten gewonnen werden. Hat ein Pferd 12 mal gewonnen und 4 mal verloren, so ergibt sich $\frac{12}{4} = 3$ als Quote. Es ist dreimal so wahrscheinlich, dass es gewinnt, als dass es nicht gewinnt (wenn alle übrigen Umstände gleich sind). Hat es demgegenüber 4 mal gewonnen und 12 mal verloren, erhält man $\frac{4}{12} = \frac{1}{3}$. Es ist dreimal wahrscheinlicher, dass es verliert, als dass es gewinnt. Quoten werden oft logarithmiert und werden dann „Logit“ oder „Logodds“ genannt.

In der Kreuztabelle

	Raucher	Nichtraucher
Männer	20	10
Frauen	15	20

ist $\frac{20}{10} = 2$ die Quote, dass ein Mann Raucher ist und $\frac{15}{20} = 0.75$ die Quote, dass eine Frau Raucherin ist. Das Verhältnis dieser Quoten (= Quotenverhältnis) beträgt

$$OR = \frac{\frac{k_{1,1}}{k_{1,2}}}{\frac{k_{2,1}}{k_{2,2}}} = \frac{k_{1,1}k_{2,2}}{k_{1,2}k_{2,1}},$$

ist im Beispiel also $\frac{20}{10} = \frac{20 \cdot 20}{15 \cdot 10} = \frac{2}{0.75} = 2.6667$. Die Quote als Mann zu rauchen ist also 2.666 mal grösser als die entsprechend Frauenquote. Die OR hängt von der Reihenfolge der Spalten oder Zeilen ab. So ergibt sich für die äquivalente Tabelle

	Raucher	Nichtraucher
Frauen	15	20
Männer	20	10

$OR = \frac{15 \cdot 10}{20 \cdot 20} = \frac{3}{8} = 0.375$. Die Quote als Frau Raucherin zu sein ist kleiner als bei den Männern. 2.6667 ist von 1 weiter weg als 0.375, obwohl die gleiche Abweichung in den Quoten gemessen wird. Rechnet man $\frac{1}{0.375} = 2.6667$ erhält man aber wiederum dieselbe Zahl. Die Frauen haben eine um 2.6667 kleinere Quote zu rauchen als die Männer. Wenn $OR = 1$, unterscheiden sich die Quoten nicht. Die Quoten streuen also um 1 herum, im Intervall $]0, \infty[$. Möchte man Abstände $1 - OR_u$ mit $OR \in]0, 1[$ und Abstände $OR_o - 1$ mit $OR_o \in]1, \infty[$ vergleichen, so berechnet man $\frac{1}{OR_u}$.

Ohne diese Umrechnung erhält man bei gleichen Unterschieden gleiche Abstände, wenn man die OR logarithmiert. Es gilt nämlich $\ln x = -\ln \frac{1}{x}$ für $x > 0$. Im Beispiel ergibt sich: $\ln(0.375) = -0.98083$ und $\ln(2.6667) = 0.98083$. Die Zahlen streuen nun statt um 1 um 0 herum und die Umkehrung der Reihenfolge in den Zeilen führt zu identischen Abständen von 0. Logarithmierte Quotenverhältnisse werden „Log-Quotenverhältnis“ genannt.

Zu beachten ist, dass die Quotenverhältnisse der Spalten mit den Zeilenquotenverhältnisse identisch sind. Es gilt

$$\frac{\frac{15}{20}}{\frac{20}{10}} = 0.375,$$

und wenn man die Reihenfolge der Spalten umkehrt

$$\frac{\frac{20}{10}}{\frac{15}{20}} = 2.6667.$$

Verändert man die Reihenfolge der Zeilen und bildet das Spaltenquotenverhältnis, erhält man wiederum

$$\frac{\frac{20}{15}}{\frac{10}{20}} = 2.666\,7$$

oder bei Umkehrung der Reihenfolge der Spalten

$$\frac{\frac{10}{20}}{\frac{20}{15}} = 0.375.$$

Das Quotenverhältnis kann auch in $m \times n$ -Tabellen verwendet werden, um jeweils 4 Zellen von zwei Ausprägungen der entsprechenden Variablen zu vergleichen. Man erhält dann allerdings nicht eine Kennzahl für die gesamte Kreuztabelle.

Um bei allenfalls leeren Zellen eine Division durch 0 zu verhindern, wird oft zu den $k_{i,i}$ die Zahl 0.5 hinzugezählt. Im Beispiel würde sich ergeben:

$$\frac{(20 + 0.5) \cdot (20 + 0.5)}{(15 + 0.5) \cdot (10 + 0.5)} = 2.582\,2$$

Bemerkung 5.3.1. Sind $\frac{k_{11}}{k_{11}+k_{12}}$ und $\frac{k_{21}}{k_{21}+k_{22}}$ sehr klein oder sehr gross, so drückt das Quotenverhältnis näherungsweise aus, um wie viele Male das Risiko grösser (kleiner) ist, dass Männer Raucher sind als dass Frauen Raucherinnen sind - im obigen Fall mit $\frac{20}{20+10} = \frac{2}{3}$ und $\frac{15}{15+20} = \frac{3}{7}$ sind diese Zahlen zu wenig nahe bei 0 oder bei 1. Das Risikoverhältnis, auch „relatives Risiko“ genannt, wird definiert durch

$$\frac{\frac{k_{11}}{k_{11}+k_{12}}}{\frac{k_{21}}{k_{21}+k_{22}}}.$$

Es drückt also aus, um wie viele Male es wahrscheinlicher ist, für Objekte in der ersten Zeile in der ersten Kategorie zu liegen, als für Objekte in der zweiten Zeile in der ersten Kategorie zu liegen. Es wird im Gegensatz zum Quotenverhältnis mit Hilfe der Zeilenrandverteilungen definiert.

Im Beispiel gilt für das relative Risiko $\frac{\frac{20}{30}}{\frac{15}{35}} = 1.555\,6$ und die OR überschätzt das Risikoverhältnis stark. $\diamond$

5.3.2 Der Kontingenzkoeffizient

Als Kennzahl für die Unabhängigkeit von zwei nominalskalierten Variablen bietet sich unter anderem der sogenannte Kontingenz-Koeffizient an. Es soll untersucht werden, ob Haar- und Augenfarbe irgendwie zusammenhängen oder unabhängig sind. Wir betrachten das folgende

Beispiel 5.3.2. Gegeben sei die folgende Kreuztabelle

		Haarfarbe			
Augenfarbe	schwarz	braun	rot	blond	Summen
braun	37	62	14	5	118
blau	11	40	7	52	110
grau/grün	12	42	13	10	77
Summen	60	144	34	67	305

Tabelle 5.3.7: Beispiel für die Berechnung des Kontingenz-Koeffizienten

Wenn die Augen- von der Haarfarbe unabhängig ist, dann müssten in den Zellen der Kreuztabellen die Häufigkeiten durch die Randverteilungen bestimmt sein: der Anteil der Personen mit

braunen Augen an allen Personen beträgt $\frac{118}{305}$. Es gibt 60 Personen mit schwarzen Haaren. Sind die Farben unabhängig, so sollten $\frac{118}{305}$ von diesen 60 Personen braune Augen haben. Dies ergibt

$$k_{11}^e = \frac{118}{305} \cdot 60 = 23.213$$

k_{ij}^e werden die (bei Unabhängigkeit) „erwarteten Häufigkeiten“ genannt. Diese Überlegung führen wir für alle Zellen i, j durch:

$$\begin{array}{llll} k_{11}^e = \frac{118}{305} \cdot 60 & k_{12}^e = \frac{118}{305} \cdot 144 & k_{13}^e = \frac{118}{305} \cdot 34 & k_{14}^e = \frac{118}{305} \cdot 67 \\ k_{21}^e = \frac{110}{305} \cdot 60 & k_{22}^e = \frac{110}{305} \cdot 144 & k_{23}^e = \frac{110}{305} \cdot 34 & k_{24}^e = \frac{110}{305} \cdot 67 \\ k_{31}^e = \frac{77}{305} \cdot 60 & k_{32}^e = \frac{77}{305} \cdot 144 & k_{33}^e = \frac{77}{305} \cdot 34 & k_{34}^e = \frac{77}{305} \cdot 67 \end{array}$$

Anschliessend berechnet man die Summe der gewichteten quadrierten Abweichungen der faktischen Häufigkeiten k_{ij} von den erwarteten Häufigkeiten k_{ij}^e :

$$\chi^2 := \sum_{i=1}^I \sum_{j=1}^J \frac{(k_{ij} - k_{ij}^e)^2}{k_{ij}^e}$$

wobei I die Anzahl Zeilen ist und J die Anzahl der Spalten. Die Abweichungen werden quadriert, um eine von 0 verschiedene Summe zu erhalten und um negative wie positive Abweichungen gleich zu behandeln. Die quadrierten Abweichungen werden mit

$$\frac{1}{k_{ij}^e}$$

gewichtet, um die quadrierten Differenzen angesichts unterschiedlich grosser erwarteter Häufigkeiten vergleichbar zu machen. Im Beispiel erhalten wir:

$$\begin{aligned} \chi^2 &= \frac{(37 - 23.213)^2}{23.213} + \frac{(62 - 55.711)^2}{55.711} + \frac{(14 - 13.154)^2}{13.154} + \frac{(5 - 25.921)^2}{25.921} \\ &+ \frac{(11 - 21.639)^2}{21.639} + \frac{(40 - 51.934)^2}{51.934} + \frac{(7 - 12.262)^2}{12.262} + \frac{(52 - 24.164)^2}{24.164} \\ &+ \frac{(12 - 15.148)^2}{15.148} + \frac{(42 - 36.354)^2}{36.354} + \frac{(13 - 8.584)^2}{8.584} + \frac{(10 - 16.915)^2}{16.915} \\ &= 74.7663 \end{aligned}$$

Um Tabellen mit verschieden grosser Gesamtanzahl n vergleichbar zu machen wird χ^2 gewichtet mit

$$\frac{1}{n + \chi^2}.$$

Es wird dann Kontingenzkoeffizient K , wie folgt, definiert:

$$K = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

Der Kontingenzkoeffizient ist offenbar eine Zahl zwischen 0 und 1, da der Nenner immer grösser als der Zähler ist. Präziser gilt:

$$0 \leq \sqrt{\frac{\chi^2}{n + \chi^2}} \leq \sqrt{\frac{\min\{I, J\} - 1}{\min\{I, J\}}}.$$

Für die untere Grenze ist dies offensichtlich. $\chi^2 = 0$ genau dann, wenn $k_{ij} = k_{ij}^e$ für alle i, j . Die obere Grenze wird erreicht, wenn pro Spalte und pro Zeile alle Häufigkeiten nur in einer Zelle sind, wobei jede Zeile und jede Spalte bei einer quadratischen Kreuztabelle genau einmal besetzt ist (s. als Beispiel die folgende Tabelle, A_{ij} für die j -te Ausprägung der i -te Variable)

	A_{21}	A_{22}	A_{23}	A_{24}	Summen
A_{11}	30	0	0	0	30
A_{12}	0	30	0	0	30
A_{13}	0	0	0	50	50
A_{24}	0	0	40	0	40
Summen	30	30	40	50	150

Um eine Kennzahl zu erhalten, die zwischen 0 und 1 variiert, wird der Kontingenzkoeffizient manchmal normiert:

$$K^* = \frac{1}{\sqrt{\frac{\min\{I, J\} - 1}{\min\{I, J\}}}} K$$

Im Beispiel gilt: $K = \sqrt{\frac{74.7663}{305 + 74.7663}} = 0.44371$ und $K^* = \frac{0.44371}{\sqrt{\frac{2}{3}}} = 0.54343$. Die beiden Variablen scheinen nicht unabhängig zu sein (völlig frei erfundene Daten). $\diamond$

5.3.3 Goodman und Kruskal's τ (= Konzentrationskoeffizient) (fakultativer Stoff)

Goodman und Kruskal's τ (τ = tau, griechisches t) wird für nominal skalierte Daten verwendet. Wir berechnen es mit Hilfe der entsprechenden Kreuztabelle der relativen Häufigkeiten.

$$\tau = \frac{\sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} - \sum_{j=1}^J h_{+j}^2}{1 - \sum_{j=1}^J h_{+j}^2}$$

Dabei ist h_{ij} die relative Häufigkeit der Zelle (i, j) , h_{i+} die Randsumme der i -ten Zeile. h_{+j} die Randsumme der j -ten Spalte.

$$1 - \sum_{j=1}^J h_{+j}^2$$

ist der nichtnormierte Simpsonindex der Spaltenrandsummen.

$$1 - \sum_{j=1}^J h_{ij}^2$$

ist der Simpsonindex für die Zeile i . Messen wir diesen an der Zeilenrandsumme, erhält man

$$\frac{1}{h_{i+}} \left(1 - \sum_{j=1}^J h_{ij}^2 \right) = \frac{1}{h_{i+}} - \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}}.$$

Es handelt sich um einen Simpsonindex, der mit Zeilenrandsumme gewichtet wird. Summieren wir

über alle Zeilen, erhält man:

$$\begin{aligned}\sum_{i=1}^I \left(\frac{1}{h_{i+}} - \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} \right) &= \sum_{i=1}^I \frac{1}{h_{i+}} - \sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} \\ &= 1 - \sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}}.\end{aligned}$$

Es handelt sich also um die Summe der mit den Zeilenrandsummen gewichteten Streumasse pro Zeile. Diese Zahl drückt aus, wie stark die Häufigkeiten innerhalb der Zeilen streuen - gewichtet mit der Bedeutung der Zeilen (bei grosser Zeilensumme wird weniger von 1 abgezählt. Auf den Zähler des obigen Ausdruck kommen wir nun durch

$$\left(1 - \sum_{j=1}^J h_{ij}^2 \right) - \left(1 - \sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} \right) = \sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} - \sum_{j=1}^J h_{ij}^2.$$

Es handelt sich um die Differenz zwischen der Streuung der Spaltenrandsummen und der Summe der gewichteten Streuungen in den Zeilen. Diese wird an der Streuung der Spaltenrandsumme gemessen.

Der Kennwert drückt intuitiv folgendes aus: Wir nehmen an, wir müssten raten, in welcher Spalte ein Objekt liegt. Wenn wir nur die Spaltenrandverteilung kennen, so wählen wir als Voraussage am besten die Spalte mit der höchsten Randsumme. Kennen wir aber zusätzlich die Verteilungen innerhalb der Zeilen und wissen wir, in welche Zeile das Objekt zu liegen kommt, wählen wir am besten die Zelle, die in der Zeile die grösste Häufigkeit hat. Das Konzentrationsmass drückt nun das Ausmass an Zusatzinformation aus, die wir dadurch gewinnen, dass wir wissen, in welcher Zeile das Objekt zu liegen kommt. Sind in einer Zeile in jeder Zelle gleich viele Objekte, gewinnen wir durch die Kenntnisse, dass das Objekt in einer spezifischen Zeile liegt, nichts hinzu; sind die Objekte in den Zeilen sehr ungleich verteilt, gewinnen wir durch die Kenntnis, in welcher Zeile das Objekt liegt, viel hinzu.

Goodman und Kruskal's τ variiert zwischen 0 und 1. Es ergibt sich ein Wert von 0, wenn in jeder Zelle einer jeden Zeile die selbe Häufigkeit auftaucht. Es ergibt sich ein Wert von 1, wenn in jeder Zeile und Spalte alle Zellen ausser einer 0 sind.

Goodman und Kruskal's τ hat den Nachteil, dass der Wert des Kennwertes mit der Anzahl der Spalten sinkt. Entsprechend kann man sinnvoller Weise nur Tabellen mit derselben Spaltenanzahl vergleichen.

Beispiel 5.3.3. Wir betrachten eine Menge von 101 Personen, und berücksichtigen dabei die Variablen „Geschlecht“ und „Rauchverhalten“. Die Resultate sind (s. 5.3.8):

◇

Geschlecht * Rauchverhalten Kreuztabelle				
Anzahl		Rauchverhalten		Gesamt
		Nichtraucher	Raucher	
Geschlecht	männlich	20	30	50
	weiblich	25	26	51
Gesamt		45	56	101

Tabelle 5.3.8: Beispieldaten in Kreuztabelle zusammengefasst

Wir berechnen den Konzentrationskoeffizienten:

Kreuztabelle mit den relativen Häufigkeiten und den entsprechenden Randverteilungen:

$\frac{20}{101} = 0.198\,02$	$\frac{30}{101} = 0.297\,03$	$\frac{20+30}{101} = 0.495\,05$
$\frac{25}{101} = 0.247\,52$	$\frac{26}{101} = 0.257\,43$	$\frac{25+26}{101} = 0.504\,95$
$\frac{20+25}{101} = 0.445\,54$	$\frac{30+26}{101} = 0.554\,46$	1

Tabelle 5.3.9: Berechnung des Simpsonindices für die Beispieldaten Abbildung 5.3.8

$$\sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} = \frac{0.19802^2}{0.49505} + \frac{0.29703^2}{0.49505} + \frac{0.24752^2}{0.50495} + \frac{0.25743^2}{0.50495} = 0.510\,00$$

$$\sum_{j=1}^J h_{+j}^2 = 0.44554^2 + 0.55446^2 = 0.505\,93$$

$$\tau = \frac{0.510\,00 - 0.505\,93}{1 - 0.505\,93} = 0.008232182$$

Es gibt weitere Kennwerte des Zusammenhangs in einer Kreuztabelle mit nominalskalierten Daten. Wir behandeln diese nicht.

5.3.4 Goodman und Kruskal's Gamma

($=\gamma$; kleiner griechischer Buchstabe für g)

Gamma ist ein Kennwert, die sich für ordinalskalierte Daten eignet, wenn pro Variable wenige Ausprägungen vorliegen. Liegen die Häufigkeiten einer Kreuztabelle zweier ordinalskalierter Variablen auf der Hauptdiagonalen, so erfolgt ein positives Gamma, liegen die Häufigkeiten auf der Nebendiagonalen, ergibt sich ein negatives Gamma. Ein positives Gamma bedeutet entsprechend, dass Objekte, die in der Kreuztabelle weiter rechts liegen, im allgemeinen auch weiter unten liegen. Als konkretes Beispiel betrachten wir die Variablen „Zufriedenheit“ mit den Ausprägungen „sehr zufrieden“, „zufrieden“, „mässig zufrieden“ und „unzufrieden“ sowie „Vermögen“ mit den Ausprägungen „sehr reich“, „reich“, „mittelständisch“ und „arm“. Gilt die Aussage „Je reicher jemand ist, desto zufriedener ist er“? Die Aussage trifft zu, wenn sich die Häufigkeiten wie in der folgenden Kreuztabelle (s. Tabelle 5.3.10) auf der Hauptdiagonalen massieren:

		Zufriedenheit			
		unzufrieden	mässig zufrieden	zufrieden	sehr zufrieden
Vermögen	arm	20	5	2	0
	mittelständisch	3	30	9	8
	reich	4	6	25	10
	sehr reich	1	2	7	15

Tabelle 5.3.10: Kreuztabelle für Beispiel Anwendung von Gamma

Gamma ist wie folgt definiert:

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d}$$

$$\Pi_c = \sum_{i=1}^I \sum_{j=1}^J k_{ij} \left(\sum_{I \geq h > i} \sum_{J \geq k > j} k_{hk} \right)$$

$$\Pi_d = \sum_{i=1}^I \sum_{j=1}^J k_{ij} \left(\sum_{I \geq h > i} \sum_{1 \leq k < j} k_{hk} \right)$$

Wir erklären die Berechnung am Beispiel: Wir berechnen zuerst die Anzahl der „konkordanten“ Paare. Es handelt sich um die Anzahl von jeweils zwei Objekten a und b für die gilt: b hat ein höheres Vermögen als a und ist zufriedener als a . Es sind die Paare, welche die obige Hypothese stützen. In der Zelle a_{11} haben wir 20 Personen. Für jede dieser 20 Personen gibt es $(30 + 9 + 8 + 6 + 25 + 10 + 2 + 7 + 15)$ Personen, die sowohl zufriedener sind als auch mehr verdienen. Derart gehen wir alle Zellen durch und bilden am Schluss die Summe aller konkordanten Paare. Dies ergibt im Beispiel:

$$\begin{aligned}\Pi_c = & 20(30 + 9 + 8 + 6 + 25 + 10 + 2 + 7 + 15) + \\ & 5(9 + 8 + 25 + 10 + 7 + 15) + \\ & 2(8 + 10 + 15) + \\ & 3(6 + 25 + 10 + 2 + 7 + 15) + \\ & 30(25 + 10 + 7 + 15) + \\ & 9(10 + 15) + \\ & 4(2 + 7 + 15) + \\ & 6(7 + 15) + \\ & 25(15) = 5409\end{aligned}$$

Analog berechnen wir die „diskordanten“ Paare. Es handelt sich um Paare, die sich nicht konkordant verhalten, d.h. b hat ein grösseres Vermögen als a und ist unzufriedener als a . In der Zelle a_{14} mit 0 Personen erhalten wir $0(3 + 30 + 9 + 4 + 6 + 25 + 1 + 2 + 7)$ Personen, die mehr verdienen und unzufriedener sind. Insgesamt:

$$\begin{aligned}\Pi_d = & 0(3 + 30 + 9 + 4 + 6 + 25 + 1 + 2 + 7) + \\ & 2(3 + 30 + 4 + 6 + 1 + 2) + \\ & 5(3 + 4 + 1) + \\ & 8(4 + 6 + 25 + 1 + 2 + 7) + \\ & 9(4 + 6 + 1 + 2) + \\ & 30(4 + 1) + \\ & 10(1 + 2 + 7) + \\ & 25(1 + 2) + \\ & 6(1) = 940\end{aligned}$$

Offenbar besteht ein „positiver“ Zusammenhang zwischen den beiden Variablen, wenn es mehr konkordante als diskordante Paare hat. Es besteht ein umgekehrter Zusammenhang zwischen den beiden Variablen, wenn es mehr diskordante Paare als konkordante Paare hat. Es besteht kein Zusammenhang, wenn es gleichviele konkordante wie diskordante Paare hat. Wir messen das Grössenverhältnis zwischen diskordanten und konkordanten Paaren, indem wir die Differenz dieser zwei Grössen an der Summe der beiden Grösse messen. Im Beispiel erhalten wir.

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d} = \frac{5409 - 940}{5409 + 940} = 0.70389$$

Im Beispiel besteht also ein positiver Zusammenhang zwischen Vermögen und Zufriedenheit: je grösser das Vermögen, desto grösser die Zufriedenheit (frei erfundene Zahlen. Gemäss Studien, die mit erhobenen Daten arbeiten, gibt es für tiefe Vermögenswerte zuerst einen positiven Zusammenhang, der sich mit zunehmendem Vermögen aber abschwächt).

Gamma variiert zwischen -1 und 1 . Da $\Pi_c, \Pi_d \geq 0$, gilt nämlich: $|\Pi_c - \Pi_d| \leq \Pi_c + \Pi_d$. Zudem gilt: Wenn $\Pi_d = 0$, dann gilt: $\frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d} = \frac{\Pi_c}{\Pi_c} = 1$.

Wenn $\Pi_c = 0$, dann gilt: $\frac{-\Pi_d}{\Pi_d} = -1$.

Bei positivem Gamma liegen die Häufigkeiten in der Kreuztabelle konzentriert auf der Hauptdiagonalen (von oben links nach unten rechts). Bei negativem Gamma liegen die Häufigkeiten konzentriert auf der Nebendiagonalen (von oben rechts nach unten links). Es ist auch möglich, dass die Häufigkeiten U-förmig konzentriert vorkommen. Dann ist die Verwendung von Gamma nicht angesagt.

Bemerkung 5.3.4. Erläuterungen zur Interpretation von Gamma (s. Tabellen 5.3.11 und 5.3.12)

	unzufrieden	mässig zufrieden	zufrieden	sehr zufrieden
arm	20	10	5	0
knapp nicht arm	10	30	20	5
wohlhabend	5	15	40	22
reich	5	10	20	30

Tabelle 5.3.11: Positives Gamma: Häufigkeiten konzentriert auf Hauptdiagonale

	sehr zufrieden	zufrieden	mässig zufrieden	unzufrieden
arm	0	5	10	20
knapp nicht arm	5	20	30	10
wohlhabend	22	40	15	5
reich	30	20	10	5

Tabelle 5.3.12: Negatives Gamma: Häufigkeiten konzentriert auf Nebendiagonale

Beachten Sie, dass in den beiden Tabellen 5.3.11 und 5.3.12 dieselben Daten verwendet wurden, die Reihenfolge der Ausprägungen in den Spalten aber umgekehrt wurde. Bei der Interpretation von Gamma ist immer die Kreuztabelle und die Beschriftung der Ausprägungen der Variablen zu berücksichtigen. $\diamond$

Es gibt weitere Kennwerte des ordinalen Zusammenhangs. Kendalls τ_b etwa ist mit Gamma nahe verwandt. Wir behandeln diese und weitere Kennzahlen nicht.

Bemerkung 5.3.5. Ist eine der beiden Variablen dichotom (d.h. sie hat zwei Ausprägungen) und die andere ordinal skaliert, so drückt Gamma einen Unterschied zwischen den beiden Ausprägungen der dichotomen Variable aus. Hat man z.B. die Variable „Geschlecht“ mit den Ausprägungen „männlich“ und „weiblich“, sowie die Variable „Zufriedenheit“ mit den Ausprägungen „sehr zufrieden“, „zufrieden“, „mässig zufrieden“ und „unzufrieden“ mit der 2×4 Kreuztabelle

	sehr zufrieden	zufrieden	mässig zufrieden	unzufrieden
männlich				
weiblich				

so sind die konkordanten Paare jene, für die die Frauen weniger zufrieden sind als Männer. Die diskordanten Paare sind jene, für die Frauen zufriedener sind als die Männer. Ein positives Gamma drückt also aus, dass die Männer im Durchschnitt zufriedener sind als Frauen, ein negatives, dass die Frauen im Durchschnitt zufriedener sind als Männer. $\diamond$

5.3.5 Der Pearson-Korrelationskoeffizient

Seien Daten zweier metrisch skalierten Variablen X und Y gegeben. Liegt die Punktwolke der Punkte (x_i, y_i) um eine Gerade herum und sind die Punkte in etwa punktsymmetrisch bezüglich dem Punkt $(\bar{y}, \bar{x})$, so wird der (Pearsonsche) Korrelationskoeffizient als Masszahl des Zusammenhangs zwischen Daten zweier metrisch skalierten Variablen X und Y verwendet (Karl Pearson, 1857 - 1936, britischer Mathematiker). Betrachten wir als Beispiel Punkte (x_i, y_i) , welche pro Person i deren Einkommen x_i und Ausgaben für Schmuck y_i ausdrücken. Liegen diese Punkte um eine steigende Gerade herum, gilt die Aussage: je mehr jemand verdient, desto mehr gibt er für Schmuck aus. Um die Stärke des Zusammenhangs zu messen wurde unter anderem die Pearson-Korrelation entwickelt, die wie folgt definiert ist:

Definition 5.3.6. Pearson-Korrelationskoeffizient:

$$r_{xy} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

Dabei ist s_x die Standardabweichung der Daten der Variable X und s_y die Standardabweichung der Daten der Variable Y . Der Pearson-Korrelationskoeffizient wird auch „Pearsonsche

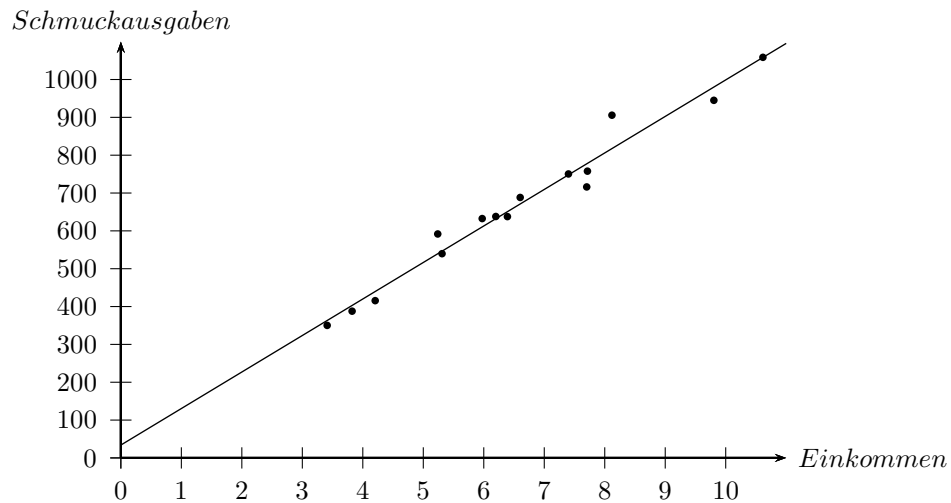


Abbildung 5.3.8: Zweidimensionales Streudiagramm der Variablen Einkommen und Ausgaben für Schmuck.

Produktmomenten-Korrelation“ genannt. Der Korrelationskoeffizient ist dimensionslos, da sich die Einheiten wegekürzen.

Diese Definition kann durch folgender Überlegungen plausibel gemacht werden: Um ein Mass für den Zusammenhang der Daten der beiden Variablen X und Y zu haben, das einen Vergleich der Zusammenhangsstärke verschiedener Datensätze erlaubt, müssen wir einen Kennwert entwickeln, die bezüglich der Mittelwerte und der Varianzen standardisiert ist. Entsprechend verwenden wir standardisierte Daten: $\frac{x_i - \bar{x}}{s_x}$ und $\frac{y_i - \bar{y}}{s_y}$.

Das Produkt

$$\frac{x_i - \bar{x}}{s_x} \frac{y_i - \bar{y}}{s_y} = \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$$

ist positiv, wenn beide Abweichungen $x_i - \bar{x}$ und $y_i - \bar{y}$ positiv oder beide negativ sind. Also genau dann, wenn die Abweichungen in dieselbe Richtung aufweisen (je grösser x_i , desto grösser y_i). Zudem wird dieses Produkt gross, wenn die Beträge der Abweichungen gross sind. Das Produkt wird noch grösser, wenn diese Beträge bei beiden Variablen gross sind.

Das obige Produkt ist negativ, wenn die Abweichungen $x_i - \bar{x}$ und $y_i - \bar{y}$ umgekehrte Vorzeichen haben. Zudem wird dieses Produkt klein, wenn die Beträge der Abweichungen gross sind. Das Produkt wird noch kleiner, wenn diese Beträge bei beiden Variablen gross sind. Diese Überlegungen zeigen: $\sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$ wird klein, wenn viele Abweichungen in unterschiedliche Richtungen weisen, d.h. wenn die beiden Variablen umgekehrt proportional zusammenhängen. (je grösser x_i , desto kleiner y_i ; je kleiner x_i , desto grösser y_i (s. Abbildung 5.3.9).

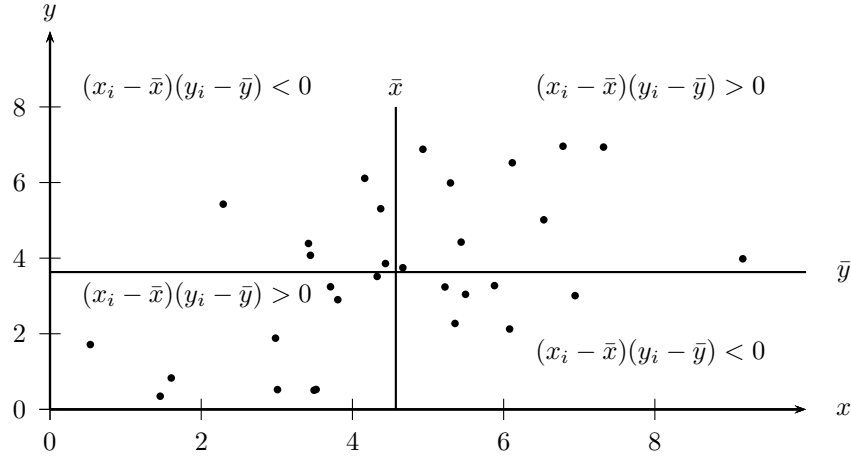


Abbildung 5.3.9: Positive und negative Produkte $(x_i - \bar{x})(y_i - \bar{y})$ je nach Lage der Punkte (x_i, y_i) bezüglich $\bar{x}$ und $\bar{y}$.

Schliesslich standardisieren wir die Grösse $\sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}$ mit $\frac{1}{n-1}$, um die Kennwerte verschieden grosser Datensätze vergleichbar zu machen. Wir erhalten:

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} &= \frac{1}{n-1} \frac{1}{s_x s_y} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \\ &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \end{aligned}$$

Man kann beweisen, dass die Korrelation einen Wert zwischen -1 und 1 annimmt. Ein Wert in der Nähe von 1 drückt eine sogenannte positive Korrelation oder einen proportionalen, linearen Zusammenhang zwischen den beiden Variablen aus. Ein Wert in der Nähe von -1 drückt eine negative Korrelation oder einen umgekehrt proportionalen linearen Zusammenhang zwischen den beiden Variablen aus. Ein Wert um 0 drückt das Fehlen eines Zusammenhangs zwischen den beiden Variablen aus. Liegen die Punkte genau auf einer Geraden (bei Daten aus Zufallsstichproben ist dies sehr unwahrscheinlich), ergibt sich ein Korrelationskoeffizient von 1 , sofern die Steigung der Geraden positiv ist und -1 , sofern die Steigung der Geraden negativ ist.

Statt eines Beweises von $r_{x,y} \in [-1, 1]$ eine Begründung: Eine maximale positive Korrelation besteht zwischen einer Variablen mit sich selber. Wir erhalten in diesem Fall:

$$\frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{s_x s_x} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}{s_x^2} = \frac{s_x^2}{s_x^2} = 1.$$

Eine maximale, negative Korrelation besteht zwischen einer Variablen und der Negation der Werte dieser Variablen. Der Mittelwert der negierten Werte ist $-\bar{x}$. Denn gilt $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, gilt

$$\frac{1}{n} \sum_{i=1}^n (-x_i) = -\frac{1}{n} \sum_{i=1}^n x_i = -\bar{x}.$$

Die Standardabweichung der Negation der Werte dieser Variablen ist s_x , denn

$$\begin{aligned} \frac{1}{n-1} \sum_{i=1}^n (-x_i - -\bar{x})^2 &= \frac{1}{n-1} \sum_{i=1}^n (-x_i + \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (-1 \cdot (x_i - \bar{x}))^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (-1)^2 (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Wir erhalten somit:

$$\begin{aligned} \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(-x_i - (-\bar{x}))}{s_x s_x} = \\ &= \frac{-\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})}{s_x s_x} = \frac{-s_x^2}{s_x s_x} = -1. \end{aligned}$$

Es gilt zu beachten, dass der Korrelationskoeffizient (wie der Mittelwert) keine robuste Masszahl ist. Ausreisser führen zu Verzerrungen. Zudem können sehr unterschiedliche Punktwolken den selben Korrelationskoeffizienten haben. So haben z.B. die folgenden Datensätze jeweils denselben Korrelationskoeffizienten von 0.7 (s. Abbildungen 5.3.10 bis 5.3.15):

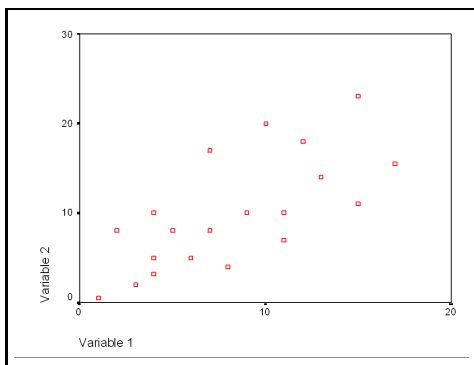


Abbildung 5.3.10: Linearer Zusammenhang

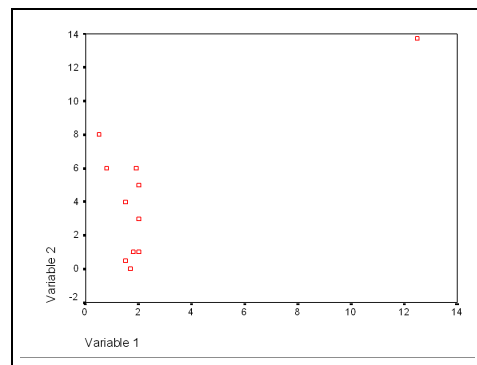


Abbildung 5.3.11: Scheinkorrelation durch Ausreisser

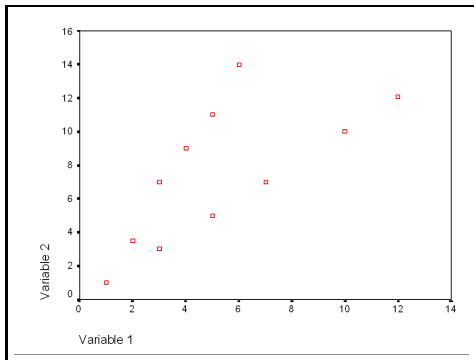


Abbildung 5.3.12: Vermutlich zwei Gruppen von Daten

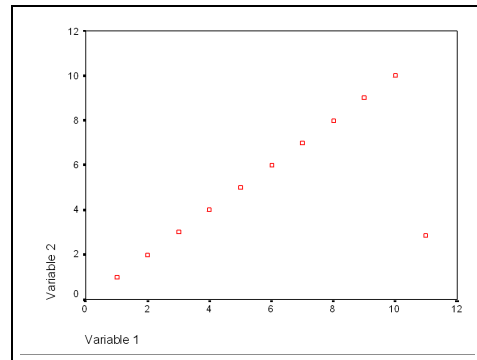


Abbildung 5.3.13: Ausreisser führt zu tieferem r

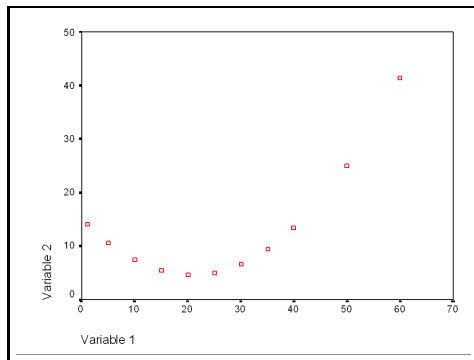


Abbildung 5.3.14: nicht linearer Zusammenhang

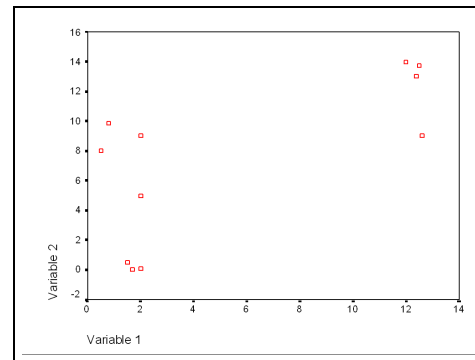


Abbildung 5.3.15: Zwei Punktwolken

Entsprechend sollte man sich nicht damit begnügen, den Korrelationskoeffizienten rechnerisch zu bestimmen. Es ist anhand einer Graphik zu überprüfen, ob:

1. die Punktwolke um eine Gerade liegt,
2. die Punkte um den Mittelpunkt $(\bar{x}, \bar{y})$ punktsymmetrisch liegen und
3. keine Ausreisser vorliegen.

Nur wenn diese Bedingungen erfüllt sind, ist der Korrelationskoeffizient für den linearen Zusammenhang zweier Variablen aussagekräftig. Liegt ein klarer, aber nicht linearer Zusammenhang vor, so kann man die Daten oft durch Transformationen in einen linearen Zusammenhang bringen (z.B. durch logarithmieren, radizieren). Auch die Verletzung der Symmetriebedingung kann man oft durch Transformationen beseitigen.

Beispiel 5.3.7. Wir betrachten ein Beispiel (s. Tabelle 5.3.13):

Objekt	x_i	y_i	$\ln y_i$
1	5	4	1.386294361
2	2	1.5	0.405465108
3	4	2	0.693147181
4	3	2.5	0.916290732
5	10	15	2.708050201
6	6	4.826809	1.574185587
7	7	7	1.945910149
8	8	8.15730721	2.098914116
9	9	11	2.397895273
10	1	1.3	0.262364264
11	11	18	2.890371758

Tabelle 5.3.13: Beispieldaten für Daten, die eine Transformation nötig machen

Wir erhalten die folgende Punktwolke (s. Abbildung 5.3.16):

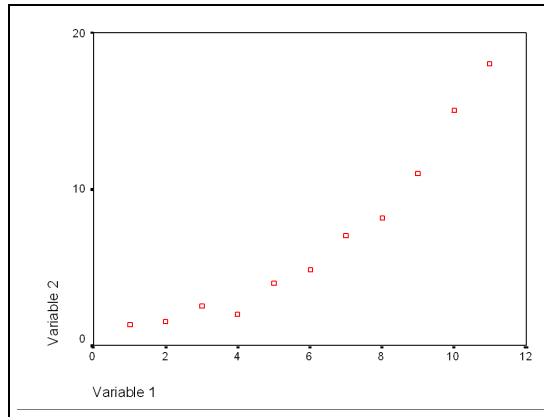


Abbildung 5.3.16: Streudiagramm der Beispieldaten Tabelle 5.3.13 (ohne Transformation)

Wir versuchen es mit Logarithmieren und erhalten (s. Abbildung 5.3.17):

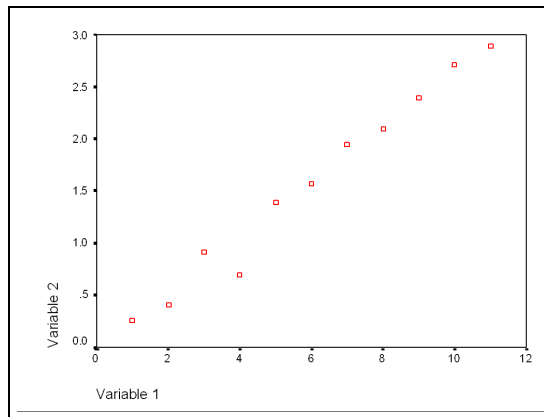


Abbildung 5.3.17: Streudiagramm der Beispieldaten Tabelle 5.3.13 (logarithmische Transformation)

Es liegt nun ein deutlich linearer Zusammenhang vor. ◇

Definition 5.3.8. Kovarianz: $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$

Die Kovarianz wird zur Varianz, wenn wir die Definition auf nur eine Variable anwenden:

$$s_{xx} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = s^2$$

Mit Hilfe der Kovarianzdefinition lässt sich der Korrelationskoeffizient wie folgt schreiben:

$$r_{xy} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} = \frac{s_{xy}}{s_x s_y}$$

5.3.6 Die Spearman-Rangkorrelation

Sind die Voraussetzungen für eine sinnvolle Interpretation des Pearson-Korrelationskoeffizienten nicht gegeben, so verwenden wir die Spearman-Rangkorrelation (Charles Edward Spearman, 1863 -

1945, britischer Psychologe). Sie ist robust (gegen Ausreisser), setzt weder Symmetrie noch lineare Tendenz voraus und kann auch auf ordinalskalierte Daten angewendet werden. Die Berechnung besteht in der Anwendung der Pearson-Korrelation auf die Ränge statt auf die Daten. Bei gleichen Ausprägungen ordnen wir jeweils das arithmetische Mittel der Ränge zu - wir würden z.B. im Beispiel unten die Ränge 6, 7.5, 7.5, 9 erhalten - d.h. 7.5 und 7.5 nehmen die Plätze von 7 und 8 ein. Allgemein kann man den Rang r_i des i -ten Datums wie folgt festlegen:

$$r_i = |\{x_j \mid x_j < x_i\}| + \frac{1}{2} (1 + |\{x_j \mid x_j = x_i\}|)$$

Beispiel 5.3.9. Die Daten von Beispiel 5.2.1 genügen der Symmetriebedingung nicht. Wir müssten sie transformieren oder die Spearman-Korrelation verwenden (s. Tabelle 5.3.14):

Objekt	Körpergrösse	Gewicht	Ränge Grösse	Ränge Gewicht
9	50	4	1	1
2	100	25	2	2
4	130	35	3	3
5	165	69	4	4
10	175	70	6	5
11	174	72	5	6
6	178	75	7.5	7
7	181	77	9	8
1	187	80	10	9
8	178	85	7.5	10
3	198	86	11	11

Tabelle 5.3.14: Beispiel für die Berechnung der Spearman-Korrelation

Wir erhalten: $r_S = 0.952166479$

◇

Entscheidungsschema (wenn die Bedingungen für die Anwendung der Pearson-Korrelation nicht erfüllt sind, denkt man als nächstes an die Spearman-Korrelation und überprüft, ob die Anwendungsbedingungen für diese erfüllt sind):

Pearson-Korrelation, wenn alle folgenden Bedingungen erfüllt sind:	Spearman-Korrelation, wenn alle folgenden Bedingungen erfüllt sind
· Beide Variablen metrisch skaliert · Punktsymmetrie der Punktwolke zu $(\bar{x}, \bar{y})$ · Punktwolke liegt ungefähr um Gerade herum · Keine Ausreisser	· Beide Variablen mindestens ordinalskaliert · Nicht zu viele Bindungen (Daten differenziert)

Daten sind differenziert, wenn es (fast) so viele Ausprägungen wie Daten hat. Wenn die Daten zu wenig differenziert sind, verwendet man Gamma.

Interpretation der Korrelation

Selbst bei starker Korrelation kann nicht geschlossen werden, dass ein Ursache-Wirkungs-Zusammenhang zwischen zwei Variablen besteht. Die Ausprägungen von X können durch die Ausprägungen von Y bestimmt sein, oder umgekehrt. Es kann auch eine Wechselwirkung bestehen oder die Ausprägungen jeder Variablen werden durch die einer dritten Variable bestimmt (Der Ursache-Wirkungs-Begriff ist allerdings so oder so unklar). Ein Beispiel für die Bestimmung durch eine dritte Variable: Bei 6- bis 10-jährigen Schulkindern wurde eine deutlich positive Korrelation zwischen Körpergrösse und manueller Geschicklichkeit festgestellt. Die dritte Variable ist das Alter

(je älter, desto grösser, je älter, desto geschickter). Es ist somit weder das Grösse, das die Geschicklichkeit „verursacht“, noch die Geschicklichkeit, die die Grösse verursacht. Besonders gefährlich ist der Fehlschluss von einer Korrelation auf eine Ursache-Wirkungs-Relation bei Zeitreihen. So beweist z.B. eine Korrelation zwischen einer Zunahme der Fernsehapparate und der Verbrechen über die Jahre hinweg nichts - jedenfalls nicht bezüglich des Einflusses von Fernsehen auf die Verbrechensrate. Diese Überlegungen führen unter Anderem zum Verfahren der partiellen Korrelation (s. S. 123).

5.3.7 Lineare Regression

Einen nahen Zusammenhang mit der Pearson-Korrelation weist die *lineare Regression* auf. Es handelt sich wie bei der Korrelation um ein Verfahren, um den affinen Zusammenhang zwischen zwei statistischen Variablen zu charakterisieren. Entsprechend sind für die Anwendung des Verfahrens die Bedingungen der Pearson-Korrelation zu beachten (s. S. 110). Betrachten wir die folgende Punktwolke (s. Abbildung 5.3.18), die den Zusammenhang zwischen Einkommen und Ausgaben für Schmuck vor Augen führt, drängt sich die Idee auf, den Zusammenhang der beiden Variablen durch eine Gerade auszudrücken (s. Abbildung 5.3.18).

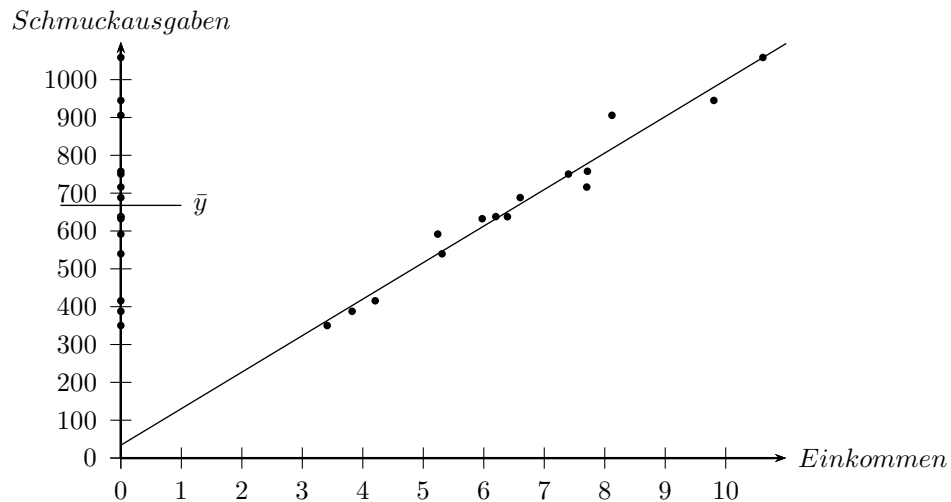


Abbildung 5.3.18: Zweidimensionales Streudiagramm mit einer möglichen Gerade für die Charakterisierung des Zusammenhangs der Variablen Einkommen und Ausgaben für Schmuck. Auf der y-Achse sind die Ausgaben für Schmuck samt deren Mittelwert $\bar{y}$ abgetragen.

Der Vorteil der Berechnung einer solchen Gerade ergibt sich beim Schätzen der mittleren Ausgaben für Schmuck. Kennen wir den Zusammenhang zwischen Einkommen und Ausgaben für Schmuck nicht, und will man vermuten, wieviel eine Person für Schmuck ausgibt, so schätzt man diese Ausgabe am besten mit dem arithmetischen Mittel der Schmuckdaten. Man liegt dann im Mittel am wenigsten neben den tatsächlichen Schmuckausgaben der Person. Wie gezeigt (s. S. 47) ist nämlich das arithmetische Mittel der Wert, bei dem die Summe der quadrierten Abweichungen minimal ist. Der mittlere Fehler (z.B. ausgedrückt durch die Varianz oder die Standardabweichung) wird bei einer solchen Schätzung aber ziemlich gross sein. Kennen wir den Zusammenhang zwischen Einkommen und Ausgaben für Schmuck durch eine Gerade $f(x) = ax + b$, so schätzen wir die Schmuckausgaben einer Person, die x Einkommen hat, durch $ax + b$. Je nach Einkommen erhalten wir andere Schätzungen für die Schmuckausgaben. Wie man an der obigen Graphik sieht, ist die Streuung um die Gerade herum viel kleiner als die Streuung der Schmuckdaten um ihr arithmetisches Mittel $\bar{y}$. Die Verwendung der Geraden für die Schätzung führt also zu einem kleineren mittleren Schätzfehler.

Der Vorteil der Berechnung einer solchen Gerade kann auch wie folgt erläutert werden. Bei einem Spiel ist eine Funktion so zu wählen, dass man möglichst wenig Strafe zahlen muss. Die

Strafe ist die Summe der quadrierten Abweichungen der Daten von der gewählten Funktion. Kennt man nur die Ausgaben für Schmuck, so muss man das Arithmetische Mittel der Schmuckausgaben wählen, um möglichst wenig Strafe zu zahlen (= konstante Funktion). Hat man aber auch die Einkommensdaten zur Verfügung, so wählt man als Funktion die Regressionsgerade und kann so die Summe der quadrierten Abweichungen minimieren.

Bei der Regression ist im Gegensatz zur Korrelation zu beachten, dass es einen Unterschied macht, welche Variable man als abhängige und welche als unabhängige betrachtet. Die unabhängige Variable, die mit x bezeichnet wird, ist jene, von der man auf Grund inhaltlicher Überlegungen davon ausgehen kann, dass sie die andere Variable, die mit y bezeichnet wird, beeinflusst. So beeinflusst z.B die Ausbildung von Personen tendenzmässig das Einkommen dieser Personen und nicht umgekehrt. Die Temperatur beeinflusst das Wachstum von Pflanzen und nicht umgekehrt. Das Gefälle und die Wassermenge in einem Wasserkraftwerk beeinflussen die Stromproduktion des Werkes und nicht umgekehrt. Das Einkommen einer Person beeinflusst den Kauf von Konsumgütern durch diese Person - und nicht umgekehrt.

Beispiel 5.3.10. Die obige Abbildung 5.3.18 ist durch die Daten der Tabelle 5.3.15 gegeben:

Einkommen (x_i) in 1000 Fr. pro Monat	Ausgaben (y_i) für Schmuck in Fr. pro Jahr	Geschätzte Werte $f(x_i) = 96.46x_i + 34$	Residuen $y_i - f(x_i)$
6.199	638.0	632.14151	5.85849
7.400	750.4	748.02600	2.37400
5.974	632.5	610.43126	22.06874
7.714	757.8	778.32386	-20.52386
6.602	688.3	671.02698	17.27302
5.310	539.7	546.36190	-6.66190
3.823	387.8	402.88127	-15.08127
8.119	905.7	817.40231	88.29769
5.239	591.8	539.51111	52.28889
10.61	1058.6	1058.24135	0.35865
9.804	945.0	979.98796	-34.98796
3.411	350.2	363.12739	-12.92739
6.391	637.7	650.66759	-12.96759
4.205	415.6	439.74045	-24.14045
7.702	716.1	777.16598	-61.06598
Mittelwert	667.68	667.669128	0
Standardabweichung	202.1203538		35.77747102

Tabelle 5.3.15: Beispiel für die Grundideen der Regression: Erhöhung der Präzision der Schätzung durch Berücksichtigung des Zusammenhangs zwischen Einkommen und Ausgaben für Schmuck

Die eingezeichnete Gerade ist gegeben durch

$$f(x) = 96.46x + 34.$$

Die Standardabweichung der der Schmuckdaten beträgt 202.120. Die Abweichung eines Schmuckdatums von seiner Schätzung ist

$$y_i - f(x_i).$$

Diese Abweichungen werden „Residuum“ (Mehrzahl „Residuen“) genannt. Die Standardabweichung der Residuen beträgt 35.77747102 - ist also viel kleiner als die Standardabweichung der Schmuckdaten. $\diamond$

Berechnung der optimalen Gerade

Offenbar ist nicht jede Gerade gleich geeignet, um den Zusammenhang der Variablen auszudrücken. Man versucht vielmehr, durch die Punktwolke der (x_i, y_i) eine *optimale* Gerade zu berechnen. Dabei kann „optimal“ unterschiedlich charakterisiert werden. Hier definieren wir als optimal die Gerade, bei welcher die Summe der quadrierten Differenzen

$$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2$$

minimal ist. Je nach der Festlegung von f ergeben sich nämlich unterschiedliche Summen quadrierter Abweichungen, wie folgendes Beispiel zeigt:

Beispiel 5.3.11. Sei der Datensatz $(1, 5), (2, 5.5)$ und $(3, 6.5)$ gegeben. Wir betrachten die Gerade $f_1(x) = 2x + 1$. Damit ergibt sich

$$\begin{aligned} \sum_{i=1}^3 (y_i - f_1(x_i))^2 &= (5 - (2 \cdot 1 + 1))^2 + (5.5 - (2 \cdot 2 + 1))^2 + (6.5 - (2 \cdot 3 + 1))^2 \\ &= 4.5 \end{aligned}$$

Für $f_2(x) = 3x + 2$ ergibt sich hingegen

$$\begin{aligned} \sum_{i=1}^3 (y_i - f_2(x_i))^2 &= (5 - (3 \cdot 1 + 2))^2 + (5.5 - (3 \cdot 2 + 2))^2 + (6.5 - (3 \cdot 3 + 2))^2 \\ &= 26.5 \end{aligned}$$

und für $f_3(x) = 0.75x + 4 + \frac{1}{6}$

$$\begin{aligned} \sum_{i=1}^3 (y_i - f_3(x_i))^2 &= \left(5 - \left(0.75 \cdot 1 + 4 + \frac{1}{6}\right)\right)^2 + \left(5.5 - \left(0.75 \cdot 2 + 4 + \frac{1}{6}\right)\right)^2 \\ &\quad + \left(6.5 - \left(0.75 \cdot 3 + 4 + \frac{1}{6}\right)\right)^2 \\ &= 0.041667 \end{aligned}$$

Von diesen drei Funktionen wäre damit $f_3(x) = 0.75x + 4 + \frac{1}{6}$ die beste. Es geht nun darum, die beste unter allen möglichen affinen Funktionen $\mathbb{R} \rightarrow \mathbb{R}$ zu finden. $\diamond$

$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2$ beschreibt eine Abbildung von der Menge aller affinen Funktionen $\mathcal{F} := \{f | f(x) = ax + b\}$ in die reellen Zahlen :

$$S : \mathcal{F} \rightarrow \mathbb{R}$$

$$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2$$

Um das Minimum dieser Funktion zu finden kann man den Umstand nutzen, dass jedem f ein geordnetes Paar von Koeffizienten (a, b) entspricht. Es gilt ja $f(x) = ax + b$. Man kann statt $S(f)$ also auch die folgende Funktion betrachten

$$S : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$S(a, b) := \sum_{i=1}^n (y_i - f(x_i))^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

Um den Satz bezüglich des Minimums dieser Funktion beweisen, noch ein paar Vorarbeiten:

Satz 5.3.12. Wenn es für einen Vektor $x = (x_1, \dots, x_n)$ mindestens ein Paar von Komponenten $x_i \neq x_j$ gibt, so ist

$$\sum_{i=1}^n x_i^2 - n\bar{x}^2 > 0.$$

Beweis.

$$\begin{aligned} \sum_{i=1}^n x_i^2 - n\bar{x}^2 &= \sum_{i=1}^n (x_i^2 - \bar{x}^2) \\ &= \sum_{i=1}^n (x_i - \bar{x})(x_i + \bar{x}) \end{aligned}$$

Für $d_i := x_i - \bar{x}$ gilt $\sum_{i=1}^n d_i = \sum_{i=1}^n (x_i - \bar{x}) = 0$ (Summe der Abweichungen der Daten von ihrem Mittelwert ist 0), sowie $x_i = d_i + \bar{x}$, womit gilt $x_i + \bar{x} = d_i + \bar{x} + \bar{x} = d_i + 2\bar{x}$. Damit gilt:

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(x_i + \bar{x}) &= \sum_{i=1}^n d_i (d_i + 2\bar{x}) \\ &= \sum_{i=1}^n (d_i^2 + d_i 2\bar{x}) \\ &= \sum_{i=1}^n d_i^2 + \sum_{i=1}^n d_i 2\bar{x} \\ &= \sum_{i=1}^n d_i^2 + 2\bar{x} \sum_{i=1}^n d_i \\ &= \sum_{i=1}^n d_i^2 + 2\bar{x} \cdot 0 \\ &= \sum_{i=1}^n d_i^2 > 0, \end{aligned}$$

wenn mindestens ein $d_i \neq 0$. Zudem gilt offensichtlich: $d_i \neq 0$ genau dann, wenn es mindestens zwei Daten x_i, x_j gibt, die verschieden sind. $\square$

Satz 5.3.13. Eine Funktion $f(x) = h(x - y)^2$ ($h, x, y \in \mathbb{R}$) nimmt in $x = y$ ein Minimum an, wenn $h > 0$, und ein Maximum, wenn $h < 0$.

Beweis. Ist $x - y \neq 0$, ist $(x - y)^2 > 0$. Wird $(x - y)^2$ mit einer positiven Zahl multipliziert, steigt damit $h(x - y)^2$ mit steigendem Betrag der Differenz von x und y . Entsprechend ist $h(x - y)^2$ minimal, wenn der Betrag der Differenz minimal ist, d.h. wenn $x - y = 0$ und damit $x = y$. Wird $(x - y)^2$ mit einer negativen Zahl multipliziert, so sinkt $h(x - y)^2$ mit steigendem Betrag der Differenz, d.h. $h(x - y)^2$ wird maximal, wenn $x = y$. $\square$

Nun kann zur Berechnung der Minimalstellen von $S(a, b)$ übergegangen werden:

Satz 5.3.14. Für $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ mit mindestens einem Datenpaar $x_i \neq x_j$, sowie $f(x) = ax + b$ ist

$$\begin{aligned} S &: \mathbb{R}^2 \rightarrow \mathbb{R} \\ S(a, b) &:= \sum_{i=1}^n (y_i - f(x_i))^2 = \sum_{i=1}^n (y_i - (ax_i + b))^2 \end{aligned}$$

minimal genau dann, wenn

$$a = \frac{\sum_{i=1}^n x_i y_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

$$b = \bar{y} - a \bar{x}$$

Beweis.

$$\begin{aligned} S(a, b) &= \sum_{i=1}^n (y_i - ax_i - b)^2 \\ &= \sum_{i=1}^n (2abx_i - 2by_i - 2ax_i y_i + b^2 + y_i^2 + a^2 x_i^2) \\ &= 2ab \sum_{i=1}^n x_i - 2b \sum_{i=1}^n y_i - 2a \sum_{i=1}^n x_i y_i + nb^2 + \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 \end{aligned}$$

Es ist z.B. $\sum_{i=1}^n x_i = \frac{n}{n} \sum_{i=1}^n x_i = n \bar{x}$, womit wir erhalten:

$$\begin{aligned} S(a, b) &= 2abn\bar{x} - 2bn\bar{y} - 2a \sum_{i=1}^n x_i y_i + nb^2 + \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 \\ &= n2ab\bar{x} - n2b\bar{y} + nb^2 - 2a \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 \end{aligned}$$

Ferner gilt - da $n(\bar{y} - a\bar{x})^2 - n(\bar{y} - a\bar{x})^2 = 0$ -

$$\begin{aligned} n2ab\bar{x} - n2b\bar{y} + nb^2 &= n2ab\bar{x} - n2b\bar{y} + nb^2 + n(\bar{y} - a\bar{x})^2 - n(\bar{y} - a\bar{x})^2 \\ &= n2b(a\bar{x} - \bar{y}) + nb^2 + n(\bar{y} - a\bar{x})^2 - n(\bar{y} - a\bar{x})^2 \\ &= -n2b(\bar{y} - a\bar{x}) + nb^2 + n(\bar{y} - a\bar{x})^2 - n(\bar{y} - a\bar{x})^2 \\ &= n(b^2 - 2b(\bar{y} - a\bar{x}) + (\bar{y} - a\bar{x})^2) - n(\bar{y} - a\bar{x})^2 \\ &= n(b^2 - 2b(\bar{y} - a\bar{x}) + (\bar{y} - a\bar{x})^2) - n\bar{y}^2 + 2n\bar{y}a\bar{x} - na^2\bar{x}^2 \\ &= n(b - (\bar{y} - a\bar{x}))^2 - n\bar{y}^2 + 2n\bar{y}a\bar{x} - na^2\bar{x}^2 \end{aligned}$$

Damit erhält man durch Einsetzen und Umstellen

$$\begin{aligned} S(a, b) &= n(b - (\bar{y} - a\bar{x}))^2 - n\bar{y}^2 + 2n\bar{y}a\bar{x} - na^2\bar{x}^2 - 2a \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 + a^2 \sum_{i=1}^n x_i^2 \\ &= \sum_{i=1}^n y_i^2 + n(b - (\bar{y} - a\bar{x}))^2 - n\bar{y}^2 + 2n\bar{y}a\bar{x} - na^2\bar{x}^2 - 2a \sum_{i=1}^n x_i y_i + a^2 \sum_{i=1}^n x_i^2. \end{aligned}$$

Es gilt

$$\begin{aligned}
 2n\bar{y}a\bar{x} - na^2\bar{x}^2 - 2a \sum_{i=1}^n x_i y_i + a^2 \sum_{i=1}^n x_i^2 &= 2n\bar{y}a\bar{x} - 2a \sum_{i=1}^n x_i y_i + a^2 \sum_{i=1}^n x_i^2 - na^2\bar{x}^2 \\
 &= 2n\bar{y}a\bar{x} - 2a \sum_{i=1}^n x_i y_i + a^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\
 &= 2a \left(n\bar{y}\bar{x} - \sum_{i=1}^n x_i y_i \right) + a^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \\
 &= -2a \left(\sum_{i=1}^n x_i y_i - n\bar{y}\bar{x} \right) + a^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right).
 \end{aligned}$$

Mit $c := \sum_{i=1}^n x_i y_i - n\bar{y}\bar{x}$, $d := \sum_{i=1}^n x_i^2 - n\bar{x}^2$ und $\frac{c^2}{d} - \frac{c^2}{d} = 0$ (gemäss obigem Satz 5.3.12 ist $d \neq 0$) gilt damit:

$$\begin{aligned}
 -2a \left(\sum_{i=1}^n x_i y_i - n\bar{y}\bar{x} \right) + a^2 \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) &= -2ac + a^2 d + \frac{c^2}{d} - \frac{c^2}{d} \\
 &= d \left(-2\frac{c}{d}a + a^2 + \frac{c^2}{d^2} \right) - \frac{c^2}{d} \\
 &= d \left(a - \frac{c}{d} \right)^2 - \frac{c^2}{d}
 \end{aligned}$$

Insgesamt erhält man

$$S(a, b) = \sum_{i=1}^n y_i^2 + n(b - (\bar{y} - a\bar{x}))^2 - n\bar{y}^2 + d \left(a - \frac{c}{d} \right)^2 - \frac{c^2}{d}$$

In dieser Formulierung von S kommen in

$$\begin{aligned}
 g &:= \sum_{i=1}^n y_i^2 - n\bar{y}^2 - \frac{c^2}{d} \\
 &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 - \frac{\left(\sum_{i=1}^n x_i y_i - n\bar{y}\bar{x} \right)^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}
 \end{aligned}$$

die Koeffizienten a und b nicht vor. Entsprechend ist g eine Konstante und hat keinen Einfluss auf die Minima oder Maxima von S - g bewirkt nur eine Verschiebung aller Funktionswerte von

$$\begin{aligned}
 S'(a, b) &:= S(a, b) - g \\
 &= n(b - (\bar{y} - a\bar{x}))^2 + d \left(a - \frac{c}{d} \right)^2
 \end{aligned}$$

um g noch oben oder nach unten. Also nimmt S' an derselben Stelle Maxima oder Minima an wie S .

Wir wenden den Satz 5.3.13 auf S' an, einer Funktion, die aus zwei Summanden der Form $h(x - y)^2$ besteht. Eine Summe von Funktionen wird an einer Stelle minimal, an der beide Summanden minimal werden. Es gilt $n > 0$ und wie oben festgehalten $d > 0$. Damit nimmt S' ein Minimum

an, und zwar bei

$$b = \bar{y} - a\bar{x}$$

$$a = \frac{c}{d} = \frac{\sum_{i=1}^n x_i y_i - n\bar{y}\bar{x}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

□

Beispiel 5.3.15. Die Berechnungsformeln sollen nun für die Berechnung der optimalen Gerade durch die Daten der Tabelle 5.3.15 verwendet werden:

Einkommen x_i	Ausgaben für Schmuck y_i	$x_i y_i$	x_i^2
6.199	638	3954.962	38.427601
7.4	750.4	5552.96	54.76
5.974	632.5	3778.555	35.688676
7.714	757.8	5845.6692	59.505796
6.602	688.3	4544.1566	43.586404
5.31	539.7	2865.807	28.1961
3.823	387.8	1482.5594	14.615329
8.119	905.7	7353.3783	65.918161
5.239	591.8	3100.4402	27.447121
10.615	1058.6	11237.039	112.678225
9.804	945	9264.78	96.118416
3.411	350.2	1194.5322	11.634921
6.391	637.7	4075.5407	40.844881
4.205	415.6	1747.598	17.682025
7.702	716.1	5515.4022	59.320804
Summen		71513.3798	706.42446
6.5672	667.68	Mittelwerte	

Tabelle 5.3.16: Berechnung der Koeffizienten fürs Beispiel

Mit $\bar{x} = 6.5672$, $\bar{y} = 667.68$ und $n = 15$ ergibt sich

$$a = \frac{71513.3798 - 15 \cdot 667.68 \cdot 6.5672}{706.42446 - 15 \cdot 6.5672^2} = 96.49236419$$

und

$$b = 667.68 - 96.49236419 \cdot 6.5672 = 33.9953459$$

◇

Zusammenhang mit der Pearson-Korrelation

Es gibt einen unmittelbaren Zusammenhang zwischen der Steigung der Regressionsgerade a und der Pearson-Korrelation r . Es gilt nämlich mit der Standardabweichung s_y der y -Daten und der Standardabweichung s_x der x -Daten:

Satz 5.3.16. $a = r \frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2}$

Beweis. Die Formel für die Steigung der Geraden kann wie folgt umgewandelt werden:

$$a = \frac{\sum_{i=1}^n x_i y_i - n \bar{y} \bar{x}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{y} \frac{1}{n} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - n \bar{x} \frac{1}{n} \sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i} = \frac{\sum_{i=1}^n (x_i y_i - x_i \bar{y})}{\sum_{i=1}^n (x_i x_i - x_i \bar{x})}. \quad (5.3.1)$$

Nun ist

$$\sum_{i=1}^n (x_i - \bar{x}) = 0.$$

Wir können somit (5.3.1) ergänzen mit

$$\begin{aligned} -\bar{y} \sum_{i=1}^n (x_i - \bar{x}) &= 0 \\ -\bar{x} \sum_{i=1}^n (x_i - \bar{x}) &= 0 \end{aligned}$$

und erhalten:

$$\begin{aligned} a &= \frac{\sum_{i=1}^n (x_i y_i - x_i \bar{y}) - \bar{y} \left(\sum_{i=1}^n (x_i - \bar{x}) \right)}{\sum_{i=1}^n (x_i x_i - x_i \bar{x}) - \bar{x} \left(\sum_{i=1}^n (x_i - \bar{x}) \right)} = \frac{\sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{y} x_i + \bar{y} \bar{x})}{\sum_{i=1}^n (x_i x_i - x_i \bar{x} - \bar{x} x_i + \bar{x}^2)} \\ &= \frac{\sum_{i=1}^n (x_i y_i - 2x_i \bar{y} + \bar{y} \bar{x})}{\sum_{i=1}^n (x_i x_i - 2x_i \bar{x} + \bar{x}^2)} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_x^2} \end{aligned}$$

mit der Kovarianz $s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$.

Durch Erweitern erhält man

$$a = \frac{s_{xy}}{s_x^2} = \frac{s_{xy} s_y}{s_x s_x s_y} = r \frac{s_y}{s_x}.$$

□

$a = r \frac{s_y}{s_x}$ wird grösser, wenn die Standardabweichung der Daten der Variable Y grösser wird oder die Standardabweichung der Daten der Variable X kleiner wird. Dies leuchtet ein: je grösser die Streuung der Daten auf der x -Achse, desto kleiner wird die Steigung der Geraden sein. Je grösser die Streuung auf der y -Achse, desto grösser wird die Steigung der Geraden sein.

Additivität der Quadratsummen und Bestimmtheitsmass

Wir definieren:

- die Summe der quadrierten Residuen oder der nicht erklärten Abweichungen

$$SQR := \sum_{i=1}^n (y_i - f(x_i))^2.$$

SQR lässt sich aus den Residuen mittels der Varianz der Residuen

$$s_R^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - f(x_i))^2$$

berechnen. Aus der Formel folgt unmittelbar durch Multiplikation

$$SQR = (n-1) s_R^2.$$

- die Summe der quadrierten Abweichungen der vorausgesagten Werte vom Mittelwert der y -Daten (erklärten Abweichungen vom Mittelwert)

$$SQE := \sum_{i=1}^n (f(x_i) - \bar{y})^2.$$

- die Summen der der quadrierten Abweichungen der y -Daten vom Mittelwert

$$SQT := \sum_{i=1}^n (y_i - \bar{y})^2.$$

Diese Summe kann direkt aus der mit $\frac{1}{n-1}$ berechneten Varianz s_T^2 der y -Daten berechnet werden, da

$$s_T^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$$

und damit

$$SQT = (n-1) s_T^2.$$

Siehe für die Veranschaulichung der verwendeten Differenzen die Abbildung [5.3.19](#).

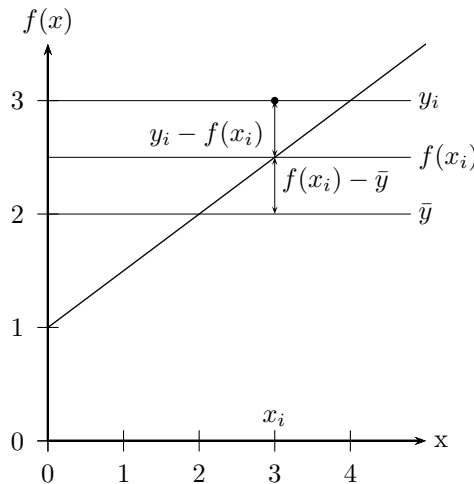


Abbildung 5.3.19: Regressionsgerade und Punkt mit den Abständen, die in die Berechnung der Quadratsummen einfließen

Es gilt nun:

Satz 5.3.17. $SQT = SQR + SQE$

Beweis. Es gilt: $y_i - \bar{y} = y_i - f(x_i) + f(x_i) - \bar{y}$
Wir quadrieren beide Seiten der Gleichung:

$(y_i - \bar{y})^2 = (y_i - f(x_i) + f(x_i) - \bar{y})^2$ und summieren auf beiden Seiten über alle i (d.h. wir addieren n Gleichungen!):

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n [(y_i - f(x_i) + (f(x_i) - \bar{y}))^2 = \\ \sum_{i=1}^n [(y_i - f(x_i))^2 + 2(y_i - f(x_i))(f(x_i) - \bar{y}) + (f(x_i) - \bar{y})^2] &= \\ \sum_{i=1}^n (y_i - f(x_i))^2 + 2 \sum_{i=1}^n [(y_i - f(x_i))(f(x_i) - \bar{y})] + \sum_{i=1}^n (f(x_i) - \bar{y})^2 \end{aligned}$$

Nun gilt: $2 \sum_{i=1}^n [(y_i - f(x_i))(f(x_i) - \bar{y})] = 0$, da mit $(y_i - f(x_i)) = e_i$ gilt:

$$\begin{aligned} 2 \sum_{i=1}^n [(y_i - f(x_i))(f(x_i) - \bar{y})] &= 2 \sum_{i=1}^n [e_i(f(x_i) - \bar{y})] = \\ 2 \sum_{i=1}^n [e_i f(x_i) - e_i \bar{y}] &= 2 \sum_{i=1}^n e_i f(x_i) - 2 \sum_{i=1}^n e_i \bar{y} = 2 \sum_{i=1}^n e_i f(x_i) - 2 \bar{y} \sum_{i=1}^n e_i = \\ 2 \sum_{i=1}^n e_i f(x_i) - 2 \bar{y} \cdot 0 &= 2 \sum_{i=1}^n e_i f(x_i) = 2 \sum_{i=1}^n e_i (ax_i + b) = 2 \sum_{i=1}^n e_i ax_i + 2 \sum_{i=1}^n e_i b = \\ 2a \sum_{i=1}^n e_i x_i + 2b \sum_{i=1}^n e_i &= 2a \cdot 0 + 2b \cdot 0 = 0 \end{aligned}$$

□

Wegen der Additivität lässt sich die durch die Gerade „erklärte“ Streuung SQE sinnvoll in % der Gesamtstreuung SQT ausdrücken: $\frac{SQE}{SQT} \cdot 100\%$. Man sagt, die Regressionsgerade „erkläre“ $\frac{SQE}{SQT} \cdot 100\%$ der Gesamtstreuung. Die Grösse

$$R^2 := \frac{SQE}{SQT}$$

wird „lineares Bestimmtheitsmass“ genannt. R^2 ist gross, wenn SQE , die durch die Regression erklärte Streuung, relativ zu SQT gross ist.

R^2 ist eine Zahl zwischen 0 und 1, denn SQT , SQR und $SQE \geq 0$. Damit ist $\frac{SQE}{SQT} \geq 0$. Aus $SQT = SQR + SQE$ folgt: $SQE \leq SQT$. Somit ist $\frac{SQE}{SQT} \leq 1$.

Man kann zeigen, dass das Quadrat des Pearson-Korrelationskoeffizienten mit dem linearen Bestimmtheitsmass übereinstimmt:

Satz 5.3.18. *Das Bestimmtheitsmass ist mit der quadrierten Pearsonkorrelation identisch, d.h.*

$$\frac{SQE}{SQT} = r_{xy}^2$$

Beweis. Es gilt:

$$a = \frac{s_{xy}}{s_x^2} \implies s_{xy} = as_x^2 \quad (5.3.2)$$

(s. Satz 5.3.16 Seite 117) und mit demselben Satz

$$r = \frac{s_{xy}}{s_x s_y} = \frac{as_x^2}{s_x s_y} = a \frac{s_x}{s_y}$$

Zudem gilt - da $SQT = SQR + SQE$, s. Satz 5.3.17:

$$\frac{SQE}{SQT} = \frac{SQT - SQR}{SQT}$$

Es ist

$$SQR = \sum (y_i - ax_i - b)^2 = \sum (y_i^2 - 2ax_i y_i - 2by_i + 2abx_i + a^2 x_i^2 + b^2).$$

Ersetzt man b durch $\bar{y} - a\bar{x}$, ergibt sich:

$$SQR = 2a(\bar{y} - a\bar{x}) \sum x_i - 2(\bar{y} - a\bar{x}) \sum y_i - 2a \sum x_i y_i + n(\bar{y} - a\bar{x})^2 + \sum y_i^2 + a^2 \sum x_i^2$$

Ersetzt man $n\bar{x}$ durch $\sum x_i$ und $n\bar{y}$ durch $\sum y_i$, erhält man:

$$SQR = 2a(\bar{y} - a\bar{x})n\bar{x} - 2(\bar{y} - a\bar{x})n\bar{y} - 2a\sum x_i y_i + n(\bar{y} - a\bar{x})^2 + \sum y_i^2 + a^2 \sum x_i^2$$

Nun gilt

$$\begin{aligned} 2a(\bar{y} - a\bar{x})n\bar{x} - 2(\bar{y} - a\bar{x})n\bar{y} + n(\bar{y} - a\bar{x})^2 &= 2a\bar{y}n\bar{x} - 2a^2\bar{x}^2n - 2\bar{y}^2n + 2a\bar{x}n\bar{y} + \bar{y}^2n - 2a\bar{y}n\bar{x} + a^2\bar{x}^2n \\ &= 2an\bar{x}\bar{y} - n\bar{y}^2 - a^2n\bar{x}^2, \end{aligned}$$

und man erhält:

$$SQR = 2an\bar{x}\bar{y} - n\bar{y}^2 - a^2n\bar{x}^2 - 2a\sum x_i y_i + \sum y_i^2 + a^2 \sum x_i^2$$

Zudem ist:

$$\begin{aligned} \sum y_i^2 - n\bar{y}^2 &= \sum y_i^2 - 2n\bar{y}^2 + n\bar{y}^2 \\ &= \sum y_i^2 - 2n\bar{y}\frac{1}{n}\sum y_i + n\bar{y}^2 \\ &= \sum (y_i^2 - 2\bar{y}y_i + \bar{y}^2) \\ &= \sum (y_i - \bar{y})^2 = SQT. \end{aligned}$$

Analog gilt:

$$\begin{aligned} a^2 \sum x_i^2 - a^2n\bar{x}^2 &= a^2 \left(\sum x_i^2 - n\bar{x}^2 \right) \\ &= a^2 \left(\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \right) \\ &= a^2 \left(\sum x_i^2 - 2n\bar{x}\frac{1}{n}\sum x_i + n\bar{x}^2 \right) \\ &= a^2 \left(\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2 \right) \\ &= a^2 \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= a^2 \sum (x_i - \bar{x})^2 \\ &= a^2 s_x^2 (n-1) \end{aligned}$$

denn $\sum (x_i - \bar{x})^2 = s_x^2 (n-1)$, da $s_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$.

Schliesslich ist

$$\begin{aligned} 2an\bar{x}\bar{y} - 2a\sum x_i y_i &= -2a \left(\sum x_i y_i - n\bar{x}\bar{y} \right) \\ &= -2as_{xy} (n-1) \\ &= -2a^2 s_x^2 (n-1), \end{aligned}$$

denn es gilt erstens

$$\begin{aligned} s_{xy} (n-1) &= \sum (x_i - \bar{x})(y_i - \bar{y}) \\ &= \sum (x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x}\bar{y}) \\ &= \sum x_i y_i - \sum x_i \bar{y} - \sum y_i \bar{x} + n\bar{x}\bar{y} \\ &= \sum x_i y_i - n\bar{x}\bar{y} - n\bar{y}\bar{x} + n\bar{x}\bar{y} \\ &= \sum x_i y_i - n\bar{x}\bar{y}, \end{aligned}$$

und zweitens mit (5.3.2) $s_{xy} = as_x^2$. Damit ist dann:

$$SQR = SQT - a^2 s_x^2 (n-1)$$

und somit

$$a^2 = \frac{SQT - SQR}{s_x^2 (n-1)},$$

woraus schliesslich mit $SQT = (n-1) s_y^2$ folgt:

$$r^2 = a^2 \frac{s_x^2}{s_y^2} = \frac{SQT - SQR}{s_x^2 (n-1)} \frac{s_x^2}{s_y^2} = \frac{SQT - SQR}{(n-1) s_y^2} = \frac{SQT - SQR}{SQT} = \frac{SQE}{SQT}$$

□

Beispiel 5.3.19. Wir berechnen das Bestimmtheitsmass der Daten der Tabelle 5.3.15.

Einkommen x_i	Ausgaben für Schmuck y_i	$f(x_i) = 96.4923642x_i + 33.9953459$	$y_i - f(x_i)$
6.199	638	632.1515115	5.848488494
7.4	750.4	748.0388409	2.361159104
5.974	632.5	610.4407296	22.05927044
7.714	757.8	778.3374433	-20.53744325
6.602	688.3	671.0379343	17.26206573
5.31	539.7	546.3697997	-6.669799743
3.823	387.8	402.8856542	-15.0856542
8.119	905.7	817.4168507	88.28314925
5.239	591.8	539.5188419	52.28115811
10.615	1058.6	1058.261792	0.33820824
9.804	945	980.0064844	-35.0064844
3.411	350.2	363.1308001	-12.93080015
6.391	637.7	650.6780454	-12.97804543
4.205	415.6	439.7457373	-24.14573732
7.702	716.1	777.1795349	-61.07953488
Varianzen	40852.63743		1280.027409

Tabelle 5.3.17: Berechnung des Bestimmtheitsmasses fürs Beispiel

Damit gilt

$$SQR = 14 \cdot 1280.027409 = 17920.38372$$

$$SQT = 14 \cdot 40852.63743 = 571936.924$$

$$SQE = SQT - SQR = 571936.924 - 17920.38372 = 554016.5403$$

$$R^2 = \frac{SQE}{SQT} = \frac{554016.5403}{571936.924} = 0.968667203$$

Im vorliegenden Fall würden also satte 96.8667203% der Streuung durch die Regressionsgerade erklärt. Zum Vergleich berechnen wir noch den Pearson-Korrelationskoeffizienten: 0.984208496, wobei $0.984208496^2 = 0.968666364$. Der kleine Unterschied ergibt sich durch Rundungsfehler. ◇

5.3.8 Partielle Korrelation

Haben wir drei Datensätze und gehen wir davon aus, dass die Korrelation zwischen den Daten des Satzes X und den Daten des Satzes Y vermutlich eine Scheinkorrelation ist und durch einen gleichwirkenden „Einfluss“ der dritten Variable Z zu erklären ist, so gibt es ein Verfahren, um die Korrelation von X und Y unter Ausschluss des Einflusses der dritten Variable zu berechnen. Wir sprechen von „partieller Korrelation“ zwischen X und Y (unter Ausschaltung des Einflusses von Z). So gibt es im Alter zwischen 6 und 14 Jahren vermutlich eine negative Korrelation zwischen Gewicht und Laufzeit beim 80-Meter-Lauf - je mehr Gewicht, desto kürzere Laufzeit. Im späteren Alter ist diese Korrelation hingegen wohl positiv - je mehr Gewicht, desto längere Laufzeit. Die beeinflussende Variable zwischen 6 und 14 ist das Alter und die Korrelation zwischen Gewicht sowie Laufzeit beim 80-Meter-Lauf ist in diesem Alterssegment eine Scheinkorrelation. Zudem kann mit der partiellen Korrelation eine zweite Einflussvariable konstant setzen: möchte man den Einfluss der Werbung auf den Umsatz ohne den Einfluss des Preises überprüfen, so kann man die Variable Preis herausrechnen. Hier würde sich allerdings eher die multiple Regression aufdrängen.

Um den Einfluss der dritten Variable Z wegzubekommen, berechnen wir die Regression zwischen Y und Z und die Residuen e_{yz} . Ebenso verfahren wir mit X und Z und erhalten e_{xz} . Anschliessend wird die Korrelation zwischen den Residuen berechnet. Formell definieren wir also:

Definition 5.3.20. $r_{xy/z} = r_{e_x e_y}$ mit den Residuen $e_{x_i} = x_i - f(z_i)$ und $e_{y_i} = y_i - g(z_i)$, wobei f die Regressionsgerade des Datenvektors x in Abhängigkeit vom Datenvektor z ist und g die Regressionsgerade des Datenvektors y in Abhängigkeit vom Datenvektor z ist (d.h. $r_{e_x e_y}$ ist die Korrelation zwischen den Residuenvektoren e_x und e_y).

Dieses Verfahren kann wie folgt abgekürzt werden, wie wir ohne Beweis festhalten wollen:

Satz 5.3.21. Die partielle Korrelation $r_{xy/z}$ zwischen X und Y unter Ausschaltung des Einflusses von Z ist:

$$r_{xy/z} = \frac{r_{xy} - (r_{xz}r_{yz})}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

r_{xy} für Pearson-Korrelation zwischen X und Y , etc.

Die so definierte partielle Korrelation lässt sich somit nur für Datensätze berechnen, für die paarweise die Voraussetzungen der Pearson-Korrelation gelten (Symmetrie, um eine Geraden streuen, metrische Daten).

Beispiel 5.3.22. Wir haben folgende Daten zur Verfügung (s. Tabelle 5.3.18):

Proband	Körpergrösse in cm	Schnelligkeit beim Lösen einer Aufgabe in Sekunden	Alter in Jahren	Residuen Alter Körpergrösse	Residuen Alter Schnelligkeit
1	90	10	5	1.919227392	0.139596137
2	100	5	9	-10.72168569	-1.513608428
3	80	12	5	-8.080772608	2.139596137
4	95	8	7	-4.401229148	-0.187006146
5	103	4	8	-2.061457419	-3.350307287
6	110	10	9	-0.721685689	3.486391572
7	120	4	13	-13.36259877	0.833187006
8	150	2	15	5.316944688	0.506584723
9	120	7	10	3.61808604	1.32309043
10	115	8	7	15.59877085	-0.187006146
11	110	6	9	-0.721685689	-0.513608428
12	130	3	10	13.61808604	-2.67690957

Tabelle 5.3.18: Beispieldaten für die Berechnung der partiellen Korrelation

Wir überprüfen zuerst, ob der Pearson-Korrelationskoeffizient sinnvoll anwendbar ist (s. Abbildungen 5.3.20 bis 5.3.23):

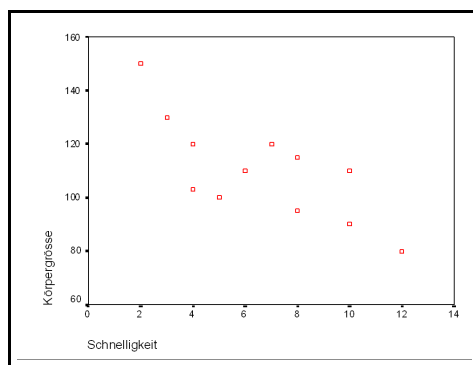


Abbildung 5.3.20: Bivariate Streudiagramme der Daten der Beispieldaten Tabelle 5.3.18

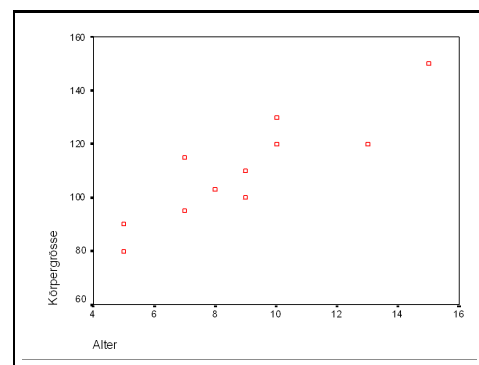


Abbildung 5.3.21: Bivariate Streudiagramme der Daten der Beispieldaten Tabelle 5.3.18

Abbildung 5.3.22: Beispieldaten für die Berechnung der partiellen Korrelation

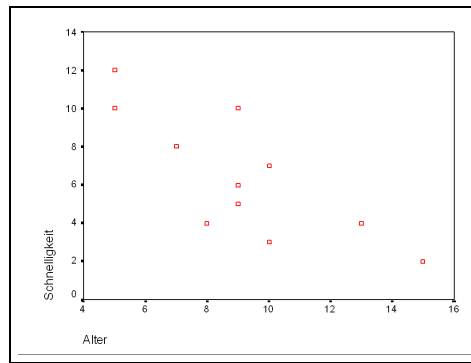


Abbildung 5.3.23: Bivariate Streudiagramme der Daten der Beispieldaten Tabelle 5.3.18

Wir kommen zum Schluss, dass die Daten ungefähr jeweils auf einer Geraden liegen und die Symmetriebedingung nicht verletzt ist. Zudem sind die Daten metrisch skaliert. Berechnen wir den Korrelationskoeffizienten zwischen Körpergröße und Schnelligkeit, erhalten wir: $r_{xy} = -0.755509151$. Somit gibt es eine starke negative Korrelation zwischen Körpergröße und Schnelligkeit. Doch wäre es wohl verfehlt, eine ursächliche Beziehung zwischen Körpergröße und Schnelligkeit herzustellen. Vermutlich ist es das Alter, das beide Größen beeinflusst. Wir berechnen somit eine partielle Korrelation mit $r_{xz} = 0.883222942$ und $r_{yz} = -0.788347528$ und erhalten mit obiger Formel:

$$r_{xy/z} = \frac{-0.755509151 - (0.883222942 \cdot -0.788347528)}{\sqrt{(1 - (0.883222942)^2)(1 - (-0.788347528)^2)}} = -0.20527$$

Das gleiche Resultat erhält man, wenn man die Residuen der Regressionsgerade

$$\text{Körper-Grösse}(\text{Alter}) = 5.66022827 \cdot \text{Alter} + 59.77963126$$

und die Residuen der Regressionsgerade

$$\text{Schnelligkeit}(\text{Alter}) = -0.836698859 \cdot \text{Alter} + 14.0438981$$

berechnet. Man erhält die Spalten in der Tabelle 5.3.18. Die Pearson-Korrelation zwischen den Residuen ist -0.20527 , wie Sie nachrechnen können. $\diamond$

5.3.9 Übungen

1. Berechnen Sie von Hand und mit Excel einen geeigneten Korrelationskoeffizienten der folgenden Daten (s. Tabelle 5.3.19):

x_i	y_i
1.5	3
2	10
3	12
4	20
5.2	25

Tabelle 5.3.19: Beispieldaten zweier metrisch skalierten Variablen

Interpretieren Sie!

2. Erstellen Sie für die folgenden Daten eine gemeinsame Häufigkeitsverteilung (in Mengenschreibweise und als Kreuztabelle). Stellen Sie mit Excel graphisch dar! (1 = sehr zufrieden,

2 = mittel zufrieden, 3 = unzufrieden; 1 = unter 20 Jahren, 2 = zwischen 20 und 40 Jahren, 3 = zwischen 40 und 60 Jahren, 4 = zwischen 60 und 80 Jahren, 5 = über 80 Jahre) (s. Tabelle 5.3.20)

Zufriedenheit	Altersgruppe
2	5
3	2
2	4
2	3
2	3
1	4
1	3
3	1
1	4
3	4
3	1
3	3
2	3
2	2
3	2
2	3
2	3
1	3
1	3

Tabelle 5.3.20: Übungsdaten von Übung 5.3.9.2

3. Stellen Sie die folgenden Daten graphisch dar (mit Excel). Berechnen Sie einen geeigneten Korrelationskoeffizienten (s. Tabelle 5.3.21)

Person	Lebensalter	Zigaretten pro Tag
1	70.8	25
2	84.3	13
3	91.9	10
4	83.9	14
5	63.7	29
6	70.4	27
7	72.5	24
8	73	23
9	91.3	10
10	68.2	28
11	89.2	10
12	74.7	21
13	82.3	14
14	85.7	12
15	84	13
16	63	33
17	78.1	17
18	79.4	15
19	76	19
20	62.2	34

Tabelle 5.3.21: Übungsdaten für Übung 5.3.9.3

Interpretieren Sie!

4. Stellen Sie die folgenden Daten (s. Tabelle 5.3.22) graphisch dar. Berechnen Sie einen geeigneten Korrelationskoeffizienten.

Person	Zigaretten pro Tag	Lebensdauer
1	39	48
2	18	72
3	36	60
4	39	45
5	39	61
6	34	52
7	27	69
8	32	56
9	32	55
10	29	65
11	32	61
12	27	73
13	14	80

Tabelle 5.3.22: Daten der Übung 5.3.9.4

Interpretieren Sie!

5. Jemand behauptet, Leute, die gut in Deutsch sind, seien im Allgemeinen schlecht in Buchhaltung und umgekehrt. Um die Behauptung zu überprüfen, werden die Noten einer Klasse untersucht (s. Tabelle 5.3.23). Berechnen Sie eine geeignete Korrelation zwischen den beiden

Person	Note Deutsch	Note Buchhaltung
1	5	4.8
2	4	4.2
3	3.5	3
4	5.5	5.2
5	4.5	4
6	4.7	5
7	4.4	4.7
8	3.8	4
9	5.2	5.8
10	4.8	4.6
11	4.2	4
12	4	4.5
13	4.6	4
14	4.9	4.5
15	5.1	4.5

Tabelle 5.3.23: Daten zur Übung 5.3.9.5

Variablen und interpretieren Sie das Ergebnis.

6. Berechnen Sie mit Excel einen geeigneten Korrelationskoeffizienten der folgenden Daten (s. Tabelle 5.3.24):

Land	Bevölkerung in Mio.	Hängige Verfahren
Frankreich	58.1	419
Italien	57.2	329
Deutschland	81.6	293
Spanien	39.8	291
Griechenland	10.5	241
Belgien	10.1	238
Portugal	9.8	199
GB	58.3	174
Österreich	8	163
Irland	3.6	146
Luxemburg	0.4	115
Niederlande	15.5	114
Schweden	8.8	101
Finnland	5.1	81
Dänemark	5.1	75

Tabelle 5.3.24: Daten der Übung 5.3.9.6 (Quelle: Hängige Verfahren, EUR-OP, 4/00, Zahlen des Jahres 1998; Hängige Verfahren sind Verfahren, die die EU-Kommission gegen EU-Mitgliedländer ergriffen hat, weil diese EU-Recht nicht umgesetzt haben.).

Wie interpretieren Sie das Ergebnis?

7. In einem Betrieb wird das Verhalten der Mitarbeiter von zwei Vorgesetzten unabhängig beurteilt (auf einer Skala von 1 bis 10). Dabei ergeben sich die folgenden Resultate (s. Tabelle 5.3.25)

Mitarbeiter	Vorgesetzter A	Vorgesetzter B
1	5	4
2	6	7
3	7	5.5
4	5	4
5	8	7.5
6	4	3.5
7	9	8
8	8.5	7
9	9	4
10	9.5	8
11	6.5	7
12	7.5	6

Tabelle 5.3.25: Daten für Übung 5.3.9.7

Sind die Beurteilungen der Vorgesetzten vergleichbar?

8. In einer Warenhauskette werden folgende Zahlen erhoben:

Verkaufs- stelle	Werbung pro Käufer	Verkauf pro Käufer in Franken	Preis pro Stück
1	5	720	8
2	6	750	7
3	7	900	6
4	2	270	13
5	3	420	11
6	8	1215	5
7	6.2	900	7
8	7.4	1080	6
9	8.3	1245	5
10	2.8	375	11.5
11	3.4	480	10
12	4	450	9.5
13	4.8	600	8
14	5.1	735	7.5
15	6.7	975	6
16	2.4	350	12
17	3.3	450	10.5
18	4.7	675	9

Tabelle 5.3.26: Daten für Übung 5.3.9.8

Untersuchen Sie, ob die verkaufte Menge eher mit der Werbung oder eher dem Preis statistisch zusammenhängt.

9. Berechnen Sie einen geeigneten Kennwert für die folgende Kreuztabelle 5.3.27

Zustand der Zähne * Rauchverhalten Kreuztabelle					
Anzahl		Rauchverhalten			Gesamt
		raucht nicht	raucht mässig	raucht viel	
Zustand der Zähne	gut	120	50	20	190
	mässig	80	60	30	170
	schlecht	30	65	50	145
Gesamt		230	175	100	505

Tabelle 5.3.27: Kreuztabelle der Übung 5.3.9.9

Interpretieren Sie das Ergebnis.

10. Die Marketing-Abteilung einer Firma vergleicht das eigene Produkt C mit zwei Konkurrenzprodukten und erhebt dazu den Zufriedenheitsgrad von Kunden mit den jeweiligen Produkten. Geben Sie an, welchen Kennwert für die resultierenden Häufigkeiten angemessen wäre (s. Tabelle 5.3.28) und interpretieren Sie:

Produkte * Zufriedenheit Kreuztabelle						
Anzahl		Zufriedenheit				Gesamt
		sehr zufrieden	zufrieden	knapp zufrieden	unzufrieden	
Produkte	Produkt A	20	30	10	5	65
	Produkt B	30	10	5	2	47
	Produkt C	10	20	30	15	75
Gesamt		60	60	45	22	187

Tabelle 5.3.28: Kreuztabelle der Übung 5.3.9.10

11. Um wieviel höher ist das näherungsweise Risiko für Raucher, an Krebs zu erkranken als für Nichtraucher, wenn die folgende Tabelle 5.3.29 gegeben ist:

	Lungenkrebs	Kein Lungenkrebs
Raucher	4	321
Nichtraucher	2	554

Tabelle 5.3.29: Übungsbeispiel für die Berechnung einer geeigneten Kennzahl

12. Berechnen Sie eine Kennzahl, mit Hilfe der Sie überprüfen können, ob die beiden betrachteten statistischen Variablen der folgenden Kreuztabelle 5.3.30 unabhängig sind.

	Wallis	Bern	übrige Schweiz	Summen
Marke A	8	6	32	46
Marke B	2	10	44	56
übrige	20	44	376	440
Summen	30	60	452	542

Tabelle 5.3.30: Kreuztabelle mit Übungsdaten für Berechnung einer Kennzahl

Interpretieren Sie

13. Überprüfen Sie für die Daten der Übung 13 (s. Excel-Tabelle), ob es sinnvoll ist, eine Regressions-Gerade zu berechnen. Berechnen Sie diese. Berechnen Sie das Bestimmtheitsmass. Wieviel % der Streuung der Einkommensdaten wird durch Variable „Ausbildung“ erklärt? Berechnen Sie das geschätzte Einkommen bei einer Ausbildung von 12.34 Jahren. Zeichnen sie die Regressionsgerade in die Punktwolke. Interpretieren Sie!

5.3.10 Lösungen

1. Wir überprüfen zuerst graphisch, ob überhaupt eine lineare Tendenz vorliegt. Ausreisser liegen keine vor. Die Punkte sind punktsymmetrisch verteilt (s. Abbildung 5.3.24):

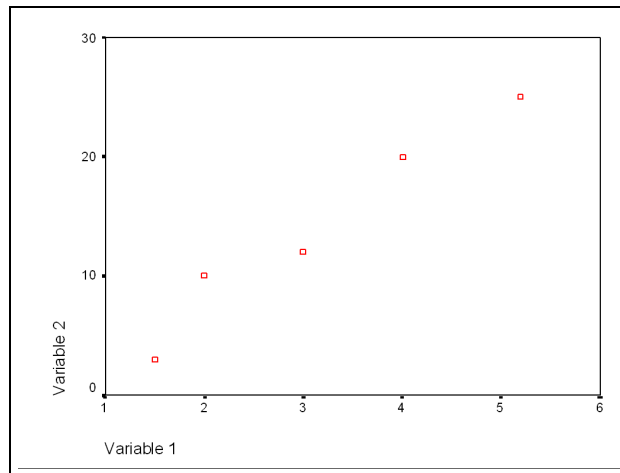


Abbildung 5.3.24: Beispieldaten Tabelle 5.3.19

Wir können also die Pearson-Korrelation verwenden - allerdings liegen (um das Beispiel von Hand - mit vertretbarem Aufwand - ausrechenbar zu machen) wenig Daten vor.

von Hand: Wir berechnen die Mittelwerte: $\bar{x} = \frac{1.5+2+3+4+5.2}{5} = 3.14$

$\bar{y} = \frac{3+10+12+20+25}{5} = 14$

Nun berechnen wir die Standardabweichungen für X und für Y:

$$s_x = \sqrt{\frac{(1.5-3.14)^2 + (2-3.14)^2 + (3-3.14)^2 + (4-3.14)^2 + (5.2-3.14)^2}{4}} = 1.4993$$

$$s_y = \sqrt{\frac{(3-14)^2 + (10-14)^2 + (12-14)^2 + (20-14)^2 + (25-14)^2}{4}} = 8.6313$$

Nun berechnen wir die Kovarianz:

$$s_{xy} = \frac{((1.5-3.14)(3-14)) + ((2-3.14)(10-14)) + ((3-3.14)(12-14)) + ((4-3.14)(20-14)) + ((5.2-3.14)(25-14))}{4} = 12.675$$

Dies ergibt einen Korrelationskoeffizienten von $r = \frac{12.675}{1.4993 \cdot 8.6313} = 0.97945$

Mit Excel erhalten wir: 0.979426112

Interpretation: Es gibt einen starken positiven Zusammenhang zwischen den beiden Variablen. Je weiter rechts die Werte der x-Variable rechts liegen, desto höher liegen die Werte der y-Variable. Je mehr von einem vorliegt, desto mehr liegt vom anderen vor.

2. (s. Tabelle 5.3.31)

4. Die Daten sind metrisch skaliert. Wir erstellen ein Streudiagramm, um die Voraussetzungen der Berechnung der Pearson-Korrelation zu überprüfen (s. Abbildung 5.3.28).

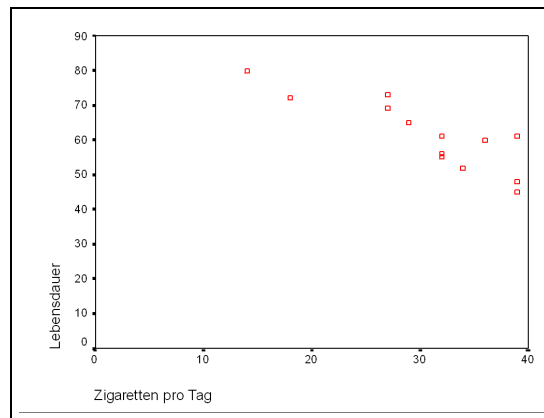


Abbildung 5.3.28: Streudiagramm der Beispieldaten Tabelle 5.3.22

Die Punkte liegen zuwenig punktsymmetrisch, um die Pearson-Korrelation zu verwenden.

Deshalb ist es besser, die Rangkorrelation zu benutzen. Wir erhalten (s. Tabelle 5.3.32): Rangkorrelation

Person	Zigaretten pro Tag	Lebensdauer	Ränge Zigaretten	Ränge Lebensdauer
4	39	45	12	1
1	39	48	12	2
6	34	52	9	3
9	32	55	7	4
8	32	56	7	5
3	36	60	10	6
11	32	61	7	7.5
5	39	61	12	7.5
10	29	65	5	9
7	27	69	3.5	10
2	18	72	2	11
12	27	73	3.5	12
13	14	80	1	13

Tabelle 5.3.32: Rangberechnung für Spearmankorrelation (Beispieldaten Tabelle 5.3.22)

von -0.844063537 (starke negative Korrelation: je mehr jemand raucht, desto weniger lange lebt er)

5. Da Noten (bestenfalls) ordinalskaliert sind, kann kein Streudiagramm gezeichnet werden. Wir müssen die Rangkorrelation verwenden (s. Tabelle 5.3.33):

Person	Note Deutsch	Note Buchhaltung	Ränge Deutsch	Ränge Buchhaltung
3	3.5	3	1	1
8	3.8	4	2	3.5
11	4.2	4	5	3.5
5	4.5	4	7	3.5
13	4.6	4	8	3.5
2	4	4.2	3.5	6
12	4	4.5	3.5	8
14	4.9	4.5	11	8
15	5.1	4.5	13	8
10	4.8	4.6	10	10
7	4.4	4.7	6	11
1	5	4.8	12	12
6	4.7	5	9	13
4	5.5	5.2	15	14
9	5.2	5.8	14	15

Tabelle 5.3.33: Berechnung Spearmankorrekation (Beispieldaten Tabelle 5.3.23)

Rangkorrelation: 0.745752968 (es gibt eine relativ starke positive Korrelation zwischen der Höhe der Noten in Deutsch und in Buchhaltung: je höher die Noten in Deutsch, desto höher sind sie in Buchhaltung).

6. Die Daten sind metrisch skaliert. Wir überprüfen zuerst graphisch, ob ein linearer, symmetrischer Zusammenhang vorliegt (s. Abbildung 5.3.29):

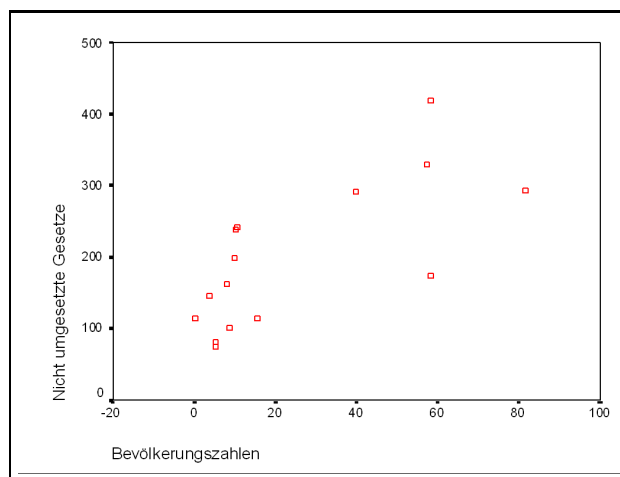


Abbildung 5.3.29: Streudiagramm der Beispieldaten 5.3.24

Die Symmetriebedingung ist kaum erfüllt. Deshalb sollte hier die Spearman-Korrelation verwendet werden (oder eine geeignete Transformation von Daten, z.B. logarithmieren einer oder beider Variablen). Wir berechnen deshalb den Spearman-Korrelationskoeffizienten (s. Tabelle 5.3.34):

	Bevölkerung in Mio.	Hängige Verfahren	Ränge Bevölkerung	Ränge Verfahren
Dänemark	5.1	75	3.5	1
Finnland	5.1	81	3.5	2
Schweden	8.8	101	6	3
Niederlande	15.5	114	10	4
Luxemburg	0.4	115	1	5
Irland	3.6	146	2	6
Österreich	8	163	5	7
GB	58.3	174	14	8
Portugal	9.8	199	7	9
Belgien	10.1	238	8	10
Griechenland	10.5	241	9	11
Spanien	39.8	291	11	12
Deutschland	81.6	293	15	13
Italien	57.2	329	12	14
Frankreich	58.1	419	13	15

Tabelle 5.3.34: Spearmankorrelation für Beispieldaten Tabelle 5.3.24

Und erhalten durch die Pearson-Korrelation auf die Ränge: $r_S = 0.731010122$

Möglicher Kommentar: Je mehr Bevölkerung ein Staat hat, desto weniger hält er die Gesetze der EU ein. Wenn man die Bevölkerungszahl bei ähnlichem Bruttosozialprodukt pro Kopf als Masszahl für die Macht eines Staates nimmt, so wäre dies ein Beleg für die folgende These: je mächtiger ein Staat, desto weniger hält er sich an internationale oder supranationale Gesetze und Regelungen.

7. Die Daten sind ordinalskaliert. Wir verwenden den Median, um die Gesamtbeurteilung zu vergleichen: Wir erhalten: 7.25 und 6.5. Der erste Vorgesetzte qualifiziert wohlwollender als der zweite. Wir verwenden die Spearman-Korrelation (s. Tabelle 5.3.35): Rangkorrelation:

Mitarbeiter	Vorgesetzter A	Vorgesetzter B	Ränge A	Ränge B
6	4	3.5	1	1
1	5	4	2.5	3
4	5	4	2.5	3
9	9	4	10.5	3
3	7	5.5	6	5
12	7.5	6	7	6
2	6	7	4	8
11	6.5	7	5	8
8	8.5	7	9	8
5	8	7.5	8	10
7	9	8	10.5	11.5
10	9.5	8	12	11.5

Tabelle 5.3.35: Berechnung Spearmankorrelation Beispieldaten Tabelle 5.3.25

0.67919726

Interpretation: in der Tendenz gibt es einen Zusammenhang zwischen den Beurteilungen: je höher die Bewertung einer Person durch den ersten Vorgesetzten, desto höher ist die Bewertung des zweiten Vorgesetzten. Dieser Zusammenhang ist aber nicht strikt!

8. Die Daten sind metrisch skaliert. Wir nehmen zuerst eine graphische Analyse vor:
Bivariate Streudiagramme für die Beispieldaten 5.3.26 (s. Abbildungen 5.3.30, 5.3.31 und

5.3.32

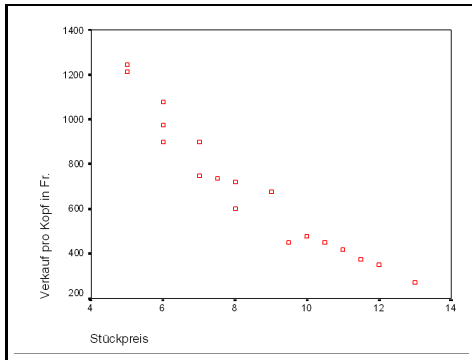


Abbildung 5.3.30: Bivariate Streudiagramme für die Beispieldaten 5.3.26

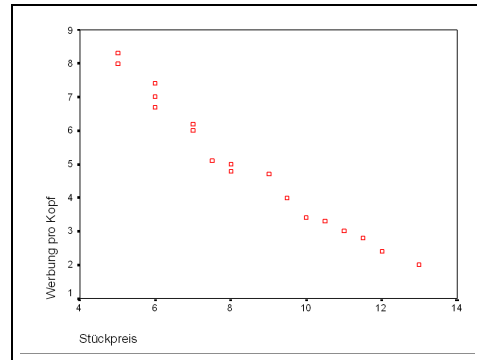


Abbildung 5.3.31: Bivariate Streudiagramme für die Beispieldaten 5.3.26

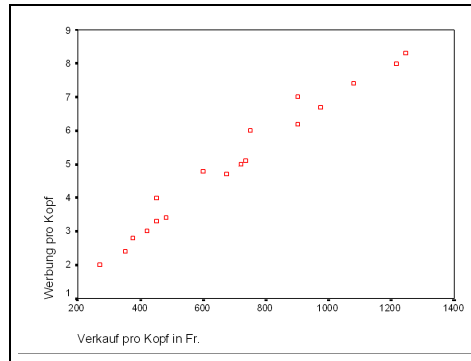


Abbildung 5.3.32: Bivariate Streudiagramme für die Beispieldaten 5.3.26

Wir kommen zum Schluss, dass die Punkte den Bedingungen (Symmetrie, um eine Geraden streuen) genügen - wobei das erste Diagramm diesen Schluss wohl nur knapp verträgt. Es lässt sich klar feststellen, dass sowohl bei höherer Werbung als auch bei tieferen Preisen, die Verkäufe ansteigen. Es stellt sich die Frage, ob der Effekt der Werbung oder den Preisen zuzuschreiben ist. Deshalb berechnen wir die partielle Korrelationen.

Es gilt: Korrelation Werbung Verkauf = $r_{wv} = 0.984695281$

Korrelation Verkauf Preis = $r_{vp} = -0.957093645$

Korrelation Werbung Preis = $r_{wp} = -0.983301791$

$$r_{w,v/p} = \frac{r_{wv} - (r_{wp}r_{vp})}{\sqrt{(1-r_{wp}^2)(1-r_{vp}^2)}} = \frac{0.984695281 - (-0.957093645 \cdot -0.983301791)}{\sqrt{(1-0.957093645^2)(1-0.983301791^2)}} = 0.826466012$$

$$r_{p,v/w} = \frac{r_{vp} - (r_{wp}r_{wv})}{\sqrt{(1-r_{wp}^2)(1-r_{wv}^2)}} = \frac{-0.957093645 - (-0.983301791 \cdot 0.984695281)}{\sqrt{(1-0.983301791^2)(1-0.984695281^2)}} = 0.351832429$$

Offenbar erklärt der Preis wenig von der Wirkung der Werbung weg, während die Werbung einiges von der Wirkung des Preises wegerklärt. Die Werbung wäre entsprechend fürs Kaufverhalten wichtiger als der Preis (frei erfundene Daten).

9. Da beide Variablen ordinal skaliert sind, können wir Gamma berechnen (s. Tabelle 5.3.33):

Zustand der Zähne * Rauchverhalten Kreuztabelle					
Anzahl		Rauchverhalten			Gesamt
		raucht nicht	mässig	raucht viel	
Zustand der Zähne	gut	120	50	20	190
	mässig	80	60	30	170
	schlecht	30	65	50	145
Gesamt		230	175	100	505

Abbildung 5.3.33: Kreuztabelle für die Berechnung von Gamma

$$\Pi_c = 120(60 + 30 + 65 + 50) + 50(30 + 50) + 80(65 + 50) + 60(50) = 40800$$

$$\Pi_d = 50(80 + 30) + 20(80 + 60 + 30 + 65) + 60(30) + 30(30 + 65) = 14850$$

$$\gamma = \frac{40800 - 14850}{40800 + 14850} = 0.46631$$

Es besteht ein Zusammenhang zwischen dem Rauchverhalten und dem Zustand der Zähne: je mehr man raucht, desto schlechter sind die Zähne.

10. siehe Tabelle 5.3.34

Produkte * Zufriedenheit Kreuztabelle						
Anzahl		Zufriedenheit				Gesamt
		sehr zufrieden	zufrieden	knapp zufrieden	unzufrieden	
Produkte	Produkt A	20	30	10	5	65
	Produkt B	30	10	5	2	47
	Produkt C	10	20	30	15	75
Gesamt		60	60	45	22	187

Abbildung 5.3.34: Kreuztabelle für die Berechnung von Gamma

Eine der Variablen ist ordinalskaliert, die andere jedoch nominal. Wir berechnen a) den Kontingenzkoeffizienten

$$\chi^2 = 50.386.$$

$$K = \sqrt{\frac{50.386}{187 + 50.386}} = 0.46071$$

$$K^* = \frac{0.46071}{\sqrt{\frac{2}{3}}} = 0.56425$$

Damit sind die beiden Variablen wohl unabhängig.

b) (fakultativ) das Konzentrationsmass (s. Tabelle 5.3.36): $n = 20 + 30 + 10 + 5 + 30 + 10 + 5 + 2 + 10 + 20 + 30 + 15 = 187$

$\frac{20}{187}$	$\frac{30}{187}$	$\frac{10}{187}$	$\frac{5}{187}$	$\frac{20+30+10+5}{187}$
$\frac{30}{187}$	$\frac{10}{187}$	$\frac{5}{187}$	$\frac{2}{187}$	$\frac{30+10+5+2}{187}$
$\frac{10}{187}$	$\frac{20}{187}$	$\frac{30}{187}$	$\frac{15}{187}$	$\frac{10+20+30+15}{187}$
$\frac{20+30+10}{187}$	$\frac{30+10+20}{187}$	$\frac{10+5+30}{187}$	$\frac{5+2+15}{187}$	1

Tabelle 5.3.36: Berechnung des Konzentrationsmasses für Kreuztabelle mit zwei nominalskalierten Variablen

$$\sum_{i=1}^I \sum_{j=1}^J \frac{h_{ij}^2}{h_{i+}} = \frac{\left(\frac{20}{187}\right)^2}{\frac{20+30+10+5}{187}} + \frac{\left(\frac{30}{187}\right)^2}{\frac{20+30+10+5}{187}} + \frac{\left(\frac{10}{187}\right)^2}{\frac{20+30+10+5}{187}} + \frac{\left(\frac{5}{187}\right)^2}{\frac{20+30+10+5}{187}} + \frac{\left(\frac{30}{187}\right)^2}{\frac{30+10+5+2}{187}} + \frac{\left(\frac{10}{187}\right)^2}{\frac{30+10+5+2}{187}} + \frac{\left(\frac{5}{187}\right)^2}{\frac{10+20+30+15}{187}} + \frac{\left(\frac{2}{187}\right)^2}{\frac{10+20+30+15}{187}} + \frac{\left(\frac{30}{187}\right)^2}{\frac{10+20+30+15}{187}} + \frac{\left(\frac{15}{187}\right)^2}{\frac{10+20+30+15}{187}} = 0.35018$$

$$\sum_{j=1}^J h_{+j}^2 = \left(\frac{20+30+10}{187}\right)^2 + \left(\frac{30+10+20}{187}\right)^2 + \left(\frac{10+5+30}{187}\right)^2 + \left(\frac{5+2+15}{187}\right)^2 = 0.277\,65$$

$$\tau = \frac{0.35018 - 0.27765}{1 - 0.27765} = 0.100\,41$$

Die Gesamttabelle ergibt eine schwache Abweichung von der Gleichverteilung auf die Zellen, obwohl das dritte Produkt in der Tendenz eindeutig schlechter abschliesst. Das Konzentrationsmass ist ein etwas rohes Mass. Kommen insbesondere eine ordinalskalierte Variable ins Spiel, so ist Vorsicht bei der Interpretation am Platz. Wichtige Informationen könnten sonst übersehen werden.

11. $\frac{4 \cdot 554}{2 \cdot 321} = 3.451713396$ oder $\frac{4.5 \cdot 554.5}{2.5 \cdot 321.5} = 3.104\,5$. Das näherungsweise Risiko, als Raucher Lungenkrebs zu erhalten ist gut 3 mal grösser als bei Nichtrauchern.

12. $\chi^2 = 17.23316322$; $K = \sqrt{\frac{17.23316322}{542 + 17.23316322}} = 0.175\,54$

$K^* = \frac{0.175\,54}{\sqrt{\frac{2}{3}}} = 0.214\,99$. Die Abweichung von der Unabhängigkeit ist nicht besonders stark, allerdings schlecht durch Zufall zu erklären, wie später zu erklären sein wird.

13. Beide Variablen sind metrisch skaliert. Die graphische Überprüfung des Streudiagramms (s. Excel-Tabelle) zeigt, dass die Punkte ungefähr auf einer Geraden liegen, dass es keine Ausreisser gibt und dass die Punktsymmetrie zu $(\bar{x}, \bar{y})$ gegeben ist. Man kann die Regressionsgerade berechnen.

Es gilt: $\bar{x} = 13.782$; $\bar{y} = 43993.881$; $n = 50$; $\sum_{i=1}^n x_i y_i = 31164619.73$; $\sum_{i=1}^n x_i^2 = 9774.07$ und damit

$$a = \frac{31164619 - 50 \cdot 13.782 \cdot 43993.881}{9774.07 - 50 \cdot 13.782^2} = 3064.121811$$

$$b = 43993.881 - 3064.121811 \cdot 13.782 = 1764.154201$$

Die Regressionsgerade ist damit $f(x) = 3064.121811x + 1764.154201$.

Für das Bestimmtheitsmass: Varianz der Einkommensdaten: 70133531.16; Varianz der Residuen: 17078178.72; $SQT = 836830757.1$; $SQR = 3436543027$; $SQE = 2599712270$

$$R^2 = \frac{2599712270}{836830757.1} = 0.756490534$$

Pearson-Korrelation: 0.869764643; deren Quadrat: 0.756490534

75.6% der Streuung der Einkommensdaten werden durch die Ausbildungsdauer erklärt.

Je mehr Ausbildung man absolviert, desto höher ist das Einkommen und zwar steigt das Einkommen durchschnittlich um 3064.12, wenn man ein Jahr mehr in die Ausbildung investiert (frei erfundene Daten).

Geschätztes Einkommen bei einer Ausbildung von 12.34 Jahre: $f(12.34) = 3064.121811 \cdot 12.34 + 1764.154201 = 39575.41735$. Für die Zeichnung, s. Excel-Tabelle.

5.4 Lernziele

- Bivariate Kreuztabellen erstellen können und entsprechende graphische Darstellungen von Hand und mit Excel zeichnen können
- Bivariate Streudiagramme erstellen können und an Hand der Zeichnung die Symmetrie und die das Um-Eine-Gerade-Streuen von Auge abschätzen können. (Für welche Skalenniveaus sind Streudiagramme sinnvoll?)
- Streudiagramme mit klassierten Daten auf einer Achse und nicht-klassierten, metrisch skalierten Daten auf der anderen Achse zeichnen können (von Hand und mit Excel). Box-Plots zeichnen können (mit einer Statistiksoftware).

- Das Chancenverhältnis berechnen und interpretieren können. Den Kontingenzkoeffizienten berechnen und interpretieren können.
- Gamma, den Pearson-Korrelationskoeffizienten und die Rangkorrelation berechnen können (die Angemessenheit der Verwendung der Masszahlen beurteilen können - nach Skalenniveau und Erfüllung der Bedingungen der Symmetrie und des Um-Eine-Gerade-Streuens bei den Korrelationen; keine Ausreisser). Die Masszahlen interpretieren können (angeben können, was sie bedeuten). Wissen, wann man das Konzentrationsmass verwendet.
- Die Regressionsgerade berechnen können - nach der Überprüfung der Voraussetzungen. Das Bestimmtheitsmass berechnen können. Schätzungen mit Hilfe der Regressionsgerade berechnen können.
- Den partiellen Korrelationskoeffizienten berechnen können und erklären können, worin die Idee hinter der Berechnung der partiellen Korrelation besteht.

Teil II

Kombinatorik

Kapitel 6

Etwas Kombinatorik

6.1 Einleitung

Bevor wir zur Wahrscheinlichkeitstheorie übergehen, behandeln wir ein Paar Ideen der Kombinatorik. Kombinatorik ist ein Zweig der Mathematik, der sich mit dem Zählen der Anzahl Elemente von spezifischen, endlichen Mengen beschäftigt. Wie viele m -Tupel gibt es, die aus den Elementen einer endlichen Menge A bildbar sind? Wie viele Teilmengen einer endlichen Mengen A gibt es? Wie viele Abbildungen von einer endlichen Menge A in eine endliche Menge B gibt es? Im Alltag kommen solche Probleme weniger mathematisch daher: Auf wie viele Arten kann ich die Karten beim Jassen (36 verschiedene Karten) auf vier Personen aufteilen? Auf wie viele Arten können 5 Personen auf einer Bank sitzen? Wie viele Möglichkeiten gibt es, aus einer Klasse von 18 Personen einen Ausschuss von 3 Personen zu bilden?

Kombinatorische Probleme können oft mit Hilfe von Urnenmodellen anschaulich gemacht werden. Das Problem „Wie viele Möglichkeiten gibt es, aus einer Klasse von 18 Personen einen Ausschuss von 3 Personen zu bilden?“ kann z.B. wie folgt beschrieben werden: Wie viele Möglichkeiten gibt es, aus einer Urne mit 18 Zetteln, die mit den Namen der 18 Personen beschriftet sind, 3 Zettel zu ziehen (ohne Zurücklegen der Zettel).

Zu beachten ist, dass Kombinatorik als Theorie mit der Wahrscheinlichkeitstheorie nichts zu tun hat. Entsprechend achten wir darauf, diese getrennt von der Wahrscheinlichkeitstheorie zu behandeln. Die Kombinatorik spielt in der Wahrscheinlichkeitstheorie als *Hilfsmittel* eine Rolle. Dies ist der Fall, wenn gleiche Wahrscheinlichkeit der Ergebnisse eines Zufallsexperimentes vorausgesetzt werden kann. Kennen wir die Anzahl der Ergebnisse eines Zufallsexperimentes, können wir dann auf Grund der gleichen Wahrscheinlichkeit die Wahrscheinlichkeit des Eintreffens eines einzelnen Ergebnisses berechnen. Die Kombinatorik erlaubt also manchmal die wirksame Berechnung der Anzahl der Ergebnisse eines Zufallsexperimentes.

6.2 Vier Grundprobleme der Kombinatorik

In der Kombinatorik können unter anderem folgende vier Grundtypen von Problemen unterschieden werden, wobei wir jeweils ein Urnenmodell zu Grunde legen und anschliessend dieselbe Problemlage mit einem Mengenmodell darstellen:

6.2.1 Ziehen mit Zurücklegen in Reihenfolge

Wir ziehen aus einer Urne mit n Objekten (die wir mit Namen versehen haben) jeweils m Objekte, die wir nach jedem Zug in die Urne zurücklegen. Wir schreiben für jedes Objekt, das wir ziehen und dann zurücklegen jeweils seinen Namen auf einer Zeile hin (von links nach rechts). Wie viele Namenfolgen sind auf diese Art möglich, wenn wir m mal ein Objekt ziehen und zurücklegen?

Wie können alle möglichen Reihenfolgen hinschreiben und diese dann zählen. Diese Vorgehensweise ist etwas mühsam. Entsprechend versuchen wir das Problem allgemein zu lösen. Beim ersten Zug haben wir n Möglichkeiten, ein Objekt zu ziehen. Im zweiten Zug haben wir für jede der n Möglichkeiten wiederum n Möglichkeiten. Dies ergibt $n \cdot n = n^2$ Möglichkeiten. Im dritten Zug haben wir für jede dieser n^2 Möglichkeiten, wiederum n Möglichkeiten, ein Objekt zu ziehen. Dies ergibt $n^2 \cdot n = n^3$ Möglichkeiten. Insgesamt gilt damit offensichtlich:

Satz 6.2.1. *Bei „Ziehen mit Zurücklegen in Reihenfolge“ gibt es n^m Möglichkeiten, m Objekte aus n Objekten zu ziehen*

Das Problem „Ziehen mit Zurücklegen in Reihenfolge“ können wir statt mit einem Urnenmodell auch mit einem mengentheoretische Modell formulieren. Wie viele Elemente enthält das m -fache kartesische Produkt der Menge A mit $|A| = n$? Oder anders formuliert: wie viele m -Tupel sind mit den Objekten einer Menge A bildbar, wenn die Objekte mehrmals im m -Tupel auftauchen können.

Beispiel 6.2.2. $A = \{1, 2, 3, 4, 5\}$. Wie viele Möglichkeiten gibt es, 2-Tupel aus A zu bilden, wobei Objekte mehrmals verwendet werden können. Offenbar erhalten wir für A , wenn $m = 2$:

$\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5),$
 $(2, 1), (2, 2), (2, 3), (2, 4), (2, 5),$
 $(3, 1), (3, 2), (3, 3), (3, 4), (3, 5),$
 $(4, 1), (4, 2), (4, 3), (4, 4), (4, 5),$
 $(5, 1), (5, 2), (5, 3), (5, 4), (5, 5)\} = A \times A.$

Die Anzahl der Elemente dieser Menge $= 5 \cdot 5 = 25$

◇

Die Problemlage „Ziehen mit Zurücklegen in Reihenfolge“ oder „Anzahl m -Tupel aus einer Menge mit n Objekten“ wird auch „ m -Variation mit Wiederholung genannt“, wobei bei dieser Redensweise vom Urnen- oder vom Mengenmodell abstrahiert wird.

6.2.2 Ziehen ohne Zurücklegen in Reihenfolge

Wir ziehen aus einer Urne mit n benannte Objekten m Objekte, ohne diese zurückzulegen, und schreiben die Namen der gezogenen Objekte von links nach rechts hin. Wie viele solcher Namenreihen sind möglich?

Offenbar gibt es n Möglichkeiten, ein erstes Objekt aus A auszuwählen. Für jede dieser n Möglichkeiten verbleiben $n - 1$ Möglichkeiten, ein zweites Objekt aus A auszuwählen (da Elemente nur einmal verwendet werden können, steht das erste nicht mehr zur Verfügung). Wenn $m = 2$, gibt es also offensichtlich $n \cdot (n - 1)$ mögliche Namensreihen. Für jede dieser $n \cdot (n - 1)$ gibt es $(n - 2)$ Möglichkeiten, ein drittes Objekt aus A auszuwählen. Somit gibt es $n \cdot (n - 1) \cdot (n - 2)$ Möglichkeiten, Namenreihen mit drei Namen aus A zu bilden. Dies führt zum folgenden Satz:

Satz 6.2.3. *Bei „Ziehen ohne Zurücklegen in Reihenfolge“ gibt es*

$$n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot (n - m + 1) = \prod_{i=0}^{m-1} (n - i)$$

Möglichkeiten, m Objekte aus einer Menge A mit n Elemente zu bilden.

Mit dem Mengenmodell können wir dieses Problem wie folgt beschreiben: Wie viele Möglichkeiten gibt es, aus Objekten einer Menge m -Tupel zu bilden, wenn die Objekte nur einmal verwendet werden können (formaler bestimmen wir $k = |\{(x_1, \dots, x_m) \mid x_i \in A \text{ und } x_i \neq x_j \text{ für } i, j \in \mathbb{N}_m\}|$).

Beispiel 6.2.4. $A = \{1, 2, 3, 4, 5\}$.

Offenbar erhalten wir bei $m = 2$:

$\{(1, 2), (1, 3), (1, 4), (1, 5),$
 $(2, 1), (2, 3), (2, 4), (2, 5),$

(3, 1), (3, 2), (3, 4), (3, 5),
 (4, 1), (4, 2), (4, 3), (4, 5),
 (5, 1), (5, 2), (5, 3), (5, 4)}

Die Anzahl der Elemente dieser Menge = $5 \cdot 4 = 20$

◇

Die Problemlage „Ziehung ohne Zurücklegen in Reihenfolge“ wird allgemein auch „Variation“ genannt.

Wenn $m = n$, nennen wir die resultierenden n -Tupel „Permutationen“. Es gilt offensichtlich, dass die Anzahl der Permutationen =

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot 2 \cdot 1 = \prod_{i=0}^{n-1} (n-i).$$

Für diesen Ausdruck führen wir eine gängige Abkürzung ein:

Definition 6.2.5. $n! := \begin{cases} \prod_{i=0}^{n-1} (n-i) & \text{für } n \in \mathbb{N}^* \\ 1 & \text{für } n = 0 \end{cases}$

„ $n!$ “ nennen wir n -te Fakultät (Mehrzahl: Fakultäten). Es gilt:

Satz 6.2.6. $1! = 1$

$n! = n(n-1)!$

Beweis. $1! = \prod_{i=0}^{1-1} (1-i) = 1 - 0 = 1$

$$n! = \prod_{i=0}^{n-1} (n-i) = n \prod_{i=1}^{n-1} (n-i) = n \prod_{i=0}^{n-2} ((n-1)-i) = n \prod_{i=0}^{(n-1)-1} ((n-1)-i) = n(n-1)! \quad \square$$

Berechnung mit Excel: Im Funktionenmenu unter *Math. und Trigonometrie* „=Fakultät(n)“. Mit der Schreibweise für Fakultäten können wir

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-m+1)$$

ohne die etwas unsaubere Schreibweise mit den Punkten oder die etwas schwerfällige Schreibweise mit dem Multiplikationszeichen wie folgt ausdrücken und berechnen:

Satz 6.2.7. $n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-m+1) = \frac{n!}{(n-m)!}$

Beweis. Durch Erweitern mit

$$\frac{(n-m) \cdot (n-m-1) \cdot \dots \cdot 1}{(n-m) \cdot (n-m-1) \cdot \dots \cdot 1} = 1$$

gilt:

$$\begin{aligned} & n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-m+1) \\ = & \frac{(n-m) \cdot (n-m-1) \cdot \dots \cdot 1}{(n-m) \cdot (n-m-1) \cdot \dots \cdot 1} \cdot n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-m+1) \\ = & \frac{((n-m) \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-m+1)) \cdot (n-m) \cdot (n-m-1) \cdot \dots \cdot 1}{(n-m) \cdot (n-m-1) \cdot \dots \cdot 1} \\ = & \frac{n!}{(n-m)!} \end{aligned}$$

□

Berechnung von $\frac{n!}{(n-m)!}$ mit Excel: Variationen($n; m$)

6.2.3 Ziehen ohne Zurücklegen ohne Reihenfolge

Wir ziehen m Objekte aus einer Urne mit n Objekten, ohne diese zurückzulegen, und legen diese auf einen Haufen. Wieviel Haufen sind bildbar (ein Haufen ist von einem anderen verschieden genau dann, wenn er nicht dieselben Objekte enthält).

Im Mengenmodell können wir das Problem wie folgt beschreiben: Wie viele Teilmengen mit $m \leq n$ Elementen sind aus einer Menge A mit n Elementen bildbar?

Beispiel 6.2.8. $A = \{1, 2, 3, 4, 5\}$.

Offenbar erhalten wir bei $m = 2$:

$\{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\},$

$\{2, 3\}, \{2, 4\}, \{2, 5\},$

$\{3, 4\}, \{3, 5\},$

$\{4, 5\}\}$

Die Anzahl der Elemente dieser Menge = 10

◇

Satz 6.2.9. Bei „Ziehen ohne Zurücklegen ohne Reihenfolge“ gibt es

$$\frac{n!}{m!(n-m)!}$$

Möglichkeiten, m Objekte aus n Objekten zu ziehen.

Beweis. Die Anzahl der m -Tupel, die aus der Menge A ohne Zurücklegen bildbar ist $\frac{n!}{(n-m)!}$. Somit ist $\frac{n!}{(n-m)!}$ die Anzahl der Elemente der Menge $S = \{(a_1, \dots, a_m) : a_1, \dots, a_m \in A \text{ und } a_i \neq a_j \text{ für } i \neq j\}$. Manche dieser m -Tupel fallen bei unserem Zählproblem jedoch zusammen, da die Reihenfolge keine Rolle spielt. Es fallen jeweils $m!$ m -Tupel zusammen, da es $m!$ Möglichkeiten gibt, die m Objekte in m -Tupel anzuordnen (Ziehen ohne Zurücklegen mit Reihenfolge). Die gesuchte Zahl $K(n, m)$ erhalten wir also, indem wir berechnen, wieviel Mengen X_i von m -Tupel es gibt, die jeweils zusammenfallen.

Dazu definieren wir folgende Äquivalenzrelation R : $(a_1, \dots, a_m)$ steht in der Relation R zu $(a'_1, \dots, a'_m)$ genau dann, wenn $\{a_1, \dots, a_m\} = \{a'_1, \dots, a'_m\}$. Laut Definition ist ein m -Tupel mit einem anderen genau dann R -äquivalent, wenn es dieselben Komponenten enthält. R ist eine Äquivalenzrelation, da „=“ eine Äquivalenzrelation ist. Sei Z die durch R bestimmte Zerlegung mit den Äquivalenzklassen X_i , d.h. $Z = \{X_i | X_i \text{ ist eine } R\text{-Äquivalenzklasse}\}$ und die X_i sind die Mengen der m -Tupel, die jeweils zusammenfallen. Da R eine Äquivalenzrelation ist, gilt für die Äquivalenzklassen X_i : $X_i \cap X_j = \emptyset$ für $j \neq i$ und $\bigcup_{i=1}^k X_i = S$ für die k der Äquivalenzklassen von Z . Da $|X_i| = m!$ für alle $i \in \mathbb{N}_k^*$, gilt

$$|Z| \cdot m! = |S| = \frac{n!}{(n-m)!}$$

Offensichtlich ist die gesuchte Anzahl $K(n, m) = |Z|$ und damit $K(n, m) = \frac{n!}{(n-m)!m!}$.

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
(1,2,3)	(1,2,4)	(1,2,5)	(1,3,4)	(1,3,5)	(1,4,5)	(2,3,4)	(2,3,5)	(2,4,5)	(3,4,5)
(1,3,2)	(1,4,2)	(1,5,2)	(1,4,3)	(1,5,3)	(1,5,4)	(2,4,3)	(2,5,3)	(2,5,4)	(3,5,4)
(2,1,3)	(2,1,4)	(2,1,5)	(3,1,4)	(3,1,5)	(4,1,5)	(3,2,4)	(3,2,5)	(4,2,5)	(4,3,5)
(2,3,1)	(2,4,1)	(2,5,1)	(3,4,1)	(3,5,1)	(4,5,1)	(3,4,2)	(3,5,2)	(4,5,2)	(4,5,3)
(3,1,2)	(4,1,2)	(5,1,2)	(4,1,3)	(5,1,3)	(5,1,4)	(4,2,3)	(5,2,3)	(5,2,4)	(5,3,4)
(3,2,1)	(4,2,1)	(5,2,1)	(4,3,1)	(5,3,1)	(5,4,1)	(4,3,2)	(5,3,2)	(5,4,2)	(5,4,3)

Tabelle 6.2.1: Beispiel zur Beweisidee: $A = \{1, 2, 3, 4, 5\}$; gesuchte Anzahl $K(5, 3)$ der Äquivalenzklassen X_i die alle gleichviele Elemente enthalten, nämlich $3! = 6$. S enthält $5 \cdot 4 \cdot 3 = 60 = ((5!)/(2!))$ Elemente. Es gilt somit: $|S| = 60 = K(5, 3) \cdot 3!$. Somit ist $K(5, 3) = ((60)/(3!)) = ((5!)/(2! \cdot 3!)) = 10$.

□

Die Problematik „Ziehungen ohne Zurücklegen und ohne Reihenfolge“ wird allgemeiner auch „Kombinationen“ genannt.

Statt

$$\frac{n!}{m!(n-m)!}$$

wird gewöhnlich die folgende Abkürzung verwendet:

Definition 6.2.10. $\binom{n}{m} := \frac{n!}{m!(n-m)!}$ für $m, n \in \mathbb{N}$

Aussprache von $\binom{n}{m}$: n tief m .

Die ersten zwei der folgenden Sätze sollten Sie kennen:

Satz 6.2.11. $\binom{0}{0} = 1 = \binom{n}{0} = \binom{n}{n}$

Beweis.

$$\begin{aligned} \binom{0}{0} &= \frac{0!}{0!(0-0)!} = \frac{1}{1} = 1 \\ \binom{n}{0} &= \frac{n!}{0!(n-0)!} = \frac{n!}{1(n)!} = 1 \\ \binom{n}{n} &= \frac{n!}{n!(n-n)!} = \frac{1}{(n-n)!} = \frac{1}{0!} = \frac{1}{1} = 1 \end{aligned}$$

□

Satz 6.2.12. $\binom{n}{m} = \binom{n}{n-m}$

Beweis.

$$\begin{aligned} \binom{n}{n-m} &= \frac{n!}{(n-(n-m))!(n-m)!} \\ &= \frac{n!}{m!(n-m)!} \\ &= \binom{n}{m} \end{aligned}$$

□

Satz 6.2.13. $\binom{n}{m+1} = \binom{n}{m} \cdot \frac{n-m}{m+1}$

Beweis.

$$\binom{n}{m+1} = \frac{n!}{(n-(m+1))!(m+1)!}$$

Nun ist:

$$(m+1)! = (m+1) \cdot m!$$

und

$$(n-(m+1))! \cdot (n-m) = (n-m)!$$

Damit erhalten wir (erster Schritt durch erweitern):

$$\begin{aligned} \frac{n!}{(n-(m+1))!(m+1)!} &= \frac{n!(n-m)}{(n-(m+1))!(n-m)m!(m+1)} \\ &= \frac{n!(n-m)}{(n-m)!m!(m+1)} \\ &= \binom{n}{m} \cdot \frac{n-m}{m+1} \end{aligned}$$

□

Satz 6.2.14. $\binom{n}{m+1} + \binom{n}{m} = \binom{n+1}{m+1}$

Beweis. (zweiter Schritt durch Erweitern)

$$\begin{aligned} \binom{n}{m+1} + \binom{n}{m} &= \frac{n!}{(n-(m+1))!(m+1)!} + \frac{n!}{(n-m)!m!} \\ &= \frac{n!(n+1)}{(n-(m+1))!(n+1)(m+1)!} + \frac{n!(n+1)(m+1)}{(n-m)!(n+1)m!(m+1)} \\ &= \frac{(n+1)!}{(n+1-(m+1))!(m+1)!} + \frac{(n+1)!(m+1)}{(n+1-m)!(m+1)!} \\ &= \frac{(n+1)!}{(n+1-(m+1))!(m+1)!} + \frac{(n+1)!}{(n+1-(m+1))!(m+1)!} \\ &= \frac{(n+1)!}{(n+1-(m+1))!(m+1)!} \\ &= \binom{n+1}{m+1} \end{aligned}$$

□

Satz 6.2.15. Wenn $m > n$, dann $\binom{n}{m} = 0$

Beweis. Der Beweis erfolgt durch eine Beweismethode, die in der Mathematik häufig angewendet wird. (1) Man zeigt, dass der Satz für die kleinste in Frage kommende natürliche Zahl i gilt. (2) Dann zeigt man, dass der Satz für eine natürliche Zahl $k+1$ gilt, wenn er für die natürliche Zahl k gilt. Wir nennen $k+1$ den "Nachfolger" von k . (3) Damit gilt der Satz für alle natürlichen Zahlen $k \geq i$, denn: Durch (2) beweisen wir, dass der Satz für jeden Nachfolger von k gilt. Da er für i gilt, gilt der Satz aber laut (2) für den Nachfolger von i , laut (2) für den Nachfolger des Nachfolgers von i , etc., d.h. für alle auch "indirekten" Nachfolger von i .

Wir wenden diese Beweismethode auf den Beweis des Satzes 5 an:

(1) Der Satz gilt für $\binom{n}{n+1}$, denn $\binom{n}{n+1} \stackrel{\text{Satz 3}}{=} \binom{n}{n} \frac{n-n}{n+1} = 0$.

(2) Der Satz gelte für $\binom{n}{m}$ für $m > n$. Es gilt nun $\binom{n}{m+1} = \binom{n}{m} \frac{n-m}{m+1} = 0$ (da laut Voraussetzung der Satz für $\binom{n}{m}$ gilt und damit $\binom{n}{m} = 0$ für $m > n$).

(3) Aus (1) und (2) folgt: $\binom{n}{m} = 0$ für $m > n$.

□

Satz 6.2.16. $\binom{n}{m} = \frac{n}{m} \binom{n-1}{m-1}$

Beweis.

$$\begin{aligned} \binom{n}{m} &= \frac{n!}{(n-m)!m!} \\ &= \frac{n \cdot (n-1)!}{m \cdot (n-1-(m-1))!(m-1)!} \\ &= \frac{n}{m} \binom{n-1}{m-1} \end{aligned}$$

□

$\binom{n}{m}$ werden „Binomialkoeffizienten“ genannt, da sie auch beim Ausrechnen von Binomen der Form $(a+b)^n$ eine Rolle spielen.

Es gilt nämlich:

Satz 6.2.17. $(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$.

Beweis. $(a+b)^n = \underbrace{(a+b) \cdot (a+b) \cdot \dots \cdot (a+b) \cdot (a+b)}_{n\text{-mal}}$. Das Produkt wird berechnet, indem die

Summe der folgenden Produkte bildet: man pickt aus jedem Faktor des Produktes $(a+b) \cdot (a+b) \cdot \dots \cdot (a+b) \cdot (a+b)$ ein a oder ein b und multipliziert diese. Diese Produkte sind die Summanden der Summe, wobei alle unterschiedlichen Möglichkeiten berücksichtigt werden müssen, aus den Faktoren derart a 's oder b 's zu picken. Für $(a+b)^3$ z.B. erhält man $(a+b)(a+b)(a+b) = aaa + aab + aba + baa + abb + bab + bba + bbb$. Die Bestimmung der Anzahl Summanden entspricht damit dem Problem „Ziehen mit Zurücklegen in Reihenfolge“. Entsprechend gibt es 2^n Summanden. Manche dieser Summanden kann man allerdings zusammenfassen, da z.B. $aab = aba = baa$ und damit $aab + aba + baa = 3a^2b$. Entsprechend ist $aaa + aab + aba + baa + abb + bab + bba + bbb = a^3 + 3a^2b + 3ab^2 + b^3$. Es stellt sich die Frage, wie viele solcher Summanden man jeweils zusammenfassen kann. Wir betrachten die a 's und die b 's vorerst als unterscheidbar und betrachten eine bestimmte Reihenfolge und eine spezifische Anzahl von a 's und b 's sowie $n = 10$: $a_1 a_2 b_1 a_3 b_2 b_3 a_4 a_5 a_6 b_4$. Es gibt $10!$ Möglichkeiten, dies anzuordnen. Manche dieser Möglichkeiten werden zusammenfallen, da die a 's und b 's eigentlich nicht unterscheidbar sind. Es gibt $6!$ Möglichkeiten, die a 's anzuordnen (indem man a 's vertauscht) und für jede dieser $6!$ Möglichkeiten $4!$ Möglichkeiten, die b 's anzuordnen. Es gibt also $6!4!$ Möglichkeiten, $a_1 a_2 b_1 a_3 b_2 b_3 a_4 a_5 a_6 b_4$ so anzuordnen, dass die jeweiligen Stellen der Zahlenserie immer noch von a 's respektive b 's eingenommen werden. Dies gilt für jede Reihe von a 's und b 's mit 6 und 4 Vorkommen. Wie viele solcher Reihen gibt es dann? Offensichtlich gilt für die Anzahl x solcher Reihen

$$x \cdot 6!4! = 10!$$

und damit

$$x = \frac{10!}{6!4!} = \binom{10}{6} = \binom{10}{4}.$$

Damit können wir jeweils $\binom{n}{m}$ (m = Anzahl der a 's) Summanden zusammenfassen. Im Beispiel ergibt sich also $\binom{10}{6} a^6 b^4 = \binom{10}{6} a^6 b^{10-4}$. Dies wird nun für alle möglichen Anzahlen a 's bestimmt, womit sich für $(a+b)^{10}$

$$\binom{10}{0} a^0 b^{10-0} + \binom{10}{1} a^1 b^{10-1} + \binom{10}{2} a^2 b^{10-2} + \dots + \binom{10}{9} a^9 b^{10-9} + \binom{10}{10} a^{10} b^{10-10} = \sum_{k=0}^{10} \binom{10}{k} a^k b^{10-k}$$

ergibt.

□

Zur weiteren Erläuterung folgen noch ein paar Berechnungsbeispiele:

$$\begin{aligned}(a+b)^2 &= \binom{2}{0}a^{2-0}b^0 + \binom{2}{1}a^{2-1}b^1 + \binom{2}{2}a^{2-2}b^2 = a^2 + 2ab + b^2 \\(a+b)^3 &= \binom{3}{0}a^{3-0}b^0 + \binom{3}{1}a^{3-1}b^1 + \binom{3}{2}a^{3-2}b^2 + \binom{3}{3}a^{3-3}b^3 = \\&= a^3 + 3a^2b + 3ab^2 + b^3 \\(a+b)^4 &= \binom{4}{0}a^{4-0}b^0 + \binom{4}{1}a^{4-1}b^1 + \binom{4}{2}a^{4-2}b^2 + \binom{4}{3}a^{4-3}b^3 + \binom{4}{4}a^{4-4}b^4 = \\&= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\end{aligned}$$

Berechnung mit Excel: „=Kombinationen(n ; m)“.

6.2.4 Ziehen mit Zurücklegen ohne Reihenfolge

Es werden m Objekte aus einer Urne mit n Elementen gezogen, wobei die Objekte mehrmals verwendet werden können und die Reihenfolge keine Rolle spielt. Um zu wissen, ob ein Objekt schon mal gezogen wurde oder nicht, können wir es bei jedem Zug entsprechend markieren.

Übung 6.2.18. $A = \{1, 2, 3, 4, 5\}$.

Offenbar erhalten wir bei $m = 2$ erhalten (wenn wir Zahlen, die wir beim zweiten Mal ziehen, mit einem Apostroph markieren, wodurch in den folgenden Mengen nicht die Zahlen selber haben, sondern Ergebnisse wie „erster Zug einer 1“ und „zweiter Zug einer 1“):

$$\begin{aligned}&\{\{1, 1'\}, \{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \\&\{2, 2'\}, \{2, 3\}, \{2, 4\}, \{2, 5\}, \\&\{3, 3'\}, \{3, 4\}, \{3, 5\}, \\&\{4, 4'\}, \{4, 5\}, \\&\{5, 5'\}\end{aligned}$$

Wir erhalten somit 15 Möglichkeiten.

Allgemein gilt, was wir ohne Beweis festhalten.:

Satz 6.2.19. Bei „Ziehen mit Zurücklegen ohne Reihenfolge“ gibt es

$$\binom{n+m-1}{m} = \frac{(n+m-1)!}{m!(n+m-1-m)!} = \frac{(n+m-1)!}{m!(n-1)!}$$

Möglichkeiten, m Objekte aus n Objekten zu ziehen.

Das Problem „Ziehen mit Zurücklegen ohne Reihenfolge“ wird auch „Kombinationen mit Zurücklegen“ genannt.

Nicht alle kombinatorischen Probleme entsprechen diesen vier Grundproblemen. Viele lassen sich jedoch mit Hilfe dieser vier Grundaufgaben lösen.

6.2.5 Übungen

1. Bilden Sie Menge der Tripel, die aus $A = \{1, 2, 3, 4, 5\}$ bildbar sind (in Reihenfolge und mit Zurücklegen).
2. Bilden Sie Menge der Tripel, die ohne Zurücklegen in Reihenfolge aus $A = \{1, 2, 3, 4, 5\}$ bildbar sind.
3. Bilden Sie Menge der Tripel, die aus $A = \{1, 2, 3, 4, 5\}$ bildbar sind - Objekte aus A dürfen nicht mehrmals verwendet werden. Wählen Sie aus den Tripeln mit identischen Komponenten jeweils einen Vertreter. Wie viele Vertreter gibt es?
4. Bilden Sie Menge der Tripel, die aus $A = \{1, 2, 3, 4, 5\}$ bildbar sind - Objekte aus A dürfen mehrmals verwendet werden. Wählen Sie aus den Tripeln mit identischen Komponenten jeweils einen Vertreter. Wie viele Vertreter gibt es?
5. Berechnen Sie die Anzahl der Reihenfolgen, in der 8 Personen auf einer Bank sitzen können?

6. Beim Jassen: 36 unterschiedliche Karten, 4 Spieler. Wie viele Möglichkeiten gibt es, die 9 Karten den Spielern zuzuweisen?
7. Wir betrachten eine Menge von 5 Sätzen: $\{p, q, r, s, t\}$. Jeder Satz sei wahr oder falsch. Wie viele Möglichkeiten gibt es, den Sätzen wahr oder falsch zuzuordnen.
8. In einer Kantine kann das Essen selber aus Beilagen (Reis, Nudeln, Kartoffeln), Fleisch (Braten, Ragout) und Salat (Grüner, Gemischter) zusammengestellt werden. Wie viele Zusammenstellungen sind möglich? (von jeder Kategorie - Beilagen, Fleisch und Salat - wird genau eine Speise gewählt).
9. Man hat drei Flüsse zu überqueren. Über den erste Fluss führen vier Brücken, über den zweiten fünf und über den dritten zwei. Wie viele mögliche Wege gibt es über die drei Flüsse (ein Weg A ist von Weg B genau dann verschieden, wenn er über mindestens eine verschiedene Brücke führt).
10. Ein Würfel wird sechsmal geworfen und man schreibt jeweils die Augenzahl auf. Wie viele Anordnungen sind denkbar. Wie viele davon enden mit 4? Wie viele enden mit 3 oder 4?
11. Aus 5 Personen A, B, C, D, E soll ein Dreierausschuss gewählt werden. Wie viele Wahlmöglichkeiten gibt es?
12. Wie viele Möglichkeiten gibt es, beim Zahlenlotto 6 Zahlen anzukreuzen (auf 45).
13. An einem Fest mit 20 Gästen stösst jeder mit jedem an. Wie oft klingen die Gläser?
14. In einer Meisterschaft mit 12 Mannschaften spielt jede Mannschaft gegen jede. Wie viele Spiele werden ausgetragen?
15. Auf wie viele Arten können 12 Personen auf 3 Vierbettzimmer verteilt werden?
16. Eine Klasse zählt 20 Schülerinnen und Schüler; sechs davon sollen die Klasse bei einer Veranstaltung der Schüler vertreten. Wie viele Auswahlen sind denkbar?
17. Ein Wort bestehe aus einer Anreihung von Buchstaben (von 2 bis 15 Buchstaben). Wie viele Wörter lassen sich aus einem Vokabular von 25 Buchstaben bilden?
18. 6 Weisswein- und 4 Rotweingläser werden in einer Reihe aufgestellt. Wieviel Möglichkeiten gibt es, eine Aufstellung vorzunehmen (Rotweingläser gelten als ununterscheidbar. Weissweingläser gelten als ununterscheidbar).
19. Beim Würfeln von 4 Würfeln (je 6 Augenzahlen). Wie viele Möglichkeiten gibt es, vier verschiedene Augenzahlen zu erhalten (mit und ohne Beachtung der Reihenfolge).
20. a) Wie viele Möglichkeiten gibt es, die Geburtstage von 18 Personen auf das Jahr zu verteilen? (inklusive Mehrfachbelegungen und 360 Tagen).
b) Wie viele Möglichkeiten gibt es, die Geburtstage auf die 360 Tage zu verteilen ohne Mehrfachbelegungen?
c) Wie viele Möglichkeiten gibt es, die Geburtstage auf 360 Tage zu verteilen, so dass mindestens eine Zweifachbelegung erfolgt.
21. Auf wie viele Arten können sich zwei nicht unterscheidbar Spatzen auf 4 Telegraphenleitungen verteilen?
22. Wieviel beträgt die Anzahl der Teilmengen der endlichen Menge A mit $|A| = n$
23. Auf einem Sporttoto-Tippschein sind 13 Fussballspiele aufgeführt. In einer danebenstehenden Kolonne ist anzugeben, ob der erste Klub gewinnt, oder der zweite oder ob unentschieden gespielt wird. Dazu wird „1“, „2“ oder „x“ verwendet. Wie viele Möglichkeiten gibt es, Voraussagen zu machen.

24. Wie viele Möglichkeiten gibt es, fünf Tassen mit Kaffee und fünf Tassen mit Tee anzuordnen (die Tassen mit Tee sowie die Tassen mit Kaffee sind ununterscheidbar).
25. Aus einer Urne mit 30 Kugeln (20 davon sind schwarz, 10 sind weiss) werden 5 Kugeln gezogen - ohne Zurücklegen. Wie viele mögliche Resultate sind dabei denkbar (bei den gezogenen Kugeln unterscheiden wir nicht zwischen den schwarzen Kugeln und nicht zwischen den weissen Kugeln).
26. Wie viele Abbildungen gibt es von A nach B ($|A| = n; |B| = m; n, m \in \mathbb{N}$). (Abbildungen verwenden alle Objekte von A als jeweils erste Komponente in den Elementen des Graphen der Funktion).
27. Wie viele vierstellige Ziffern lassen sich aus den Ziffern $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ bilden, wenn keine Ziffer zweimal verwendet werden kann.
28. Wie viele Elemente enthält das kartesische Produkt: $((A \times B) \times C) \times D \times E$, wenn $|A| = n; |B| = m; |C| = o; |D| = p; |E| = r$.

6.2.6 Lösungen

1. Ziehen mit Zurücklegen in Reihenfolge ($m = 3$). $A = \{1, 2, 3, 4, 5\}$.

- 1) (1,1,1), (1,1,2), (1,1,3), (1,1,4), (1,1,5)
- 2) (1,2,1), (1,2,2), (1,2,3), (1,2,4), (1,2,5)
- 3) (1,3,1), (1,3,2), (1,3,3), (1,3,4), (1,3,5)
- 4) (1,4,1), (1,4,2), (1,4,3), (1,4,4), (1,4,5)
- 5) (1,5,1), (1,5,2), (1,5,3), (1,5,4), (1,5,5)
- 6) (2,1,1), (2,1,2), (2,1,3), (2,1,4), (2,1,5)
- 7) (2,2,1), (2,2,2), (2,2,3), (2,2,4), (2,2,5)
- 8) (2,3,1), (2,3,2), (2,3,3), (2,3,4), (2,3,5)
- 9) (2,4,1), (2,4,2), (2,4,3), (2,4,4), (2,4,5)
- 10) (2,5,1), (2,5,2), (2,5,3), (2,5,4), (2,5,5)
- 11) (3,1,1), (3,1,2), (3,1,3), (3,1,4), (3,1,5)
- 12) (3,2,1), (3,2,2), (3,2,3), (3,2,4), (3,2,5)
- 13) (3,3,1), (3,3,2), (3,3,3), (3,3,4), (3,3,5)
- 14) (3,4,1), (3,4,2), (3,4,3), (3,4,4), (3,4,5)
- 15) (3,5,1), (3,5,2), (3,5,3), (3,5,4), (3,5,5)
- 16) (4,1,1), (4,1,2), (4,1,3), (4,1,4), (4,1,5)
- 17) (4,2,1), (4,2,2), (4,2,3), (4,2,4), (4,2,5)
- 18) (4,3,1), (4,3,2), (4,3,3), (4,3,4), (4,3,5)
- 19) (4,4,1), (4,4,2), (4,4,3), (4,4,4), (4,4,5)
- 20) (4,5,1), (4,5,2), (4,5,3), (4,5,4), (4,5,5)
- 21) (5,1,1), (5,1,2), (5,1,3), (5,1,4), (5,1,5)
- 22) (5,2,1), (5,2,2), (5,2,3), (5,2,4), (5,2,5)
- 23) (5,3,1), (5,3,2), (5,3,3), (5,3,4), (5,3,5)
- 24) (5,4,1), (5,4,2), (5,4,3), (5,4,4), (5,4,5)
- 25) (5,5,1), (5,5,2), (5,5,3), (5,5,4), (5,5,5)

Anzahl: $25 \cdot 5 = 125$.

Mit Hilfe der Formel: $5^3 = 125$

2. Ziehen ohne Zurücklegen in Reihenfolge ($m = 3$). $A = \{1, 2, 3, 4, 5\}$.

- 1) (1,2,3), (1,2,4), (1,2,5)
- 2) (1,3,2), (1,3,4), (1,3,5)
- 3) (1,4,2), (1,4,3), (1,4,5)
- 4) (1,5,2), (1,5,3), (1,5,4)
- 5) (2,1,3), (2,1,4), (2,1,5)
- 6) (2,3,1), (2,3,4), (2,3,5)

- 7) (2,4,1),(2,4,3),(2,4,5)
- 8) (2,5,1),(2,5,3),(2,5,4)
- 9) (3,1,2),(3,1,4),(3,1,5)
- 10) (3,2,1),(3,2,4),(3,2,5)
- 11) (3,4,1),(3,4,2),(3,4,5)
- 12) (3,5,1),(3,5,2),(3,5,4)
- 13) (4,1,2),(4,1,3),(4,1,5)
- 14) (4,2,1),(4,2,3),(4,2,5)
- 15) (4,3,1),(4,3,2),(4,3,5)
- 16) (4,5,1),(4,5,2),(4,5,3)
- 17) (5,1,2),(5,1,3),(5,1,4)
- 18) (5,2,1),(5,2,3),(5,2,4)
- 19) (5,3,1),(5,3,2),(5,3,4)
- 20) (5,4,1),(5,4,2),(5,4,3)

Anzahl: $20 \cdot 3 = 60$.

Mit der Formel: $5 \cdot 4 \cdot 3 = 60$.

3. Ziehen ohne Zurücklegen ohne Reihenfolge ($m = 3$). $A = \{1, 2, 3, 4, 5\}$. Wir wählen folgende „Repräsentanten“:

- 1) (1,2,3),(1,2,4),(1,2,5)
- 2) (1,3,4),(1,3,5)
- 3) (1,4,5)
- 4) (2,3,4),(2,3,5)
- 5) (2,4,5)
- 6) (3,4,5)

Anzahl: 10:

Mit der Formel: $\frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 10$

4. Ziehen mit Zurücklegen ohne Reihenfolge ($m = 3$). $A = \{1, 2, 3, 4, 5\}$. Wir wählen folgende „Repräsentanten“:

- 1) (1,1,1), (1,1,2), (1,1,3), (1,1,4),(1,1,5)
- 2) (1,2,2), (1,2,3),(1,2,4),(1,2,5)
- 3) (1,3,3),(1,3,4),(1,3,5)
- 4) (1,4,4),(1,4,5)
- 5) (1,5,5)
- 6) (2,2,2),(2,2,3),(2,2,4),(2,2,5)
- 7) (2,3,3),(2,3,4),(2,3,5)
- 8) (2,4,4),(2,4,5)
- 9) (2,5,5)
- 10) (3,3,3),(3,3,4),(3,3,5)
- 11) (3,4,4),(3,4,5)
- 12) (3,5,5)
- 13) (4,4,4),(4,4,5)
- 14) (4,5,5)
- 15) (5,5,5)

Anzahl: $15 + 10 + 6 + 3 + 1 = 35$

Mit der Formel: $\binom{n+m-1}{m} = \binom{5+3-1}{3} = \binom{7}{3} = \frac{7!}{3!(7-3)!} = 35$

5. Eine Person kann nicht in einer Reihenfolge an zwei Orten sitzen: Es geht somit um das Problem „ohne Zurücklegen in Reihenfolge“. Da wir alle anordnen, geht es darum, die Anzahl der Permutationen zu bestimmen. $8! = 40320$
6. Es handelt sich um ein Problem, das nicht einem der obigen vier Grundprobleme entspricht, das jedoch mit Hilfe dieser Grundprobleme gelöst werden kann. Wir können uns die Frage stellen, wie viele Möglichkeiten es gibt, dem ersten Spieler Karten zuzuweisen. Es geht somit

um das Problem „Ziehen ohne Zurücklegen ohne Reihenfolge“ für den ersten ($m = 9$, $n = 36$). Es gibt für ihn $\binom{36}{9}$ Möglichkeiten, Karten zugewiesen zu bekommen. Für den zweiten Spieler bleiben dann noch 27 Karten: $\binom{27}{9}$. Für den Dritten bleiben 18: $\binom{18}{9}$. Für den letzten sind seine Karten durch die den übrigen zugewiesenen Karten bestimmt. Somit beträgt die Anzahl der Möglichkeiten: $\binom{36}{9} \cdot \binom{27}{9} \cdot \binom{18}{9} = 2.14527522663 \cdot 10^{19}$ (Es wird multipliziert und nicht addiert, da es für jede Möglichkeit des ersten, Karten zugewiesen zu bekommen, $\binom{27}{9}$ Möglichkeiten für den zweiten gibt, Karten zugewiesen zu bekommen, etc.)

7. „ p “ kann „wahr“ oder „falsch“ zugeordnet werden (2 Möglichkeiten). „ q “ kann „wahr“ oder „falsch“ zugeordnet werden. Somit gibt es für q auch 2 Möglichkeiten. Für „ p “ und „ q “ ergibt dies $2 \cdot 2 = 4$. Wir betrachten eine Menge von 5 Sätzen: $\{p, q, r, s, t\}$. Jeder Satz sei wahr oder falsch. Wie viele Möglichkeiten gibt es, den Sätzen wahr oder falsch zuzuordnen. Es geht um 5-maliges Ziehen mit Reihenfolge mit Zurücklegen. ($n = 2$, $m = 5$). Dies ergibt: $2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 = 2^5 = 32$. Für n Sätze gibt es 2^n Möglichkeiten, „wahr“ oder „falsch“ zuzuordnen.
8. $3 \cdot 2 \cdot 2 = 12$
9. $4 \cdot 5 \cdot 2 = 40$
10. Mit Reihenfolge und Zurücklegen: $6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 = 6^6 = 46\,656$ (= Anzahl Zahlenfolgen) Es enden auf 4? $6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 = 7776$ (ebenso viele enden auf 1, 2, 3, 5 und 6). Es enden auf 3 oder 4? $6 \cdot 6 \cdot 6 \cdot 6 \cdot 6 \cdot 2 = 15\,552$
11. Problemlage: ohne Reihenfolge ohne Zurücklegen: $\binom{5}{3} = 10$
12. Problemlage: ohne Reihenfolge ohne Zurücklegen: $\binom{45}{6} = 8145\,060$
13. *Lösungsweg 1:* Die erste Person stösst mit 19 an, die zweite mit 18 (das Anstossen mit der 1. hat schon stattgefunden), die dritte mit 17, etc. Dies ergibt: $19 + 18 + 17 + \dots + 1 = \frac{19 \cdot (20)}{2} = 190$
Dabei haben wir die Formel verwendet: $\sum_{i=1}^n i = \frac{n(n+1)}{2}$ (mit $n = 19$).
Beweis der Formel durch Induktion: (1) $1 = \frac{1(1+1)}{2}$. Damit gilt der Satz für $n = 1$.
(2) Der Satz gelte für n , d.h. es gilt: $\sum_{i=1}^n i = \frac{n(n+1)}{2}$. Damit erhalten wir: $\sum_{i=1}^{n+1} i = (n+1) + \sum_{i=1}^n i = (n+1) + \frac{n(n+1)}{2} = \frac{2(n+1) + n(n+1)}{2} = \frac{(n+1)(n+2)}{2} = \frac{(n+1)((n+1)+1)}{2}$
Damit gilt der Satz für $n+1$, wenn er für n gilt. Daraus folgt: (3) Der Satz gilt für alle n .

Man kann sich die Formel auch wie folgt überlegen: Jede Person stösst mit jeder an (ohne sich selber). Dies ergibt $20 \cdot 19$. Allerdings berücksichtigen wir zu viele Paare: jedes Paar wird zweimal gezählt. Dies ergibt: $\frac{20 \cdot 19}{2} = 190$

Lösungsweg 2: Ziehen ohne Zurücklegen ohne Reihenfolge $\binom{20}{2} = 190$.

Beweis der Äquivalenz der beiden Lösungswege: Damit Personen anstossen können, muss es mindestens 2 Personen geben. Somit ist das minimale, in Frage kommende $n = 2$. Wir beweisen durch Induktion $\binom{n}{2} = \frac{(n-1)(n)}{2}$ (Achtung hier $n = 20$):

- (1) Der Satz gilt für $n = 2$: $\binom{n}{2} = \binom{2}{2} = 1 = \frac{1(2)}{2} = \frac{(n-1)(n)}{2}$
- (2) Der Satz gelte für ein beliebiges n , d.h. $\binom{n}{2} = \frac{(n-1)(n)}{2}$. Wir versuchen zu zeigen, dass er für $n+1$ gilt: $\binom{n+1}{2} = \frac{(n+1)!}{(n+1-2)!2!} = \frac{n!(n+1)}{(n-1)!2!} = \frac{n!(n+1)}{(n-2)!(n-1)!2!} = \frac{n!}{(n-2)!2!} \cdot \frac{n+1}{n-1} = \frac{(n-1)(n)}{2} \cdot \frac{n+1}{n-1} = \frac{(n+1-1)(n+1)}{2}$. Somit gilt
- (3) $\binom{n}{2} = \frac{(n-1)(n)}{2}$ für alle $n > 1$. q.e.d.

14. Die erste Mannschaft spielt gegen 11, die zweite gegen 10, etc. Somit gilt: $11 + 10 + 9 + \dots + 2 + 1 = \frac{11 \cdot 12}{2} = 66$ (oder äquivalent: $\binom{12}{2} = 66$).
15. Auf's erste Zimmer: Ziehen ohne Reihenfolge ohne Zurücklegen: $\binom{12}{4} = 495$
 Es verbleiben 8 Personen: Auf's zweite Zimmer gibt es noch Möglichkeiten: $\binom{8}{4} = 70$
 Es verbleiben 4 Personen (= 1 Möglichkeit). Somit gilt: $\binom{12}{4} \cdot \binom{8}{4} = 34650$
16. Eine Klasse zählt 20 Schülerinnen und Schüler; sechs davon sollen die Klasse bei einer Veranstaltung der Schüler vertreten. Wie viele Auswahlen sind denkbar? Ziehen ohne Reihenfolge ohne Zurücklegen: $\binom{20}{6} = 38760$
17. Mit Zurücklegen (da ein Buchstabe mehrmals im Wort auftauchen kann) mit Reihenfolge:
 $25^2 + 25^3 + 25^4 + 25^5 + 25^6 + 25^7 + 25^8 + 25^9 + 25^{10} + 25^{11} + 25^{12} + 25^{13} + 25^{14} + 25^{15}$
 $= 970\,127\,681\,891\,123\,453\,750$
18. Wenn die Gläser alle unterscheidbar wären, gäbe es $10!$ mögliche Reihenfolgen. Da die Rotweingläser einerseits, die Weissweingläser andererseits nicht unterscheidbar sind, fallen jedoch immer verschiedene Reihenfolgen zusammen: Wir betrachten eine spezifische Reihenfolge $A = wrrwwrwwrw$. Die Weissweingläser können in dieser Reihe untereinander auf $6!$ Arten angeordnet sein, die Rotweingläser können in dieser Reihe auf $4!$ Arten angeordnet sein. Dies bedeutet, dass für die Reihe A jeweils $6! \cdot 4!$ Reihenfolgen zusammenfallen. Offensichtlich gilt dies für jede Reihenfolge A . Somit gilt für die k Reihenfolgen ununterscheidbarer Rot- bzw. Weissweingläser:
 $k \cdot 6! \cdot 4! = 10! \implies k = \frac{10!}{6! \cdot 4!} = \binom{10}{6} = 210$
 Allgemein gilt: Bei n Objekten, von denen m und $(n - m)$ ununterscheidbar sind, gibt es $\frac{n!}{m!(n-m)!} = \binom{n}{m}$ mögliche Anordnungen. Somit fällt das Problem mit dem Problem „Ziehen ohne Zurücklegen ohne Reihenfolge“ zusammen.
19. Ziehen ohne Zurücklegen mit Reihenfolge: $6 \cdot 5 \cdot 4 \cdot 3 = 360$
 b) Ziehen ohne Zurücklegen ohne Reihenfolge: $\binom{6}{4} = 15$
20. a) 360^{18} b) $\frac{360!}{(360-18)!}$
 c) Die Anzahl der Möglichkeiten mit mindestens einer Mehrfachbelegung sind die Anzahl der Möglichkeiten von a) weniger die Anzahl der Möglichkeiten ohne Mehrfachbelegung:
 $360^{18} - \frac{360!}{(360-18)!}$
21. $4 + 3 + 2 + 1 = 10$
 Begründung: Sei einer der Spatzen auf dem ersten Draht. Dann kann der andere Spatz auf einem der vier Drähte sitzen (= 4 Möglichkeiten). Ist nun einer der Spatzen auf dem zweiten Draht, kann der andere nicht auf dem ersten sitzen, da wir diese Möglichkeit bereits berücksichtigt haben. Er kann jedoch auf dem zweiten, dritten oder vierten sitzen (3 Möglichkeiten). etc..
22. Wir definieren: $\mathcal{P}(A) := \{B : B \subset A\}$ ($\mathcal{P}(A)$ ist also die Menge aller Teilmengen von A . $\mathcal{P}(A)$ wird „Potenzmenge“ genannt. Es gilt: $|\mathcal{P}(A)| = 2^n$.
 Beweis: (1) $|\mathcal{P}(A)| = 2^n$ gilt für $n = 1$. Ist $|A| = 1$ gibt es ein x , so dass $A = \{x\}$. Die Anzahl der Teilmengen ist 2, da $\mathcal{P}(A) = \{\emptyset, \{x\}\}$. Nun ist $2^1 = 2$. Somit gilt der Satz für $n = 1$.
 (2) Wir nehmen an, dass der Satz für n gilt und versuchen zu zeigen, dass er für $n + 1$ gilt. Wenn der Satz für n gilt, ist $|\mathcal{P}(A)| = 2^n$ für $|A| = n$. Für $A' = A \cup \{x\}$ (und $x \notin A$) gilt $|A'| = n + 1$. Für $\mathcal{P}(A')$ gilt: $\mathcal{P}(A') = \mathcal{P}(A) \cup B$ mit $B = \{C : C = D \cup \{x\} \text{ und } D \in \mathcal{P}(A)\}$. Nun weist B und $\mathcal{P}(A)$ gleichviele Element auf, da B aus den Elementen von $\mathcal{P}(A)$ durch Vereinigung mit $\{x\}$ entsteht. Somit gilt dann $|\mathcal{P}(A')| = |\mathcal{P}(A) \cup B| = 2|\mathcal{P}(A)| = 2 \cdot 2^n = 2^{n+1}$. Damit gilt der Satz für $n + 1$, wenn er für n gilt.
 (3) Daraus können wir (durch Induktion) schliessen, dass der Satz für alle n gilt.
23. Es gibt drei Möglichkeiten pro Fussballspiel. Dies ergibt 3^{13} Möglichkeiten.

24. $\binom{10}{5} = 252$
25. Es werden nur Schwarze gezogen: Dann gibt es $\binom{20}{5} = 15504$. Anzahl der Möglichkeiten aus den schwarzen schwarze Kugeln zu ziehen.
 Es werden eine weisse und vier schwarze gezogen: Es gibt $\binom{10}{1} = 10$ Möglichkeiten, aus 10 Weissen eine Weisse zu ziehen und $\binom{20}{4} = 4845$ Möglichkeiten aus den Schwarzen 4 Schwarze zu ziehen. Insgesamt: 48450 Möglichkeiten:
 Es werden zwei weisse und drei schwarze gezogen: $\binom{10}{2} \binom{20}{3} = 51300$
 Es werden drei weisse und zwei schwarze gezogen: $\binom{10}{3} \binom{20}{2} = 22800$
 Es werden vier weisse und eine schwarze gezogen: $\binom{10}{4} \binom{20}{1} = 4200$
 Es werden nur weisse gezogen: $\binom{10}{5} = 252$
 Damit ergibt sich: $15504 + 48450 + 51300 + 22800 + 4200 + 252 = 1.4251 \times 10^5$
26. Für die Menge der Abbildungen von A nach B schreiben wir: $Abb(A, B)$. Es gibt m^n Abbildungen f mit $f \in Abb(A, B)$; ($|A| = n$; $|B| = m$). Wir beweisen dies durch Induktion:
 (1) Der Satz gilt für $n = 1$. In diesem Fall gibt es ein x , so dass $A = \{x\}$. Für jedes Element y aus B gibt es genau eine Abbildung mit einem Graphen $\{(x, y)\}$. Somit gibt es $m = m^n = m^1$ Abbildungen. Der Satz gilt also für $n = 1$.
 (2) Der Satz gelte für n . Somit gilt: Die Anzahl der Abbildungen von A nach B beträgt: m^n . Sei nun $A' = A \cup \{z\}$, so dass $|A'| = n + 1$ (d.h. $z \notin A$). Wir können nun für jede Abbildung f von A nach B eine Abbildung f' von A' nach B bilden: Graph von $f' :=$ Graph von $f \cup \{(z, y)\}$ mit $y \in B$. Alle m^n Abbildungen können mit je einem der m Elemente (z, y) aufgefüllt werden. Damit gibt es dann $m^n \cdot m = m^{n+1}$ Abbildungen von A' nach B . Damit gilt der Satz für $n + 1$ wenn er für n gilt.
 (3) Durch Induktion schliessen wir: Der Satz gilt für alle n .
 Zu beweisen wäre noch, dass der Satz für alle $m \in \mathbb{N}^*$ gilt.
27. Bei Ziffern spielt die Reihenfolge eine Rolle (ohne Zurücklegen): $\frac{9!}{(9-4)!} = 3024$
28. $|A| \cdot |B| \cdot |C| \cdot |D| \cdot |E| = n \cdot m \cdot o \cdot p \cdot r$.

6.2.7 Lernziele

- die vier Problemlagen „Ziehen mit/ohne Zurücklegen mit/ohne Reihenfolge“ kennen und die jeweiligen Anzahlen für beliebige n und m berechnen können (letzteres ohne die 4. Problemlage).
- kombinatorische Probleme von der Art der Übungsaufgaben lösen können (ohne jene mit Induktionsbeweisen). Insbesondere die Übung 18 spielt später noch eine wichtige Rolle.
- Die Definition von „Fakultät“ kennen und Fakultäten berechnen können.
- Die Definition von $\binom{n}{m}$ kennen, den Ausdruck lesen können und entsprechende Ausdrücke berechnen können.

Teil III

Einführung in die Wahrscheinlichkeitstheorie

Kapitel 7

Einleitung

Das Wort „wahrscheinlich“ gehört zum täglichen Sprachgebrauch, wobei wir das Wort in unterschiedlichen Bedeutungen verwenden. Im Satz „Wahrscheinlich gibt es ausserhalb unseres Sonnensystems keine Menschen“ äussere ich eine Vermutung. Im Satz „Mit einer grossen Wahrscheinlichkeit gewinne ich den nächsten Jackpot nicht“ drücke ich aus, dass Anteil der Gewinner an den Lotto-Spielern sehr klein ist. Entsprechend klein ist dann die Wahrscheinlichkeit, dass ich der Gewinner bin. Die Wahrscheinlichkeitstheorie interessiert sich nur für Wahrscheinlichkeiten, die als theoretisches Modell für relative Häufigkeiten aufgefasst werden können. Andere Wahrscheinlichkeitsbegriffe werden nicht diskutiert.

Die Wahrscheinlichkeitsrechnung als mathematische Disziplin stellt sich die Aufgabe, auf dem Hintergrund von relativen Häufigkeiten eine Theorie für die Eintrittschancen von Ereignissen aufzubauen. Man muss sich dabei auf Ereignisse beschränken, von denen vorausgesetzt werden kann, dass ihre Eintrittschancen gemessen werden können.

Der Geburtsort der Wahrscheinlichkeitstheorie war der Spieltisch. Ein leidenschaftlicher Spieler kontaktierte vor mehr als 300 Jahren den Franzosen Blaise Pascal (1623-1662; Philosoph und Mathematiker), damit ihm dieser errechne, wann er ein bestimmtes Glücksspiel vorzeitig abbrechen müsse, damit seine Gewinnchancen höher als jene des Gegners blieben. Pascal schuf in einem Briefkontakt mit dem Mathematiker Pierre Fermat (1601 - 1665) als erster heute noch verwendete grundlegende Formeln und Regeln der Wahrscheinlichkeitsrechnung. Christiaan Huygens (1629 - 1695) führte das Konzept des *Erwartungswertes* ein.

Einen wesentlichen Beitrag zur Weiterentwicklung der Wahrscheinlichkeitstheorie machte dann der Basler Mathematiker Jacob Bernoulli (1654 - 1705) mit seinem erst nach seinem Tode veröffentlichten und unvollendet gebliebenen Hauptwerk „Ars conjectandi“ - die „Kunst des Vermutens“. Mit dieser Schrift wurde das Zeitalter der Statistik und der Wahrscheinlichkeitsrechnung in der Mathematik eingeläutet, obwohl er seine Absicht, die entwickelte Wahrscheinlichkeitstheorie auf politische, rechtliche und wirtschaftliche Entscheidungen anzuwenden, nicht verwirklichen konnte. 1698 hatte Jacob Bernoulli das sogenannte „Gesetz der grossen Zahlen“ aufgestellt. Er definierte als erster Wahrscheinlichkeiten als Zahlen zwischen 0 und 1.

Die Wahrscheinlichkeitstheorie wird hier behandelt, um Vorarbeiten für die schliessende Statistik zu leisten. Bisher haben wir verschiedene statistische Kennzahlen kennen gelernt, die uns dazu dienen, Datensätze zu vergleichen (z.B. Welche Maschine hat die kleinere Varianz und führt damit zu einem kleineren Ausschuss. Welche Beurteilung von Mitarbeitern ist im Mittel wohlwollender. etc.) Wir haben aber bisher keine Mittel in der Hand, um solche Unterschiede als solche zu bewerten (sind sie „bedeutsam“ oder durch den Zufall einer spezifischen Datensammlung zu erklären?). Die schliessende Statistik dient dazu, die Bedeutsamkeit des Unterschiedes von Kennzahlen zu evaluieren. Dabei geht es nicht um inhaltliche Bedeutsamkeit (z.B. ökonomische Bedeutung), sondern um eine statistische Bedeutsamkeit: ist der Unterschied durch die Zufälligkeiten der Datensätze erklärbar oder ist der Unterschied eher schlecht durch die Zufälligkeiten der Datensätze erklärbar. Um solche Fragen zu entscheiden, müssen wir einen Umweg über die Wahrscheinlichkeitstheorie machen. Entsprechend werden wir vorerst einmal vor allem Wahrscheinlichkeitstheorie betreiben.

Wir behandeln zuerst nur sogenannte diskrete Wahrscheinlichkeitsräume (Kapitel 8 - Kapitel 10). Anschliessend werden die entwickelten Konzepte auf stetige Wahrscheinlichkeitsräume erweitert.

Kapitel 8

Diskrete Wahrscheinlichkeitsräume - Grundbegriffe

8.1 Zufallsexperimente, Ergebnisse, diskrete Ergebnismengen und Ereignisse

Lässt man einen Stein aus einer bestimmten Höhe fallen, so kann der zurückgelegte Weg zu jedem Zeitpunkt vorausgesagt werden. Das Ergebnis des Experimentes ist unter gleichbleibenden Voraussetzungen mit Sicherheit vorhersehbar. Wirft man einen Würfel, so kann die Augenzahl 1, 2, 3, 4, 5 oder 6 eintreffen. Das Ergebnis, die eintreffende Augenzahl, kann nicht vorausgesagt werden. Man nennt solche Experimente Zufallsexperimente. Bei Zufallsexperimenten hängt das Ergebnis vom Zufall ab. Zufallsexperimente, auf die man sich in der Wahrscheinlichkeitstheorie bezieht, müssen zudem beliebig oft unter denselben Bedingungen wiederholt werden können, so dass die relativen Häufigkeiten der Ergebnisse der Experimente in Erfahrung gebracht werden können. Der Begriff des „Zufallsexperimentes“ ist nicht mathematischer Art, dient jedoch dazu, den intuitiven Hintergrund der mathematischen Wahrscheinlichkeitstheorie zu beleuchten.

Definition 8.1.1. (Ergebnisse und Ergebnismenge): Die Menge aller Ergebnisse ω eines Zufallsexperimentes heisst Ergebnismenge Ω . Eine abzählbare (endliche oder nicht) Ergebnismenge heisst „diskret“.

Dabei definieren wir nicht, was Ergebnisse genau sind. Wir können Ergebnisse im Allgemeinen sprachlich recht unterschiedlich repräsentieren. Wir betrachten die jeweiligen Abkürzungen jeweils als Bezeichnungen von sonst nicht näher spezifizierten Objekten, die wir „Ergebnisse“ nennen. Wir verlangen von Ergebnissen nur, dass sie klar unterscheidbar sind. Man kann Ergebnisse etwa auch durch Ziffern repräsentieren, wobei die Ziffern dann nicht die entsprechenden Zahlen, sondern die entsprechenden Ergebnisse repräsentieren. Wir untersuchen zuerst nur diskrete Ergebnismengen.

Beispiel 8.1.2. 1) Beim Experiment „Werfen eines Würfels mit Augen“ ist die Ergebnismenge $\Omega = \{\text{ein Auge, zwei Augen, drei Augen, vier Augen, fünf Augen, sechs Augen}\}$. Diese Ergebnismenge können wir auch wiedergeben mit $\Omega = \{1, 2, 3, 4, 5, 6\}$, wobei wir die Ziffern n wie die Ausdrücke „ n Augen“ als Abkürzungen für das Ergebnis betrachten, dass die Seite des Würfels nach oben schaut, die n Augen aufweist. Entsprechend betrachten wir nicht die Zahlen selber oder die Ausdrücke „zwei Augen“ als Ergebnisse.

(2) Beim Experiment „Werfen eines Würfels mit Farben“ (wobei die verschiedenen Seiten des Würfels die Farben grün, gelb, rot schwarz, weiss und violett aufweisen) kann die Ergebnismenge z.B. durch $\Omega = \{\text{grün, gelb, rot, schwarz, weiss, violett}\}$ ausgedrückt werden.

(3) Beim Werfen einer Münze sind die Ergebnisse Kopf oder Zahl möglich. Die Ergebnismenge kann in der Form $\Omega = \{K, Z\}$ dargestellt werden.

(4) Wir ziehen aus den Studierenden der Hochschule Wallis nach Zufallsprinzip eine Person, wobei wir uns für die Variable „Geschlecht“ mit den Ausprägungen „männlich“ und „weiblich“ interessieren. Die gezogene Person ist weiblich oder männlich. Die Ergebnismenge ist damit $\Omega = \{\text{gezogene Person ist männlich, gezogene Person ist weiblich}\}$.

(5) Wir ziehen aus den Wallisern nach Zufallsprinzip eine Person und befragen sie nach der Zufriedenheit mit dem Leben, wobei vier Antwortmöglichkeiten vorgelegt werden (sehr zufrieden, zufrieden, mässig zufrieden, unzufrieden). Die Ergebnismenge $\Omega = \{\text{gezogene Person ist unzufrieden, gezogene Person ist mässig zufrieden, gezogene Person ist zufrieden, gezogene Person ist sehr zufrieden}\}$.

(6) Wir ziehen aus der Menge der in der Schweiz wohnhaften Personen nach Zufallsprinzip eine Person und messen deren Körpergewicht in kg. Die Ergebnismenge ist $\Omega = \{\text{Ergebnis, dass die gezogene Person ein Körpergewicht von } x \text{ kg hat} \mid x \in \mathbb{R}^+\}$. Die Ergebnismenge hat überabzählbar viele Ergebnisse. $\diamond$

Mit Hilfe von Ergebnissen können wir nicht alle möglichen Ereignisse ausdrücken, die wir verwenden möchten. Wir interessieren uns z.B. für das Ereignis, beim Würfeln eine gerade Zahl zu erhalten. Dieses Ereignis tritt genau dann ein, wenn beim Zufallsexperiment eine 2, eine 4 oder eine 6 gewürfelt wird. Oder wir interessieren uns für das Ereignis, eine zufriedene oder sehr zufriedene Person bei einer Zufallsziehung zu erwischen? Um solche Ereignisse ausdrücken zu können, nehmen wir die folgende Definition vor:

Definition 8.1.3. (Ereignis): A ist ein Ereignis genau dann, wenn $A \in \mathcal{P}(\Omega)$ ($\mathcal{P}(\Omega)$ ist die Potenzmenge der diskreten Ergebnismenge Ω , d.h. die Menge aller Teilmengen von Ω).

Alternativ könnte man „Ereignis“ auch wie folgt definieren: A ist ein Ereignis genau dann, wenn $A \subset \Omega$. Man beachte: Ergebnisse sind keine Ereignisse. Eine Menge, die ein Ergebnis als Element enthält, ist demgegenüber ein Ereignis. Zudem beachte man, dass die Potenzmenge von Ω sowohl die leere Menge als auch die Ergebnismenge Ω als Element enthält.

Definition 8.1.4. (Ereignis trifft ein): Ein Ereignis trifft genau dann ein, wenn eines seiner Elemente (= Ergebnisse) eintritt.

So trifft das Ereignis, eine gerade Zahl zu würfeln ein, wenn man ein 2, eine 4 oder eine 6 würfelt.

Aus dieser Definition folgt folgendes:

Merke:

- Eine Ereignis A und ein Ereignis B tritt genau dann ein, wenn ein Ergebnis ω eintritt, das Element von A und von B ist (d.h. $\omega \in A \cap B$).
- Ein Ereignis A oder ein Ereignis B tritt genau dann ein, wenn ein Ergebnis ω eintritt, das Element von A oder von B ist (d.h. $\omega \in A \cup B$).
- Ein Ereignis A tritt ein ohne dass das Ereignis B eintritt, genau dann, wenn ein Ergebnis ω eintritt, das Element von A ist, ohne Element von B zu sein (d.h. $\omega \in A \setminus B$, wobei $A \setminus B = \{\omega : \omega \in A \text{ und } \omega \notin B\}$).

Beispiel 8.1.5. (für Ereignisse)

(1) Sei $\Omega = \{K, Z\}$ (Ergebnismenge des Zufallsexperimentes des Werfens einer Münze). Dann sind folgende Mengen Ereignisse: $\emptyset, \{K\}, \{Z\}, \{K, Z\}$. Die Menge der Ereignisse ist $\mathcal{P}(\Omega) = \{\emptyset, \{K\}, \{Z\}, \{K, Z\}\}$.

(2) Sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ die Ergebnismenge, der dem Zufallsexperiment des Werfens eines Würfels entspreche. Dann gibt es $2^6 = 64$ Ereignisse: (man beachte: in Mengen spielt die Reihenfolge der Aufzählung ihrer Elemente keine Rolle)

$\mathcal{P}(\Omega) =$
 $\{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\},$

$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\},$
 $\{2, 3\}, \{2, 4\}, \{2, 5\}, \{2, 6\}, \{3, 4\}, \{3, 5\}, \{3, 6\}, \{4, 5\}, \{4, 6\}, \{5, 6\},$
 $\{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 2, 6\}, \{1, 3, 4\}, \{1, 3, 5\}, \{1, 3, 6\}, \{1, 4, 5\}, \{1, 4, 6\}, \{1, 5, 6\},$
 $\{2, 3, 4\}, \{2, 3, 5\}, \{2, 3, 6\}, \{2, 4, 5\}, \{2, 4, 6\}, \{2, 5, 6\},$
 $\{3, 4, 5\}, \{3, 4, 6\}, \{3, 5, 6\}, \{4, 5, 6\}$
 $\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 3, 6\}, \{1, 2, 4, 5\}, \{1, 2, 4, 6\}, \{1, 2, 5, 6\},$
 $\{1, 3, 4, 5\}, \{1, 3, 4, 6\}, \{1, 3, 5, 6\}, \{1, 4, 5, 6\},$
 $\{2, 3, 4, 5\}, \{2, 3, 4, 6\}, \{2, 3, 5, 6\}, \{2, 4, 5, 6\}, \{3, 4, 5, 6\}$
 $\{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \{1, 2, 3, 5, 6\}, \{1, 2, 4, 5, 6\}, \{1, 3, 4, 5, 6\}, \{2, 3, 4, 5, 6\},$
 $\{1, 2, 3, 4, 5, 6\}$

Das Ereignis, beim Würfeln eine gerade Augenzahl zu erhalten, ist $\{2, 4, 6\}$. Das Ereignis, die Augenzahlen 5 oder 3 zu erhalten ist $\{3, 5\}$. $\diamond$

Definition 8.1.6. Sei $A \subset B$. Dann gilt per Definition: $A_B^c := \{\omega : \omega \in B \text{ und } \omega \notin A\}$

A_B^c wird das Komplement von A bezüglich B genannt. Sind alle betrachteten Mengen Teilmenge einer Grundmenge G , so schreiben wir $A^c := A_G^c$.

Definition 8.1.7. Bezüglich Ereignissen nennen wir A^c (mit $A \subset \Omega$) *Gegenereignis* von A (bezüglich Ω).

Es folgen fünf Sätze - mengentheoretische, nicht wahrscheinlichkeitstheoretische -, die wir später brauchen.

Satz 8.1.8. $A \cap A^c = \emptyset$

Beweis. Wir nehmen an, es gebe ein $\omega \in A \cap A^c$. Damit gilt dann $\omega \in A$ und $\omega \in A^c$. Ist nun aber $\omega \in A^c$, so ist gemäss Definition von A^c ω nicht Element von A . Wir erhalten einen Widerspruch. $\square$

Satz 8.1.9. $A \cup A^c = \Omega$

Beweis. Sei $\omega \in A \cup A^c$. Dann gilt: $\omega \in A$ oder $\omega \in A^c$. Sei $\omega \in A$, dann ist $\omega \in \Omega$, da $A \subset \Omega$. Sei $\omega \in A^c$. Dann ist $\omega \in \Omega$ (gemäss Definition von A^c). Damit gilt: $A \cup A^c \subset \Omega$

Sei nun $\omega \in \Omega$. Zudem gilt: $\omega \in A$ oder $\omega \notin A$. Sei $\omega \in A$. Dann gilt: $\omega \in A \cup A^c$.

Sei $\omega \notin A$. Dann gilt auf Grund der Definition von A^c : $\omega \in A^c$. Damit gilt: $\omega \in A \cup A^c$. Somit gilt: $\Omega \subset A \cup A^c$

Aus $A \cup A^c \subset \Omega$ und $\Omega \subset A \cup A^c$ folgt: $A \cup A^c = \Omega$. $\square$

Bemerkung 8.1.10. Da $A \cup A^c = \Omega$ und $A^c \cap A = \emptyset$, stellt $\{A^c, A\}$ eine Zerlegung von Ω dar. $\diamond$

Beispiel 8.1.11. (für Gegenereignisse:)

1) Werfen eines Würfels: Ergebnismenge $\Omega = \{1, 2, 3, 4, 5, 6\}$

Ist $A = \{1, 2\}$, dann ist $A^c = \{3, 4, 5, 6\}$

Ist $A = \{1, 3, 5\}$, dann ist $A^c = \{2, 4, 6\}$

2) Werfen einer Münze:

Ist $A = \{K\}$, dann ist $A^c = \{Z\}$

Ist $A = \{Z\}$, dann ist $A^c = \{K\}$ $\diamond$

Satz 8.1.12. Sei $B \subset \Omega$ und $A \subset \Omega$. Dann gilt: $(B \cap A) \cup (B \cap A^c) = B$

Beweis. (i) Sei $\omega \in (B \cap A) \cup (B \cap A^c)$. Dann gilt $\omega \in (B \cap A)$ oder $\omega \in (B \cap A^c)$. Wir nehmen an, $\omega \in (B \cap A)$. Daraus folgt $\omega \in B$.

Wir nehmen an $\omega \in (B \cap A^c)$. Daraus folgt $\omega \in B$. Entsprechend gilt so oder so $\omega \in B$.

(ii) Sei $\omega \in B$. Damit ist $\omega \in \Omega$. Da A und A^c eine Zerlegung von Ω darstellen, gilt $\omega \in A$

oder $\omega \in A^c$. Sei $\omega \in A$. Damit gilt dann $\omega \in B \cap A^c$ und damit $\omega \in (B \cap A) \cup (B \cap A^c)$. Sei $\omega \in A^c$. Dann gilt $\omega \in (B \cap A^c)$ und damit $\omega \in (B \cap A) \cup (B \cap A^c)$. Entsprechend gilt so oder so $\omega \in (B \cap A) \cup (B \cap A^c)$. $\square$

Satz 8.1.13. Sei $B \subset \Omega$ und $A \subset \Omega$. Dann gilt: $(B \cap A) \cap (B \cap A^c) = \emptyset$.

Beweis. Wir nehmen an, es gebe ein $\omega \in (B \cap A) \cap (B \cap A^c)$. Damit gilt: $\omega \in (B \cap A)$ und $\omega \in (B \cap A^c)$. Damit gilt $\omega \in B$ und $\omega \in A$ und $\omega \in A^c$. Nun ist aber $\omega \in A$ genau dann, wenn $\omega \notin A^c$. Wir erhalten also einen Widerspruch und es gibt kein $\omega \in (B \cap A) \cap (B \cap A^c)$. Somit gilt $(B \cap A) \cap (B \cap A^c) = \emptyset$. $\square$

Satz 8.1.14. Für $B, A \subset \Omega$ gilt: $B^c \cap A^c = (B \cup A)^c$

Beweis. (a) Sei $\omega \in B^c \cap A^c$. Damit gilt $\omega \in B^c$ und $\omega \in A^c$ und damit $\omega \notin B$ und $\omega \notin A$. Sei nun $\omega \in (B \cup A)$. Dann wäre $\omega \in B$ oder $\omega \in A$. Wir erhalten einen Widerspruch. Damit gilt $\omega \notin (B \cup A)$ und damit $\omega \in (B \cup A)^c$.
(b) Sei $\omega \in (B \cup A)^c$. Damit gilt $\omega \notin (B \cup A)$. Damit gilt $\omega \notin B$ und $\omega \notin A$ und weiter $\omega \in B^c$ und $\omega \in A^c$. Damit gilt $\omega \in B^c \cap A^c$. $\square$

8.1.1 Übungen

1. Geben Sie für $\Omega = \{K, Z\}$ (K = Kopf liegt nach oben, Z = Zahl liegt nach oben beim Münzwurf) das Ereignis A an, Kopf zu werfen.
2. Geben Sie für $\Omega = \{1, 2, 3, 4, 5, 6\}$ ($\omega \in \Omega$ gibt die Anzahl Augenzahlen, die nach oben liegen, beim Würfeln an) das Ereignis A an,
 - (i) eine 5 zu würfeln;
 - (ii) eine ungerade Zahl zu würfeln.
3. Sei $\Omega = \{1, 2, 3, 4, 5, 6\}$ ($i \in \Omega$ gibt die Anzahl Augenzahlen, die nach oben liegen, beim Würfeln an). Geben Sie das Ereignis A an, eine in den natürlichen Zahlen durch 3 teilbare Augenzahl zu würfeln.
4. Geben Sie das Ereignis A an: „Die Augenzahl beträgt mindestens 5“ (Beim Würfeln mit einem Würfel)?
5. Die Zahlen von 1 bis 10 werden auf je ein Kärtchen geschrieben und diese in eine Urne gelegt. Geben Sie die Ergebnismenge für das Ziehen einer Karte an. Geben Sie die Ereignisse an, die beschrieben werden durch:
 - (a) Die gezogene Zahl ist eine Primzahl.
 - (b) Die gezogene Zahl ist eine ungerade Zahl.
 - (c) Die gezogene Zahl ist grösser als 7.
6. Geben Sie das Gegenereignis des folgenden Ereignisses an: $A = \{\omega : \omega \in \mathbb{N}^* \text{ und } \omega < 20\}$ wenn $\Omega = \{\omega : \omega \in \mathbb{N}^* \text{ und } \omega < 30\}$
7. Gegeben sei $\Omega = \{1, 2, 3, 4, 5\}$ und $A = \{1, 2\}$, $B = \{2, 3, 4\}$, $C = \{2, 5\}$
Welches der folgenden Ereignisse tritt auf, wenn 1 eintritt:
 $A; A \cup B; A \cap B; B \setminus A; A \setminus B; A \cup C; A \cap C; C \setminus A; B \cup C; B \cap C; B \setminus C; C \setminus B$
wenn 2 eintritt?

8.1.2 Lösungen

1. $A = \{K\}$
2. (i) $A = \{5\}$;
(ii) $A = \{1, 3, 5\}$.

3. $A = \{3, 6\}$
4. $A = \{5, 6\}$
5. $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
 - (a) $\{2, 3, 5, 7\}$
 - (b) $\{1, 3, 5, 7, 9\}$
 - (c) $\{8, 9, 10\}$
6. $A^c = \{20, 21, 22, 23, 24, 25, 26, 27, 28, 29\} = \{\omega : \omega \in \mathbb{N}^* \text{ und } 20 \leq \omega < 30\}$
7. Wenn 1 eintritt, so tritt $A, A \cup B, A \cup C, A \setminus B$ ein (die anderen Ereignisse treten nicht ein)
Wenn 2 eintritt, so tritt $A, A \cup B, A \cap B, A \cup C, B \cup C, B \cap C, A \cap C$ ein (die anderen Ereignisse treten nicht ein).

8.1.3 Lernziele

- Die Begriffe „Zufallsexperiment“, „diskrete Ergebnismenge“, „Ergebnis“, „Ereignis“ und „Gegenereignis“ erklären können.
- Erläutern können, wenn ein Ereignis eintritt (siehe obiges „Merke“).
- Die mengentheoretischen Sätze (8.1.8 und 8.1.9) anwenden können.
- Übungen der obigen Art lösen können.

8.2 Der Begriff der Wahrscheinlichkeit

Der Begriff der Wahrscheinlichkeit dient als theoretisches Modell für relative Häufigkeiten. Entsprechend weisen die beiden Grössen gewissen Gemeinsamkeiten auf, die wir in Kürze nachweisen werden. Doch vorgängig führen wir die Begriffe einer Wahrscheinlichkeitsfunktionen f und einer Wahrscheinlichkeitsverteilung P ein:

Definition 8.2.1. (Wahrscheinlichkeitsfunktion): Sei Ω eine diskreter Ergebnismenge. f ist eine Wahrscheinlichkeitsfunktion auf Ω genau dann, wenn gilt:

$$f : \Omega \rightarrow [0, 1]$$

$$\sum_{\omega \in \Omega} f(\omega) = 1$$

Die Wahrscheinlichkeitsfunktion ordnet somit jedem Ergebnis eine Zahl zwischen 0 und 1 zu, so dass die Summe der Funktionswerte 1 ergibt. Die von der Wahrscheinlichkeitsfunktion zugeordneten Zahlen nennen wir künftig „Wahrscheinlichkeiten“. Statt Wahrscheinlichkeitsfunktion schreiben wir auch „W-Funktion“.

Beispiel 8.2.2. (i) Die Wahrscheinlichkeit, bei einer Münze Kopf zu erhalten, sei 0.6, die Wahrscheinlichkeit Zahl zu erhalten sei 0.4. Damit ist $\Omega = \{K, Z\}$ und $f(K) = 0.6$ und $f(Z) = 0.4$. Der Graph der Wahrscheinlichkeitsfunktion sieht fürs Beispiel wie folgt aus: $\{(K, 0.6), (Z, 0.4)\}$
(ii) Die Wahrscheinlichkeit, bei einem Würfel eine spezifische Augenzahl zu erhalten, sei $\frac{1}{6}$. Kürzen wir die Augenzahlen mit den entsprechenden Zahlen ab, so erhalten wir den folgenden Graphen der Wahrscheinlichkeitsfunktion: $\{(1, \frac{1}{6}), (2, \frac{1}{6}), (3, \frac{1}{6}), (4, \frac{1}{6}), (5, \frac{1}{6}), (6, \frac{1}{6})\}$. $\diamond$

Wir brauchen allerdings Wahrscheinlichkeiten bezüglich der Ereignisse. Uns würde z.B. interessieren, wie hoch die Wahrscheinlichkeit ist, beim Würfeln eine gerade Zahl zu würfeln. Durch die Wahrscheinlichkeitsfunktion wird bei diskreten Ergebnismengen ein sogenanntes Wahrscheinlichkeitsmass auf die Menge der Ereignisse definiert, das uns Aussagen über solche Wahrscheinlichkeiten erlaubt:

Definition 8.2.3. (Diskrete Wahrscheinlichkeitsverteilung P auf Ω):

$$P : \mathcal{P}(\Omega) \rightarrow [0, 1]$$

$$P(A) := \sum_{\omega \in A} f(\omega)$$

(für $A \subset \Omega$; Ω ist eine diskrete Ergebnismenge ,

$\mathcal{P}(\Omega)$ = Potenzmenge von Ω)

Die Wahrscheinlichkeitsverteilung wird auch „Wahrscheinlichkeitsmass“ genannt - wobei wir auch die Ausdrücke W-Verteilung und W-Mass brauchen werden. Die durch die Wahrscheinlichkeitsverteilung zugeordneten Zahlen nennen wir ebenfalls „Wahrscheinlichkeiten“. Beachten Sie, dass P eine Mengenfunktion ist: sie ordnet jeder Teilmenge von Ω eine Wahrscheinlichkeit zu. Demgegenüber ist f eine Funktion, die nicht unbedingt Mengen, sondern jedem Element von Ω eine Wahrscheinlichkeit zuordnet. Aus Gründen des einfacheren Sprechens verwenden wir aber im Allgemeinen den Ausdruck „Wahrscheinlichkeitsverteilung auf Ω “ statt auf $\mathcal{P}(\Omega)$, obwohl P auf $\mathcal{P}(\Omega)$ definiert ist. Im Speziellen gilt:

Satz 8.2.4. Sei f eine Wahrscheinlichkeitsfunktion auf die diskrete Ergebnismenge Ω und P die durch f gemäss Definition 8.2.3 bestimmte Wahrscheinlichkeitsverteilung auf $\mathcal{P}(\Omega)$. Dann gilt:

$$f(\omega) = P(\{\omega\}).$$

Beweis. $P(\{\omega\}) \stackrel{8.2.3}{=} \sum_{\omega \in \{\omega\}} f(\omega) = f(\omega).$ □

Der Satz hat zur Folge, dass aus einer diskreten Wahrscheinlichkeitsverteilung P die Wahrscheinlichkeitsfunktion berechnet werden kann: Wir suchen alle einelementigen Teilmengen von Ω und übernehmen die Wahrscheinlichkeiten, die diesen von P zugeordnet werden. z.B. Graph von $P = \{(\{1\}, \frac{1}{3}), (\{2\}, \frac{1}{3}), (\{4\}, \frac{1}{3}), (\emptyset, 0), (\{1, 2, 4\}, 1), (\{1, 2\}, \frac{2}{3}), (\{1, 4\}, \frac{2}{3}), (\{2, 4\}, \frac{2}{3})\}$. Wir erhalten: Graph von $f = \{(1, \frac{1}{3}), (2, \frac{1}{3}), (4, \frac{1}{3})\}$. Damit können wir aus einer Wahrscheinlichkeitsfunktion die Verteilung berechnen und umgekehrt. Ist P gegeben, so nennen wir die aus P berechenbare W-Funktion f „die durch P bestimmte W-Funktion f “. Ist f gegeben, so nennen wir das aus f berechenbare W-Mass „das durch f bestimmte W-Mass P “.

Definition 8.2.5. (diskreter Wahrscheinlichkeitsraum): Das Tripel $(\Omega, \mathcal{P}(\Omega), P)$ wird diskreter Wahrscheinlichkeitsraum genannt (Ω = diskrete Ergebnismenge, $\mathcal{P}(\Omega)$ = Potenzmenge von Ω , P Wahrscheinlichkeitsverteilung auf Ω).

Für die Wahrscheinlichkeitsverteilung gilt nun auf Grund der obigen Definition:

Satz 8.2.6. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und f die durch P bestimmte W-Funktion.

(a) $P(\Omega) = \sum_{\omega \in \Omega} f(\omega) = 1$

(b) $P(A) \geq 0$ (für $A \subset \Omega$)

(c) Wenn $A_i \cap A_j = \emptyset$, dann $P(A_i \cup A_j) = P(A_i) + P(A_j)$

Beweis. (a) folgt unmittelbar aus den Definitionen: $P(\Omega) \stackrel{\text{Definition 8.2.3}}{=} \sum_{\omega \in \Omega} f(\omega) \stackrel{\text{Definition 8.2.1}}{=} 1$

(b) $P(A) = \sum_{\omega \in A} f(\omega) \geq 0$, da $f: \Omega \rightarrow [0, 1]$ und damit $\sum_{\omega \in A} f(\omega)$ eine Summe von nichtnegativen Zahlen ist.

(c) Sei $A_i \cap A_j = \emptyset$. Dann gilt: $P(A_i \cup A_j) = \sum_{\omega \in A_i \cup A_j} f(\omega) \stackrel{\text{da } A_i \cap A_j = \emptyset}{=} \sum_{\omega \in A_i} f(\omega) + \sum_{\omega \in A_j} f(\omega) = P(A_i) + P(A_j).$ □

Dabei ist (c) ein Spezialfall von: $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ sofern für alle $i \neq j$: $A_j \cap A_i = \emptyset$.

Wir zeigen nun die nahe Verwandtschaft von Wahrscheinlichkeiten und relativen Häufigkeiten, indem wir zum obigen Satz analoge Sätze für relative Häufigkeiten festhalten. Seien G eine

Grundmenge und x_i die m möglichen Skalenwerte bezüglich der Variable V , $k(x_i)$ die Anzahl Objekte aus G mit dem Skalenwert x_i , $h(x_i)$ die relative Häufigkeit der Objekte aus G mit dem Skalenwert x_i und $n = |G|$. Dann gilt:

Satz 8.2.7. (1) $\frac{|G|}{|G|} = \frac{n}{n} = 1 = \sum_{k=1}^n h(x_i)$ (Summe der relativen Häufigkeiten)

(2) $h(x_i) = \frac{k(x_i)}{n} \geq 0$ für alle $1 \leq i \leq m$

(3) Fassen wir zwei Ausprägungen mit den Skalenwerten x_i, x_j zu einer Ausprägung mit einem einzigen Skalenwert x_l zusammen, so gilt: $h(x_l) = h(x_i) + h(x_j)$ (ein Analogon zu $A_i \cap A_j = \emptyset$ ist hier nicht nötig, da Variablen mit ihren Ausprägungen eine Zerlegung von G produzieren!).

Da die leere Menge $\emptyset$ eine Teilmenge von Ω ist (und damit Element von $\mathcal{P}(\Omega)$), muss die Wahrscheinlichkeitsverteilung auch dieser Teilmenge einen Wert zuordnen. Es gilt:

Satz 8.2.8. $P(\emptyset) = 0$

Beweis. Es gilt: $\emptyset = A \cap \emptyset$. Damit gilt dann gemäss Satz 8.2.6(c) $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$. Somit gilt $P(\emptyset) = 0$. $\square$

Definition 8.2.9. Wir nennen $\emptyset$ das „unmögliche Ereignis“ und Ω das „sichere Ereignis“.

Beachten Sie: Aus $P(A) = 0$ kann man nicht schliessen $A = \emptyset$. Es kann also Ereignisse geben, die nicht das unmögliche Ereignis sind, obwohl ihnen die Wahrscheinlichkeit 0 zugeordnet ist.

Satz 8.2.10. $P(A \cup A^c) = P(A) + P(A^c)$

Beweis. Da gemäss Satz 8.1.8 $A \cap A^c = \emptyset$, gilt laut Satz 8.2.6(c) der Satz unmittelbar. $\square$

Satz 8.2.11. $P(A) = 1 - P(A^c)$

$P(A^c) = 1 - P(A)$

Beweis. Da gemäss Satz 8.1.9

$$A \cup A^c = \Omega,$$

gilt (Satz 8.2.6(c))

$$P(A \cup A^c) = P(\Omega) = 1$$

Zudem gilt nach Satz 8.2.10:

$$P(A \cup A^c) = P(A) + P(A^c).$$

Damit gilt:

$$P(A) + P(A^c) = 1,$$

woraus die obigen Sätze sofort folgen. $\square$

Satz 8.2.12. Wenn $A \subset B$ dann $P(A) \leq P(B)$

Beweis.

$$P(A) = \sum_{\omega \in A} f(\omega)$$

Nun ist $A \subset B$ und $f(\omega) \geq 0$ für $\omega \in \Omega$. Damit folgt

$$\sum_{\omega \in A} f(\omega) \leq \sum_{\omega \in B} f(\omega) = P(B)$$

$\square$

Definition 8.2.13. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum mit $|\Omega| \in \mathbb{N}^*$. P ist eine diskrete uniforme Wahrscheinlichkeitsverteilung genau dann, wenn für alle $\omega \in \Omega$ gilt: $P(\{\omega\}) = c$ ($c \in]0, 1]$ ist eine Konstante)

Damit sprechen wir von einer uniformen Wahrscheinlichkeitsverteilung genau dann, wenn die Wahrscheinlichkeitsfunktion allen Ergebnissen dieselbe Wahrscheinlichkeit zuordnet. Die uniforme Wahrscheinlichkeitsverteilung nennen wir auch „Gleichverteilung“. Viele Beispiele der einfachen Wahrscheinlichkeitstheorie (Würfel, Kartenspiele, etc.) setzen die gleiche Wahrscheinlichkeit der Ergebnisse voraus. In diesem Fall lassen sich die Wahrscheinlichkeiten für beliebige Ereignisse einfach berechnen:

Satz 8.2.14. Wenn P eine diskrete uniforme Wahrscheinlichkeitsverteilung ist, dann gilt:

$$P(A) = \frac{|A|}{|\Omega|}$$

Beweis. Da P eine diskrete uniforme Wahrscheinlichkeitsfunktion ist, gilt

$$P(\{\omega\}) = c = f(\omega)$$

für alle $\omega \in \Omega$. Nun gilt auf Grund der Definition der Wahrscheinlichkeitsfunktion:

$$P(\Omega) = \sum_{\omega \in \Omega} f(\omega) = 1$$

Da eine uniforme Verteilung vorliegt, gilt damit:

$$\sum_{\omega \in \Omega} f(\omega) = \sum_{\omega \in \Omega} c = |\Omega| c = 1.$$

Daraus folgt:

$$c = \frac{1}{|\Omega|}$$

Für $P(A)$ gilt damit:

$$\begin{aligned} P(A) &= \sum_{\omega \in A} f(\omega) \\ &= \sum_{\omega \in A} c \\ &= |A| c \\ &= |A| \frac{1}{|\Omega|}. \end{aligned}$$

□

$|\Omega|$ wird oft „Anzahl der möglichen Ergebnisse“ genannt. $|A|$ wird oft „Anzahl der günstigen Ergebnisse“ genannt. Mit dieser Sprechweise entspricht der eben festgehaltene Satz dem sogenannten Laplace-Wahrscheinlichkeitsbegriff (Pierre-Simon Laplace, 1749 - 1827, französischer Mathematiker, Physiker und Astronom):

$$P(A) = \frac{\text{Anzahl günstiger Ergebnisse}}{\text{Anzahl möglicher Ergebnisse}} = \text{relative Häufigkeit der günstigen Ergebnisse.}$$

Man beachte: diese Redensweise setzt die gleiche Wahrscheinlichkeit der Ergebnisse voraus! Damit erfasst diese Definition nur den - keineswegs umfassenden - Spezialfall der diskreten, uniformen Verteilung.

Bei grösseren Mengen erlauben es manchmal kombinatorische Hilfsmittel, $|A|$ sowie $|\Omega| = n$ zu berechnen.

Beispiel 8.2.15. 1) Wir nehmen an, die möglichen Ergebnisse eines Würfels seien gleich wahrscheinlich. Die Ergebnismenge weist damit 6 Ergebnisse auf, d.h. $|\Omega| = n = 6$. Das Ereignis, eine gerade Zahl zu würfeln, weist 3 Elemente auf: $\{2, 4, 6\}$. Somit ist $|A| = |\{2, 4, 6\}| = 3$. Damit gilt für die Wahrscheinlichkeit, eine gerade Zahl zu würfeln: $\frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$.

Die Wahrscheinlichkeit, 2 oder 4 zu erhalten, ist mit $A = \{2, 4\}$: $\frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}$.

2) Wir nehmen an, bei einer Münze sei die Wahrscheinlichkeit, Kopf oder Zahl zu werfen, identisch. Die Ergebnismenge erfasst zwei Elemente. Das Ereignis, Kopf zu werfen, umfasst ein Element. In diesem Fall ist damit die Wahrscheinlichkeit, Kopf zu werfen ($A = \{K\}$): $\frac{|A|}{|\Omega|} = \frac{1}{2}$.

3) Beispiel mit Kombinatorik: wie hoch ist die Wahrscheinlichkeit, bei vier Zügen ohne Zurücklegen aus einem Kartenset von 36 Karten mit 4 Typen Schaufel, Herz, Kreuz und Egge genau die Karten Schaufel König, Herz Dame, Herz Zehn, Kreuz Neun zu ziehen? Wir haben nur eine „günstige Möglichkeit“, womit $|A| = 1$. Zur Bestimmung von $|\Omega|$, gehen wir von der Überlegung aus, dass es um Ziehen ohne Reihenfolge geht. Es gibt also $\binom{36}{4}$ Möglichkeiten, vier Karten zu ziehen, d.h. $|\Omega| = \binom{36}{4} = 58\,905$. Jede dieser Möglichkeiten hat dieselbe Wahrscheinlichkeit, realisiert zu werden - nehmen wir vernünftiger Weise an. Damit beträgt die Wahrscheinlichkeit, genau diese vier Karten zu ziehen

$$\frac{|A|}{|\Omega|} = \frac{1}{\binom{36}{4}} = \frac{1}{58\,905} = 0.000016976$$

◇

Wahrscheinlichkeiten sind eine Art Modell für die relativen Häufigkeiten, wobei ein Modell zu einem Zufallsexperiment passt, wenn die relativen Häufigkeiten nicht zu stark vom Modell abweichen. Ein Modell kann angemessen sein, wenn ein Zufallsexperiment genügend oft durchgeführt zu relativen Häufigkeiten führt, die vom Modell nur unbedeutend abweichen. Betrachten wir das Beispiel der Münze. Wir werfen eine Münze 1000 mal. Wir erhalten 490 mal Kopf und 510 mal Zahl. Die relativen Häufigkeiten sind dann $\frac{490}{1000} = 0.49$ und $\frac{510}{1000} = 0.51$. Die Ergebnisse des Zufallsexperimentes scheinen nicht schlecht zum Modell $P(\{K\}) = 0.5$ und $P(\{Z\}) = 0.5$ zu passen. In diesem Falle ist es durchaus vernünftig, die gleiche Wahrscheinlichkeit der Ergebnisse vorauszusetzen. Man könnte aber auch Münzen machen, die auf einer Seite schwerer sind als auf der anderen. Die schwerere Seite wird häufiger nach unten schauen als nach oben - fällt ein Butterbrot auf den Boden, fällt es fast immer mit der schwereren Seite nach unten. Damit wäre das Postulat der gleichen Wahrscheinlichkeit der Ergebnisse nicht mehr angemessen. In den 70er Jahren wurde dem Schweizerischen Zahlenlotto nachgewiesen, dass nicht alle Kugeln mit gleicher Wahrscheinlichkeit gewählt werden. Es erwies sich, dass dies an den unterschiedlichen chemischen Zusammensetzungen der Farben lag, die für die Kugeln damals verwendet wurden. Die Farben hatten nicht dieselben elektrostatischen Eigenschaften, was sich auf die Wahrscheinlichkeit der Auswahl auswirkte.

Ist das Postulat gleicher Wahrscheinlichkeiten nicht haltbar, sind wir auf relative Häufigkeiten angewiesen, um Wahrscheinlichkeiten festzulegen: wir führen ein Zufallsexperiment genügend oft durch und wählen als Wahrscheinlichkeiten die gefundenen relativen Häufigkeiten. Trotzdem ist es wichtig, relative Häufigkeiten scharf von Wahrscheinlichkeiten zu unterscheiden. Relative Häufigkeiten sind empirische Kennwerte, während Wahrscheinlichkeiten theoretische, nicht-empirische Größen darstellen („empirisch“ heisst: mit Hilfe der Erfahrung, durch Erfahrung gewonnen).

Beispiel 8.2.16. Wir werfen einen Reissnagel auf eine, harte, glatte Unterlage. Dabei sind empirisch zwei Resultate möglich. Der Reissnagel fällt auf den Kopf (die Spitze ragt senkrecht nach oben) oder er kommt seitlich zu liegen (so dass die Spitze und der Rand des Kopfes aufliegen). Es ist hier kaum möglich, Gleichwahrscheinlichkeit der Ergebnisse vorauszusetzen. Nur ein Experiment könnte hier weiterführen, um zu einer näherungsweisen Bestimmung angemessener Wahrscheinlichkeiten zu kommen (s. Abbildung 8.2.1).

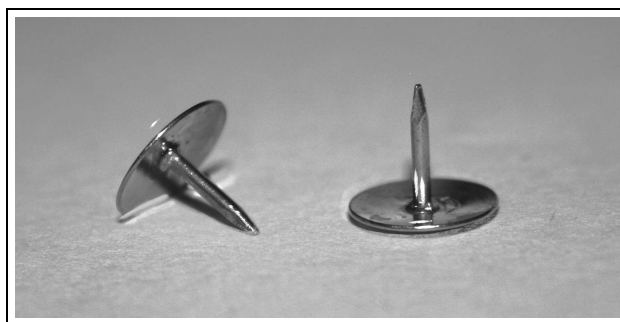


Abbildung 8.2.1: Beispiel für die möglichen Ergebnisse eines Zufallsexperimentes mit ungleichwahrscheinlichen Ausgängen

◇

8.2.1 Übungen

1. Zufälliges einmaliges Ziehen von Karten aus 36 Jasskarten (es wird jeweils eine uniforme Verteilung vorausgesetzt).
 - (a) Wie gross ist die Wahrscheinlichkeit, eine Herzkarte zu ziehen?
 - (b) Wie gross ist die Wahrscheinlichkeit, eine Zehn oder eine Acht zu ziehen?
2. Eine Urne enthält 9 weisse, 3 rote und 4 schwarze Kugeln. Es wird eine Kugel herausgenommen. Wie gross ist die Wahrscheinlichkeit, dass
 - (a) eine weisse Kugel gezogen wird?
 - (b) eine weisse oder schwarze Kugel gezogen wird?
 - (c) keine schwarze Kugel gezogen wird?
 (Wir setzen bei beiden Frage gleiche Wahrscheinlichkeiten der Ergebnisse voraus).
3. Berechnen Sie die Wahrscheinlichkeit, aus einem Spiel von 36 Karten ein As zu ziehen (Gleichverteilung vorausgesetzt).
4. In einer Lotterie sind 20'000 Lose mit 1400 Treffern. Welche Wahrscheinlichkeit, einen Treffer zu erhalten, hat der Besitzer eines Loses (Gleichverteilung vorausgesetzt)?
5. Wie gross ist die Wahrscheinlichkeit, beim Werfen eines Würfels mehr als 3 zu würfeln (Gleichverteilung vorausgesetzt)?
6. Finden Sie 2 Ergebnismengen, so dass für einen die Annahme der Gleichverteilung realistisch ist, für den anderen realistischer weise nicht.

8.2.2 Lösungen

1. (a) Die Ergebnismenge enthält 36 Elemente. Damit gilt $|\Omega| = n = 36$. Ein Spiel enthält 9 Herzkarten. Damit enthält das Ereignis A , eine Herzkarte zu ziehen, 9 Elemente. Damit gilt $P(A) = \frac{|A|}{|\Omega|} = \frac{9}{36} = \frac{1}{4}$
 (b) Die Ergebnismenge wie oben. Es gibt 4 Zehner und 4 Achter. Damit gilt: $|A| = 8$. Wir erhalten somit $P(A) = \frac{|A|}{|\Omega|} = \frac{8}{36} = \frac{2}{9} = 0.222\ 22$
2. (Voraussetzung: gleiche Wahrscheinlichkeit der Ergebnisse)
 - a) In der Urne befinden sich 16 Kugeln, d.h. $|\Omega| = 16$. Es hat 9 weisse Kugeln. Damit ist $|A| = 9$.
 $P(A) = \frac{9}{16} = 0.562\ 5$

- b) Ω wie oben. $|A| = 9 + 4 = 13$
 $P(A) = \frac{13}{16} = 0.8125$
 c) Ω wie oben. $|A| = 9 + 3$
 $P(A) = \frac{12}{16} = \frac{3}{4} = 0.75$
3. $P(A) = \frac{|A|}{|\Omega|} = \frac{4}{36} = \frac{1}{9}$ (Voraussetzung: gleiche Wahrscheinlichkeit der Ergebnisse)
4. $P(A) = \frac{|A|}{|\Omega|} = \frac{1400}{20000} = \frac{7}{100} = 0.07$ (Voraussetzung: gleiche Wahrscheinlichkeit der Ergebnisse)
5. Mehr als drei werden gewürfelt, wenn 4, 5 oder sechs gewürfelt werden. Damit ergibt sich bei gleicher Wahrscheinlichkeit der Ergebnisse: $P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$

8.2.3 Lernziele

- Den Zusammenhang von Wahrscheinlichkeiten und relativen Häufigkeiten wie im Text dargestellt beschreiben können.
- Die Definition der Wahrscheinlichkeitsfunktion hinschreiben können.
- Die Definition der Wahrscheinlichkeitsverteilung (= Wahrscheinlichkeitsmass) hinschreiben können und den Unterschied zwischen der Wahrscheinlichkeitsfunktion und der Wahrscheinlichkeitsverteilung erläutern können.
- Den Begriff des diskreten Wahrscheinlichkeitsraums erläutern können.
- Die drei Eigenschaften des Wahrscheinlichkeitsmasses hinschreiben können ((1) $P(\Omega) = 1$, (2) $P(B) \geq 0$ (für $B \subset \Omega$); (3) für $A_i \cap A_j = \emptyset$ gilt: $P(A_i \cap A_j) = P(A_j) + P(A_i)$)
- Die erwähnten Sätze kennen und in Rechnungen anwenden können, insbesondere die Sätze $P(A^c) = 1 - P(A)$ und $P(A) = 1 - P(A^c)$.
- Übungen von der obigen Art rechnen können.

8.3 Das Rechnen mit Wahrscheinlichkeiten

8.3.1 Addition von Wahrscheinlichkeiten

Die Wahrscheinlichkeit, dass entweder das eine oder das andere von zwei sich einander ausschließenden Ereignissen eintritt, errechnet sich durch die Addition der beiden Einzelwahrscheinlichkeiten (s. Satz 8.2.6(c))

Beispiel 8.3.1. (1) Es gelte beim Würfeln: A = Ereignis, eine gerade Zahl zu würfeln mit $P(A) = \frac{4}{6}$ (nicht uniforme Verteilung). A' = Ereignis, eine 1 zu würfeln mit $P(A') = \frac{1}{12}$. Gesucht ist die Wahrscheinlichkeit dafür, eine gerade Zahl oder 1 zu würfeln.

Da $A \cap A' = \emptyset$, gilt $P(A \cup A') = P(A) + P(A') = \frac{4}{6} + \frac{1}{12} = \frac{3}{4}$

(2) Es gelte beim Würfeln: A = Ereignis, eine gerade Zahl zu würfeln mit $P(A) = \frac{3}{6}$ (wir wissen nicht, ob uniform verteilt oder nicht). A' = Ereignis, eine 1 oder 2 zu würfeln mit $P(A') = \frac{2}{6}$. Gesucht ist die Wahrscheinlichkeit dafür, eine gerade Zahl oder 1 zu würfeln. Da $A \cap A' \neq \emptyset$, können wir die Wahrscheinlichkeit nicht bestimmen.

Wenn wir allerdings z.B. wüssten, dass $P(\{2\}) = \frac{1}{6}$, wüssten wir damit dass $P(\{1\}) = \frac{1}{6}$ (Da $\{1\} \cap \{2\} = \emptyset$ und $P(\{1\} \cup \{2\}) = \frac{2}{6} = P(\{1\}) + \frac{1}{6} \implies P(\{1\}) = \frac{1}{6}$. In diesem Fall gälte dann: $P(A \cup A') = P(A \cup \{1\}) = \frac{3}{6} + \frac{1}{6} = \frac{4}{6}$ (da $A \cap \{1\} = \emptyset$).

(3) Wie gross ist die Wahrscheinlichkeit, aus einer Gesamtkartei aller Erwachsenen in der Schweiz zufällig eine 20-29 jährige (Ereignis A) oder eine 30-39 jährige Person (Ereignis B) zu ziehen? (22,4% aller Erwachsenen sind 20-29jährig und 20,6 % 30-39 jährig). Die beiden Ereignisse

schliessen sich aus. Entsprechend ist die gesuchte Wahrscheinlichkeit: $\frac{22.4+20.6}{100} = 0.43$

(4) Wie gross ist die Wahrscheinlichkeit, aus einer Gesamtkartei aller Erwachsenen in der Schweiz zufällig eine 20-34 jährige (Ereignis A) oder eine 25-39 jährige Person (Ereignis B) zu ziehen? (32,4% aller Erwachsenen sind 20-34jährig und 31,6 % 25-39 jährig). Die beiden Ereignisse A und B schliessen sich nicht mehr aus. Entsprechend können wir die Wahrscheinlichkeiten ohne zusätzliche Angaben nicht berechnen. Dieses Problem führt uns zu den folgenden Überlegungen. $\diamond$

Überschneiden sich zwei Ereignisse A und B , so könnten wir die Wahrscheinlichkeit, dass A oder B eintritt offenbar berechnen mit:

Satz 8.3.2. (Additionssatz): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Beweis. (i) $A \cup B = A \setminus B \cup (A \cap B) \cup B \setminus A$ (siehe Abbildung 8.3.2)

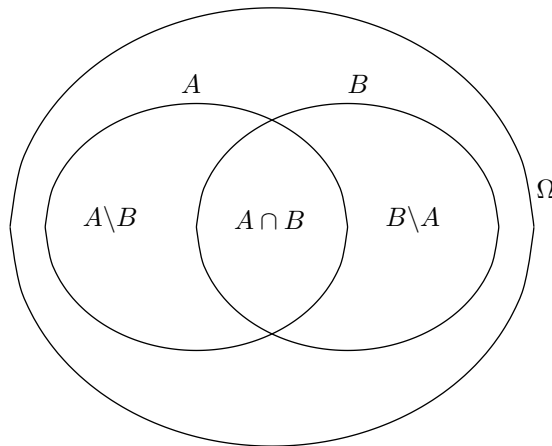


Abbildung 8.3.2: Skizze zum Beweis des Additionsgesetzes 8.3.2

Da Mengen $A \setminus B$, $(A \cap B)$ und $B \setminus A$ paarweise fremd sind, gilt:

(ii) $P(A \cup B) = P(A \setminus B) + P(A \cap B) + P(B \setminus A)$

(iii) Zudem gilt: $P(A \setminus B) = P(A) - P(A \cap B)$ und $P(B \setminus A) = P(B) - P(A \cap B)$.

(iv) Setzen wir (iii) in (ii) ein, erhalten wir:

$$P(A \cup B) = P(A \setminus B) + P(A \cap B) + P(B \setminus A) =$$

$$P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) = P(A) + P(B) - P(A \cap B). \quad \square$$

Beachten Sie: Aus der Formel folgt: wenn $A \cap B = \emptyset$, dann $P(A \cup B) = P(A) + P(B)$, denn aus $A \cap B = \emptyset$ folgt gemäss bereits bewiesenem Satz $P(A \cap B) = 0$.

Beispiel 8.3.3. Wir nehmen an, wir hätten einen Würfel mit 7 Seiten, der unregelmässig ist, wobei man auf die Ergebnismenge den folgenden Graph der Wahrscheinlichkeitsfunktion festlegen kann: $f = \{(1, \frac{1}{6}), (2, \frac{1}{6}), (3, \frac{1}{7}), (4, \frac{1}{7}), (5, \frac{1}{5}), (6, \frac{1}{10}), (7, \frac{17}{210})\}$

Wir überprüfen zuerst, ob es sich tatsächlich um eine Wahrscheinlichkeitsfunktion handelt: $\frac{1}{6} + \frac{1}{6} + \frac{1}{7} + \frac{1}{7} + \frac{1}{5} + \frac{1}{10} + \frac{17}{210} = 1$.

Es gilt nun für $A = \{3, 4, 5\}$ und $B = \{4, 5, 6, 7\}$

$$P(A) = \frac{1}{7} + \frac{1}{7} + \frac{1}{5} = \frac{17}{35} \text{ und } P(B) = \frac{1}{7} + \frac{1}{5} + \frac{1}{10} + \frac{17}{210} = \frac{11}{21}$$

$$\text{Zudem gilt } P(A \cap B) = P(\{4, 5\}) = \frac{1}{7} + \frac{1}{5} = \frac{12}{35}$$

$$\text{Damit gilt gemäss Satz: } P(A \cup B) = \frac{17}{35} + \frac{11}{21} - \frac{12}{35} = \frac{2}{3}$$

$$\text{Wir überprüfen: } A \cup B = \{3, 4, 5, 6, 7\} \text{ und } P(A \cup B) = \frac{1}{7} + \frac{1}{7} + \frac{1}{5} + \frac{1}{10} + \frac{17}{210} = \frac{2}{3}. \quad \diamond$$

8.3.2 Die Multiplikation von Wahrscheinlichkeiten

Multiplikation bei wiederholten Zufallsexperimenten

Oft betrachten wir mehrere Zufallsexperimente hintereinander. So können wir zweimal eine Münze werfen, wobei ein Münzwurf als ein Zufallsexperiment gilt. Wir können uns dann z.B. fragen, wie hoch die Wahrscheinlichkeit ist, zuerst Kopf zu werfen und dann eine Zahl, wenn auf die Ergebnismengen der einzelnen Wurfexperimente eine Gleichverteilung definiert ist.

Wir untersuchen zuerst die Frage, was bei wiederholten Zufallsexperimenten als Ergebnismenge zu wählen ist. Die möglichen Ergebnisse beim zweimaligen Wiederholen des Wurfszufallsexperimentes ergeben offensichtlich die folgende Ergebnismenge:

$$\Omega = \Omega_1 \times \Omega_1 = \{(K, K), (K, Z), (Z, K), (Z, Z)\} \text{ mit } \Omega_1 = \{K, Z\}$$

Wir erhalten bei zweimaligem Zufallsexperiment mit Ergebnismenge Ω_1 fürs einzelne Zufallsexperiment als neuer Ergebnismenge das kartesische Produkt $\Omega_1 \times \Omega_1$. Wir nennen $\Omega = \Omega_1 \times \Omega_1$ „Produkt von Ergebnismengen“. Bei n -maligem Wiederholen desselben Zufallsexperimentes erhalten wir als Ergebnismenge $\Omega = \Omega_1^n$. So würden wir bei dreifachem Werfen mit $\Omega_1 = \{K, Z\}$ als neue Ergebnismenge $\Omega = \Omega_1 \times \Omega_1 \times \Omega_1 =$

$$\{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), (K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\} \text{ erhalten.}$$

Im Beispiel verwendeten wir jeweils das gleiche Zufallsexperiment. Man könnte aber auch einen Würfelexperiment und ein Münzexperiment betrachten und sich die Frage stellen, wie wahrscheinlich es ist, eine 3 zu würfeln und Kopf zu werfen. Wir erhalten als „neue“ Ergebnismenge mit $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$ und $\Omega_2 = \{K, Z\}$:

$$\Omega_1 \times \Omega_2 = \{(1, K), (2, K), (3, K), (4, K), (5, K), (6, K), (1, Z), (2, Z), (3, Z), (4, Z), (5, Z), (6, Z)\}.$$

Allgemein erhalten wir also bei n -maligem Zufallsexperiment als Ergebnismenge $\Omega = \bigtimes_{i=1}^n \Omega_i$ mit n Ergebnismengen Ω_i . Auf die Ergebnismenge $\bigtimes_{i=1}^n \Omega_i$ können nun Wahrscheinlichkeitsmasse definiert werden, wobei wir - wenn wir einen Zusammenhang mit den Massen auf den einzelnen Ω_i wahren wollen, nicht alle berücksichtigen. Wir definieren für den zweidimensionalen Fall:

Definition 8.3.4. Seien 2 diskrete Ergebnismengen Ω_i gegeben. Dann nennen wir eine Wahrscheinlichkeitsfunktion $f_{\Omega_1 \times \Omega_2} : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ „gemeinsame Wahrscheinlichkeitsfunktion von f_{Ω_1} und f_{Ω_2} auf $\Omega_1 \times \Omega_2$ “ genau dann, wenn

$$f_{\Omega_1}(\omega_i) = \sum_{\omega_j \in \Omega_2} f_{\Omega_1 \times \Omega_2}(\omega_i, \omega_j) \quad (\text{für alle } \omega_i \in \Omega_1)$$

$$f_{\Omega_2}(\omega_j) = \sum_{\omega_i \in \Omega_1} f_{\Omega_1 \times \Omega_2}(\omega_i, \omega_j) \quad (\text{für alle } \omega_j \in \Omega_2).$$

Wir nennen in diesem Zusammenhang f_{Ω_1} und f_{Ω_2} die Rand-W-Funktionen von $f_{\Omega_1 \times \Omega_2}$. Diese bestimmen sogenannte Randverteilungen, nämlich die Wahrscheinlichkeitsmasse P_{Ω_1} und P_{Ω_2} , die durch f_{Ω_1} und f_{Ω_2} bestimmt sind. Die durch $f_{\Omega_1 \times \Omega_2}$ definierte Wahrscheinlichkeitsverteilung $\mathcal{P}(\Omega_1 \times \Omega_2) \rightarrow [0, 1]$ nennen wir „gemeinsame Wahrscheinlichkeitsverteilung“ oder „gemeinsames Wahrscheinlichkeitsmass von P_{Ω_1} und P_{Ω_2} auf $\Omega_1 \times \Omega_2$ “.

Die Definition kann zwanglos auf $\bigtimes_{i=1}^n \Omega_i$ erweitert werden. Die Idee hinter der Definition ist die folgende: $f_{\Omega_1}(\omega_i)$ ist die Wahrscheinlichkeit, dass sich ω_i ereignet, wobei sich ω_i ereignet, wenn sich ein geordnetes Paar von $\Omega_1 \times \Omega_2$ ereignet, das an erster Stelle ω_i enthält. Damit ereignet sich ω_i , wenn sich (ω_i, ω_j) mit $\omega_j \in \Omega_2$ ereignet. Es ist also die Summe der Wahrscheinlichkeiten $f_{\Omega_1 \times \Omega_2}(\omega_i, \omega_j)$ über alle $\omega_j \in \Omega_2$ zu bilden.

Beispiel 8.3.5. Sei $\Omega_1 = \{a, b\}, \Omega_2 = \{c, d, e\}$ mit $f_{\Omega_1}(a) = \frac{1}{3}, f_{\Omega_1}(b) = \frac{2}{3}; f_{\Omega_2}(c) = \frac{2}{10}, f_{\Omega_2}(d) = \frac{5}{10}, f_{\Omega_2}(e) = \frac{3}{10}$.

Dann ist eine Gemeinsame W -Funktion von f_{Ω_1} und f_{Ω_2} auf $\Omega_1 \times \Omega_2$ gegeben durch

$$\begin{aligned} f_{\Omega_1 \times \Omega_2}((a, c)) &= \frac{1}{10} \\ f_{\Omega_1 \times \Omega_2}((a, d)) &= \frac{2}{10} \\ f_{\Omega_1 \times \Omega_2}((a, e)) &= \frac{1}{30} \\ f_{\Omega_1 \times \Omega_2}((b, c)) &= \frac{1}{10} \\ f_{\Omega_1 \times \Omega_2}((b, d)) &= \frac{3}{10} \\ f_{\Omega_1 \times \Omega_2}((b, e)) &= \frac{4}{15} \end{aligned}$$

Es gilt nämlich:

$$\begin{aligned} f_{\Omega_1}(a) &= f_{\Omega_1 \times \Omega_2}((a, c)) + f_{\Omega_1 \times \Omega_2}((a, d)) + f_{\Omega_1 \times \Omega_2}((a, e)) = \frac{1}{10} + \frac{2}{10} + \frac{1}{30} = \frac{1}{3}. \\ f_{\Omega_1}(b) &= f_{\Omega_1 \times \Omega_2}((b, c)) + f_{\Omega_1 \times \Omega_2}((b, d)) + f_{\Omega_1 \times \Omega_2}((b, e)) = \frac{1}{10} + \frac{3}{10} + \frac{4}{15} = \frac{2}{3}. \\ f_{\Omega_2}(c) &= f_{\Omega_1 \times \Omega_2}((a, c)) + f_{\Omega_1 \times \Omega_2}((b, c)) = \frac{1}{10} + \frac{1}{10} = \frac{2}{10}. \\ f_{\Omega_2}(d) &= f_{\Omega_1 \times \Omega_2}((a, d)) + f_{\Omega_1 \times \Omega_2}((b, d)) = \frac{2}{10} + \frac{3}{10} = \frac{5}{10}. \\ f_{\Omega_2}(e) &= f_{\Omega_1 \times \Omega_2}((a, e)) + f_{\Omega_1 \times \Omega_2}((b, e)) = \frac{1}{30} + \frac{4}{15} = \frac{3}{10}. \end{aligned}$$

Im vorliegenden zweidimensionalen Fall lässt sich das Kriterium für Gemeinsamkeit der W -Funktion einfach in einer Kreuztabelle überprüfen: Es liegt genau dann eine Gemeinsame W -Funktion und damit eine Gemeinsame Verteilung vor, wenn die Randsummen der Zellen der folgenden Kreuztabelle mit den Werten der W -Funktionen f_{Ω_1} und f_{Ω_2} übereinstimmen:

	a	b	Randsummen
c	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$
d	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$
e	$\frac{1}{30}$	$\frac{4}{15}$	$\frac{3}{10}$
Randsummen	$\frac{1}{3}$	$\frac{2}{3}$	1

◇

Es gibt überabzählbar unendlich viele gemeinsame Verteilungen auf $\mathcal{P}(\Omega_1 \times \dots \times \Omega_n)$, da es überabzählbar unendlich viele Möglichkeiten gibt, mit Hilfe einer Wahrscheinlichkeitsfunktion den Elementen von $\Omega_1 \times \dots \times \Omega_n$ Zahlen zwischen 0 und 1 zuzuordnen, so dass deren Randsummen mit gegebenen Randverteilungen übereinstimmen.

Wir wenden uns in der Folge einer spezifischen Gemeinsamen Verteilung zu, die bedeutsam ist, weil sie einer häufig angetroffenen Situation entspricht und weil diese Verteilung aus den Verteilungen auf die einzelnen Ergebnismengen - d.h. aus den Randverteilungen - berechnet werden kann. Beim zweifachen Münzwurf können wir davon ausgehen, dass das Eintreffen eines Ergebnisses im ersten Münzwurf die Wahrscheinlichkeiten des Eintreffens von Ergebnissen im zweiten Münzwurf nicht beeinflusst. In diesem Fall kann die gemeinsame Verteilung auf

$$\{(K, K), (K, Z), (Z, K), (Z, Z)\} = \Omega_1 \times \Omega_1$$

durch Multiplikation mit Hilfe der Verteilung auf $\{K, Z\}$ berechnet werden. Gehen wir nämlich z.B. von einer Gleichverteilung auf $\{K, Z\}$ aus, ist offenbar auf $\Omega_1 \times \Omega_1$ ebenfalls eine Gleichverteilung definiert, d.h. es gilt

$$f_{\Omega_1 \times \Omega_1}((K, Z)) = \frac{1}{4}.$$

Zudem können wir beobachten, dass offenbar

$$f_{\Omega_1 \times \Omega_1}((K, Z)) = f_{\Omega_1}(K) \cdot f_{\Omega_1}(Z) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

gilt. Wir erhalten $f_{\Omega_1 \times \Omega_1}((K, Z))$ also durch die Multiplikation von $f_{\Omega_1}(K)$ und $f_{\Omega_1}(Z)$.

Die Multiplikation von Wahrscheinlichkeiten der Ergebnisse aus Ω_1 und Ω_2 ist nur eine geeignete Operation, wenn wir garantieren können, dass dadurch auf $\Omega_1 \times \Omega_2$ wiederum eine Wahrscheinlichkeitsfunktion entsteht, und zwar eine Gemeinsame. Dies ist die Aussage des folgenden Satzes:

Satz 8.3.6. *Seien zwei Wahrscheinlichkeitsfunktionen f_{Ω_1} und f_{Ω_2} gegeben. Die durch*

$$\begin{aligned} f_{\Omega_1 \times \Omega_2} : \Omega_1 \times \Omega_2 &\rightarrow [0, 1] \\ f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j)) &= f_{\Omega_1}(\omega_i) \cdot f_{\Omega_2}(\omega_j) \quad (\omega_i \in \Omega_1, \omega_j \in \Omega_2) \end{aligned}$$

definierte Abbildung $f_{\Omega_1 \times \Omega_2}$ ist eine Gemeinsame Wahrscheinlichkeitsfunktion von f_{Ω_1} und f_{Ω_2} auf $\Omega_1 \times \Omega_2$.

Beweis. Wir zeigen zuerst, dass die Summe der Wahrscheinlichkeit 1 ergibt: $1 = \sum_{\omega_i \in \Omega_1} f_{\Omega_1}(\omega_i) \cdot$

$\sum_{\omega_j \in \Omega_2} f_{\Omega_2}(\omega_j) = \sum_{\omega_i \in \Omega_1} \left(\sum_{\omega_j \in \Omega_2} f_{\Omega_1}(\omega_i) f_{\Omega_2}(\omega_j) \right) = \sum_{\omega_i \in \Omega_1} \sum_{\omega_j \in \Omega_2} (f_{\Omega_1}(\omega_i) f_{\Omega_2}(\omega_j)).$
Es bleibt noch zu zeigen, dass $f_{\Omega_1}(\omega_i) = \sum_{\omega_j \in \Omega_2} f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j))$. Es gilt $\sum_{\omega_j \in \Omega_2} f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j)) = \sum_{\omega_j \in \Omega_2} f_{\Omega_1}(\omega_i) \cdot f_{\Omega_2}(\omega_j) = f_{\Omega_1}(\omega_i) \sum_{\omega_j \in \Omega_2} f_{\Omega_2}(\omega_j) = f_{\Omega_1}(\omega_i) \cdot 1 = f_{\Omega_1}(\omega_i)$. Dies gilt für alle $\omega_i \in \Omega_1$ und kann analog für alle $\omega_j \in \Omega_2$ die Aussage $f_{\Omega_2}(\omega_j) = \sum_{\omega_i \in \Omega_1} f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j))$ bewiesen werden. $\square$

Der Satz kann auf das Produkt von n Ergebnismengen verallgemeinert werden (für die entsprechende Formulierung s. Lösungen zu den Übungen 8.3.3.21 (a)).

Definition 8.3.7. Seien zwei diskrete Ergebnismengen Ω_1 und Ω_2 mit zwei Wahrscheinlichkeitsfunktionen f_{Ω_1} und f_{Ω_2} gegeben. Wir nennen die gemeinsame Wahrscheinlichkeitsfunktion $f_{\Omega_1 \times \Omega_2}$ mit

$$\begin{aligned} f_{\Omega_1 \times \Omega_2} : \Omega_1 \times \Omega_2 &\rightarrow [0, 1] \\ f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j)) &= f_{\Omega_1}(\omega_i) \cdot f_{\Omega_2}(\omega_j), \omega_i \in \Omega_1, \omega_j \in \Omega_2 \end{aligned}$$

die Produktwahrscheinlichkeitsfunktion auf $\Omega_1 \times \Omega_2$.

(Diese Namengebung gilt auch für die durch Multiplizieren der Wahrscheinlichkeiten erzeugte Wahrscheinlichkeitsfunktion auf das Produkt von n Ergebnismengen)

Ereignisse aus $\Omega_1 \times \Omega_2$ sind wiederum definiert als Teilmengen von $\Omega_1 \times \Omega_2$. Deren Wahrscheinlichkeiten sind durch die Wahrscheinlichkeitsfunktion $f_{\Omega_1 \times \Omega_2}$ bestimmt:

Definition 8.3.8. Seien zwei diskrete Ergebnismengen Ω_1 und Ω_2 mit zwei Wahrscheinlichkeitsfunktionen f_{Ω_1} und f_{Ω_2} gegeben. Das durch die Produktwahrscheinlichkeitsfunktion $f_{\Omega_1 \times \Omega_2}$ definierte Wahrscheinlichkeitsmass $P_{\Omega_1 \times \Omega_2}$ mit

$$\begin{aligned} P_{\Omega_1 \times \Omega_2} : \mathcal{P}(\Omega_1 \times \Omega_2) &\rightarrow [0, 1] \\ P_{\Omega_1 \times \Omega_2}(A) &= \sum_{(\omega_i, \omega_j) \in A} f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j)) \quad \text{mit } A \subset \Omega_1 \times \Omega_2 \end{aligned}$$

nennen wir Produktmass von P_1 und P_2 .

Diese Begriffe können zwanglos auf das Produkt von n Ergebnismengen verallgemeinert werden (für die entsprechende Formulierung s. Lösungen zu den Übungen 8.3.3.21 (b)).

Beispiel 8.3.9. (i) *Wie hoch ist die Wahrscheinlichkeit, beim zweifachen Münzwurf zweimal Kopf oder zweimal Zahl zu werfen, wenn $f_{\Omega_1}(K) = 0.2$?*

$$\Omega_1 = \{K, Z\}. \quad \Omega_1 \times \Omega_1 = \{(K, K), (K, Z), (Z, K), (Z, Z)\}.$$

Das interessierende Ereignis ist

$$\{(K, K), (Z, Z)\}.$$

Wegen der fehlenden gegenseitigen Beeinflussung gilt:

$$\begin{aligned} f_{\Omega_1 \times \Omega_1}((K, K)) &= f_{\Omega_1}(K)f_{\Omega_1}(K) = 0.2 \cdot 0.2 = 0.04 \\ f_{\Omega_1 \times \Omega_1}((Z, Z)) &= f_{\Omega_1}(Z)f_{\Omega_1}(Z) = 0.8 \cdot 0.8 = 0.64. \end{aligned}$$

Damit gilt dann

$$\begin{aligned} P_{\Omega_1 \times \Omega_1}(\{(K, K), (Z, Z)\}) &= f_{\Omega_1 \times \Omega_1}((K, K)) + f_{\Omega_1 \times \Omega_1}((Z, Z)) \\ &= 0.04 + 0.64 \\ &= 0.68 \end{aligned}$$

(ii) Wie hoch ist die Wahrscheinlichkeit, beim dreifachen Münzwurf mindestens einmal Kopf zu werfen, wenn $f_{\Omega_1}(K) = 0.3$? Es ist $\Omega_1 = \{K, Z\}$ und

$$\Omega_1 \times \Omega_1 \times \Omega_1 = \{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), (K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\}.$$

Das interessierende Ereignisse ist

$$A := \{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), (K, Z, Z), (Z, K, Z), (Z, Z, K)\},$$

wobei dieses das Gegenereignis von $\{(Z, Z, Z)\}$ ist. Wir berechnen somit effizienter

$$\begin{aligned} P_{\Omega_1 \times \Omega_1 \times \Omega_1}(\{(Z, Z, Z)\}) &= f_{\Omega_1 \times \Omega_1 \times \Omega_1}((Z, Z, Z)) \\ &= (f_{\Omega_1}(Z))^3 \\ &= 0.7^3 \\ &= 0.343. \end{aligned}$$

Damit gilt dann:

$$P_{\Omega_1 \times \Omega_1 \times \Omega_1}(A) = 1 - 0.343 = 0.657.$$

(Wir setzen dabei fehlende gegenseitige Beeinflussung der Ergebnisse voraus). $\diamond$

Für Ereignisse aus $\Omega_1 \times \Omega_2$ gilt bezüglich des Produktemasses auf $\Omega_1 \times \Omega_2$:

Satz 8.3.10. Sei das Produktemass $P_{\Omega_1 \times \Omega_2}$ von P_1 und P_2 auf die diskrete Ergebnismenge $\Omega_1 \times \Omega_2$ gegeben und $A \subset \Omega_1$ sowie $B \subset \Omega_2$. Dann gilt: $P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1}(A) \cdot P_{\Omega_2}(B)$

$$\begin{aligned} \text{Beweis. } P_{\Omega_1 \times \Omega_2}(A \times B) &= \sum_{(\omega_i, \omega_j) \in A \times B} f_{\Omega_1 \times \Omega_2}((\omega_i, \omega_j)) = \sum_{(\omega_i, \omega_j) \in A \times B} f_{\Omega_1}(\omega_i) f_{\Omega_2}(\omega_j) = \\ &= \sum_{\omega_i \in A} \sum_{\omega_j \in B} f_{\Omega_1}(\omega_i) f_{\Omega_2}(\omega_j) = \sum_{\omega_i \in A} f_{\Omega_1}(\omega_i) \sum_{\omega_j \in B} f_{\Omega_2}(\omega_j) = P_{\Omega_1}(A) \cdot P_{\Omega_2}(B). \quad \square \end{aligned}$$

Der Satz kann auf das Produktemass von n Ergebnismengen verallgemeinert werden (s. für eine allgemeine Formulierung des Satzes die Lösung 21(c) der 8.3.3).

Definition 8.3.11. Ereignisse A und B beeinflussen sich gegenseitig nicht genau dann, wenn $P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1}(A) \cdot P_{\Omega_2}(B)$.

Eine allgemeine Formulierung ergibt sich zwanglos (für die entsprechende Formulierung s. Lösungen zu den Übungen 8.3.3.21 (d)). In der Definition ist „Ereignisse beeinflussen sich gegenseitig nicht“ als abkürzende Redensweise für „Das Eintreffen des Ereignisses A verändert die

Wahrscheinlichkeit des Eintreffens des Ereignisses B nicht und das Eintreffen von B verändert die Wahrscheinlichkeit des Eintreffens von A nicht“ zu verstehen.

Wie ausgeführt, stellt das Produktmass eine *spezifische gemeinsame Wahrscheinlichkeitsverteilung* auf das Produkt von Ergebnismengen dar. Das Produktmass nimmt in der Wahrscheinlichkeitstheorie und in der Statistik eine zentrale Rolle ein. Es hat den Vorteil, dass die gemeinsame Verteilung aus der Verteilung der einzelnen Ergebnismengen berechnet werden kann. Dabei muss beachtet werden, dass es durchaus Situationen gibt, wo die Voraussetzung der fehlenden gegenseitigen Beeinflussung nicht sinnvoll ist. So wird bei ansteckenden Krankheiten die Wahrscheinlichkeit, einen Krankheitsfall in einer Population zu finden, durch das Auffinden eines Krankheitsfalles beeinflusst. Die Wahrscheinlichkeit, dass jemand eine Information hat, wird durch das Vorkommen dieser Information bei Bekannten dieser Person beeinflusst. Wir vergleichen noch Produktmasse mit anderen gemeinsamen Verteilungen mit Hilfe des nächsten Beispiels:

Beispiel 8.3.12. $\Omega_1 = \{1, 2\}$, Graph von $f_{\Omega_1} = \{(1, 0.4), (2, 0.6)\}$

$\Omega_2 = \{a, b, c\}$. Graph von $f_{\Omega_2} = \{(a, 0.2), (b, 0.3), (c, 0.5)\}$

$\Omega_1 \times \Omega_2 = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, c)\}$

Graph der Wahrscheinlichkeitsfunktion f' einer möglichen Verteilung auf $\Omega_1 \times \Omega_2$:

$\{((1, a), 0.1), ((1, b), 0.2), ((1, c), 0.3), ((2, a), 0.05), ((2, b), 0.05), ((2, c), 0.3)\}$

Probe: $0.1 + 0.2 + 0.3 + 0.05 + 0.05 + 0.3 = 1$

Dieser Graph ist nicht der Graph der W -Funktion einer Gemeinsamen Verteilung auf $\Omega_1 \times \Omega_2$ bezüglich f_{Ω_1} und f_{Ω_2} , da z.B.

	a	b	c	Summen
1	0.1			
2	0.05			
Summen	0.15			

und $f_{\Omega_2}(a) = 0.2$. Man kann allerdings für f' W -Funktionen f'_{Ω_1} und f'_{Ω_2} finden, so dass f' eine Gemeinsame W -Funktion von f'_{Ω_1} und f'_{Ω_2} ist. Es genügt die Rand- W -Funktionen von f' zu berechnen.

Der folgende Graph der Wahrscheinlichkeitsfunktion f'' mit

$\{(1, a, 0.15), (1, b, 0.1), (1, c, 0.15), (2, a, 0.05), (2, b, 0.2), (2, c, 0.35)\}$

ist eine Gemeinsame W -Funktion auf $\Omega_1 \times \Omega_2$ bezüglich f_{Ω_1} und f_{Ω_2} , da

	a	b	c	Summen
1	0.15	0.1	0.15	0.4
2	0.05	0.2	0.35	0.6
Summen	0.2	0.3	0.5	1

Die dadurch gegebene gemeinsame Verteilung ist aber nicht das Produktmass, da z.B. $0.4 \cdot 0.2 = 0.08 \neq 0.15$ (Zelle $(1, 1)$)

Graph der Produkte- W -Funktion wäre vielmehr:

$\{((1, a), 0.08), ((1, b), 0.12), ((1, c), 0.2), ((2, a), 0.12), ((2, b), 0.18), ((2, c), 0.3)\}$, da die Zahlen in den Zellen immer das Produkt der entsprechenden Randsummen sind:

	a	b	c	Summen
1	0.08	0.12	0.2	0.4
2	0.12	0.18	0.3	0.6
Summen	0.2	0.3	0.5	1

Nicht jede Gemeinsame W -Verteilung zweier Ergebnismengen fällt also mit dem Produktmass zusammen. $\diamond$

Es folgen noch ein paar Beispiele für die Verwendung von Produktmassen:

Beispiel 8.3.13. Wie gross ist die Wahrscheinlichkeit, dass beim Werfen zweier Würfel, der erste die 6 und der zweite Würfel die 5 zeigen? (Fehlende gegenseitige Beeinflussung der Ereignisse bei wiederholtem Zufallsexperiment mit $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$).

Sei $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$ die Ergebnismenge fürs Werfen mit einem Würfel. Damit gilt $\Omega = \Omega_1 \times \Omega_1 = \{(i, j) : i, j \in \mathbb{N}_6^*\}$. $A = \{(6, 5)\}$. Zudem sei P_1 das auf Ω_1 definierte Wahrscheinlichkeitsmass (Gleichverteilung), f_{Ω_1} die auf Ω_1 definierte Wahrscheinlichkeitsfunktion und P_Ω das Produktmass auf Ω mit der Produktwahrscheinlichkeitsfunktion f_Ω auf Ω . Wegen des Fehlens gegenseitiger Beeinflussung gilt: $P_\Omega(A) = f_\Omega((6, 5)) = f_{\Omega_1}(6)f_{\Omega_1}(5) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$.

Wir betrachten noch das folgende Ereignis $B := \{(2, 3), (5, 6), (2, 6), (5, 3)\}$. Damit gilt:

$$\begin{aligned} P_\Omega(B) &= f_\Omega((2, 3)) + f_\Omega((5, 6)) + f_\Omega((2, 6)) + f_\Omega((5, 3)) = \\ &= f_{\Omega_1}(2)f_{\Omega_1}(3) + f_{\Omega_1}(5)f_{\Omega_1}(6) + f_{\Omega_1}(2)f_{\Omega_1}(6) + f_{\Omega_1}(5)f_{\Omega_1}(3) = \\ &= \left(\frac{1}{6} \cdot \frac{1}{6}\right) + \left(\frac{1}{6} \cdot \frac{1}{6}\right) + \left(\frac{1}{6} \cdot \frac{1}{6}\right) + \left(\frac{1}{6} \cdot \frac{1}{6}\right) = \frac{1}{9}. \end{aligned}$$

◇

Beispiel 8.3.14. Wir werfen fünf mal eine Münze. Wie hoch ist die Wahrscheinlichkeit, 3 mal Zahl zu werfen, wenn $f_{\Omega_1}(K) = 0.4$? (Fehlende gegenseitige Beeinflussung der Ereignisse bei wiederholtem Zufallsexperiment mit $\Omega_1 = \{K, Z\}$). Das Produkt der Ergebnismengen ist $\Omega := \Omega_1^5$. Wir betrachten ein spezielles Ergebnis aus Ω , das genau drei mal Zahl enthält, nämlich (Z, Z, Z, K, K) .

Dann gilt wegen der fehlenden gegenseitigen Beeinflussung: $f_\Omega((Z, Z, Z, K, K)) = (f_{\Omega_1}(Z))^3 (f_{\Omega_1}(K))^2 = 0.6^3 0.4^2 = 0.03456$. Es stellt sich nun die Frage, wie viele solcher Ergebnisse es gibt. Mit Hilfe von Kombinatorik erhalten wir $\binom{5}{3} = 10$. Alle haben dieselbe Wahrscheinlichkeit. Damit erhalten wir die Wahrscheinlichkeit des Ereignisses A , genau drei mal Zahl zu würfeln: $P_\Omega(A) = 10 \cdot 0.03456 = 0.3456$

◇

Beispiel 8.3.15. (1) Wir betrachten drei Urnen. Die erste enthalte 50 Steine, die von 1 bis 50 durchnummeriert sind. Die zweite enthalte 35 Steine, die von 100 bis 134 durchnummeriert sind, die dritte enthalte 10 Steine, die von 1001 bis 1010 durchnummeriert sind. Wir setzen jeweils gleiche Wahrscheinlichkeiten der Ergebnisse voraus. Wir ziehen aus jeder Urne einen Stein. $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$. Wieviel beträgt die Wahrscheinlichkeit des Ereignisses $\{(25, 104, 1002)\}$? $P_\Omega(\{(25, 104, 1002)\}) = f_\Omega((25, 104, 1002)) = f_{\Omega_1}(25)f_{\Omega_2}(104)f_{\Omega_3}(1002) = \frac{1}{50} \cdot \frac{1}{35} \cdot \frac{1}{10} = \frac{1}{17500}$ (Es ist vernünftig, davon auszugehen, dass die Ergebnisse der verschiedenen Züge sich gegenseitig nicht beeinflussen.)

(2) Es werde aus einer Urne mit 99 Steinen (mit 99 verschiedenen Zahlen i : $i \in \mathbb{N}_{99}^*$) zweimal je ein Stein gezogen (mit Zurücklegen, gutes Durchmischen nach jedem Zug). Wie hoch ist die Wahrscheinlichkeit, eine 3 und eine 4 zu erhalten (Reihenfolge spielt keine Rolle, auf Ω_i sind Gleichverteilungen definiert). $\Omega_1 = \mathbb{N}_{99}^*$. $\Omega = \Omega_1 \times \Omega_1$.

$$A = \{(3, 4), (4, 3)\}.$$

$$\begin{aligned} P_\Omega(A) &= f_\Omega((3, 4)) + f_\Omega((4, 3)) = f_{\Omega_1}(3)f_{\Omega_1}(4) + f_{\Omega_1}(4)f_{\Omega_1}(3) = \\ &= \left(\frac{1}{99} \cdot \frac{1}{99}\right) + \left(\frac{1}{99} \cdot \frac{1}{99}\right) = \frac{2}{9801} \text{ (wir setzen wiederum voraus, dass sich die Ergebnisse bei wiederholten Zufallsexperimenten nicht beeinflussen).} \end{aligned}$$

(3) Wir nehmen (kontrafaktisch) an, dass es genau so wahrscheinlich ist, einen Knaben wie ein Mädchen zu gebären und dass die entsprechenden Ergebnisse sich nicht beeinflussen, wenn zwei Kinder geboren werden. Für den Zweikinderfall erhalten wir die Ergebnismenge $\Omega = \{K, M\} \times \{K, M\} = \{(K, K), (K, M), (M, K), (M, M)\}$. Wie hoch ist die Wahrscheinlichkeit, einen Knaben und ein Mädchen unabhängig von der Reihenfolge auf die Welt zu bringen? Es gilt $A = \{(K, M), (M, K)\}$ und $P_\Omega(A) = f_\Omega((M, K)) + f_\Omega((K, M)) = f_{\Omega_1}(M)f_{\Omega_1}(K) + f_{\Omega_1}(K)f_{\Omega_1}(M) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$.

◇

Multiplikation und bedingte Wahrscheinlichkeiten

Unter „bedingter Wahrscheinlichkeit“ versteht man die Wahrscheinlichkeit eines Ereignisses unter der Bedingung, dass ein anderes Ereignis eingetreten ist. Wie hoch ist z.B. die Wahrscheinlichkeit, eine 5 gewürfelt zu haben, wenn man weiss, dass man eine ungerade Zahl gewürfelt hat?

Definition 8.3.16. Sei der diskrete Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ mit $A_i \subset \Omega$ und mit $P(A_1) > 0$ gegeben. Dann wird definiert:

$$P(A_2|A_1) := \frac{P(A_2 \cap A_1)}{P(A_1)}.$$

Wir lesen „ $P(A_2|A_1)$ “ als „die Wahrscheinlichkeit von A_2 unter der Bedingung dass A_1 “.

Die Formel kann man wie folgt verstehen: wir messen die Wahrscheinlichkeit, dass das Ereignis $A_2 \cap A_1$ eintritt an der Wahrscheinlichkeit, dass A_1 eintritt.

Man kann die bedingte Wahrscheinlichkeiten $P(A_2|A_1)$ auch wie folgt verstehen: Wir schränken die Ergebnismenge auf A_1 ein und berechnen dann die Wahrscheinlichkeit, dass A_2 eintritt. Die Einschränkung der Ergebnismenge auf A_1 wird oft wie folgt beschrieben. Wir wissen von einem Ergebnis, dass es in A_1 liegt. Wir wissen aber nicht, ob es auch in A_2 liegt. Wie hoch ist dann die Wahrscheinlichkeit, dass es in A_2 liegt.

Beispiel 8.3.17. Wie hoch ist die Wahrscheinlichkeit aus 36 Karten, die durchnummeriert sind, eine Karte mit einer Zahl kleiner-gleich 10 zu ziehen, unter der Bedingung, dass die gezogene Zahl gerade ist (Gleiche Wahrscheinlichkeit der Ergebnisse).

$$A_1 = \{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36\}.$$

$$A_2 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

$$A_1 \cap A_2 = \{2, 4, 6, 8, 10\}$$

$$\text{Es gilt } P(A_2|A_1) = \frac{P(A_2 \cap A_1)}{P(A_1)} = \frac{\frac{5}{36}}{\frac{18}{36}} = 0.27778.$$

◇

Satz 8.3.18. (Satz von der totalen Wahrscheinlichkeit): Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein Wahrscheinlichkeitsraum, $(A_1, \dots, A_n)$ eine Zerlegung von Ω und $B \subset \Omega$, dann gilt:

$$P(B) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i)$$

$$\text{Beweis. } P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B|A_i) \cdot P(A_i)$$

da aus der Definition der bedingten Wahrscheinlichkeit $P(B|A_i) \cdot P(A_i) = P(B \cap A_i)$ folgt.

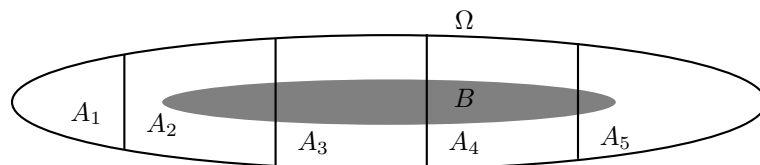


Abbildung 8.3.3: Graphik zum Satz von der totalen Wahrscheinlichkeit

□

Es folgt ein Beispiel für die Anwendung des Satzes:

Beispiel 8.3.19. Tennisspieler a weiss, dass er gegen den Spieler

b eine Gewinnwahrscheinlichkeit von	0.2
c eine Gewinnwahrscheinlichkeit von	0.3
d eine Gewinnwahrscheinlichkeit von	0.7
e eine Gewinnwahrscheinlichkeit von	0.9

Sein Gegenspieler wird durch Los bestimmt. Wie hoch ist die Wahrscheinlichkeit, dass er die nächste Runde des Turniers erreicht?

Mit

$$\begin{aligned} ab &:= a \text{ spielt gegen Spieler } b \text{ und gewinnt} \\ \overline{ab} &:= a \text{ spielt gegen Spieler } b \text{ und verliert} \end{aligned}$$

und analog für das Spiel gegen die anderen Spieler gilt

$$\Omega = \{ab, \overline{ab}, ac, \overline{ac}, ad, \overline{ad}, ae, \overline{ae}\}$$

Das Ereignis, gegen den Spieler $i \in \{b, c, d, e\} := C$ zu spielen ist dann:

$$A_i = \{ai, \overline{ai}\}.$$

Offensichtlich ist $A_i \subset \Omega$ und $\{A_i \mid i \in \{b, c, d, e\}\}$ ist eine Zerlegung von Ω . Sei G das Ereignis zu gewinnen:

$$G = \{ab, ac, ad, ae\} \subset \Omega$$

Die Wahrscheinlichkeit, dass er gegen Spieler i spielt, ist:

$$P(A_i) = \frac{1}{4}$$

(Wahl durch Los, uniforme Verteilung bei vier möglichen Gegenspielern). Die Wahrscheinlichkeit, dass er gegen Spieler i gewinnt, ist die Wahrscheinlichkeit, dass er gewinnt, unter der Bedingung, dass er gegen den entsprechenden Spieler spielt:

$$P(G|A_b) = 0.2$$

$$P(G|A_c) = 0.3$$

$$P(G|A_d) = 0.7$$

$$P(G|A_e) = 0.9$$

Damit ist dann gemäss Satz 8.3.18 die Wahrscheinlichkeit, dass er gewinnt:

$$\begin{aligned} P(G) &= \sum_{i \in C} P(G|A_i) \cdot P(A_i) \\ &= 0.2 \cdot \frac{1}{4} + 0.3 \cdot \frac{1}{4} + 0.7 \cdot \frac{1}{4} + 0.9 \cdot \frac{1}{4} \\ &= 0.525 \end{aligned}$$

denn die A_i stellen eine Zerlegung von Ω dar. ◇

Satz 8.3.20. (Satz von Bayes, Thomas Bayes [Aussprache: bei:z], 1701 - 1761, englischer Mathematiker und presbyterianischer Pfarrer) Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein Wahrscheinlichkeitsraum, $(A_1, \dots, A_n)$ eine Zerlegung von Ω und $B \subset \Omega$, $P(A_j) > 0$ und $P(B|A_i) > 0$ für mindestens ein i , dann gilt:

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)} = \frac{P(B|A_j) \cdot P(A_j)}{P(B)}$$

Beweis. Gemäss Definition gilt

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)}$$

und mit Satz von der totalen Wahrscheinlichkeit

$$P(A_j|B) = \frac{P(A_j \cap B)}{\sum_{i=1}^n P(B|A_i) \cdot P(A_i)}.$$

Aus der Definition der bedingten Wahrscheinlichkeit folgt zudem $P(A_j \cap B) = P(B|A_j) \cdot P(A_j)$, wodurch der Satz unmittelbar folgt. $\square$

Es folgt ein Beispiel für die Anwendung des Satzes von Bayes:

Beispiel 8.3.21. 2% einer Population weisen eine bestimmte Krankheit auf. Es wurde ein Test entwickelt, um die Krankheit nachzuweisen. Dieser zeigt bei 97% der Kranken an, dass sie krank sind und bei 1% der Gesunden, dass sie krank sind. Wie hoch ist die Wahrscheinlichkeit, dass eine Person, bei der der Test „krank“ anzeigt, krank ist?

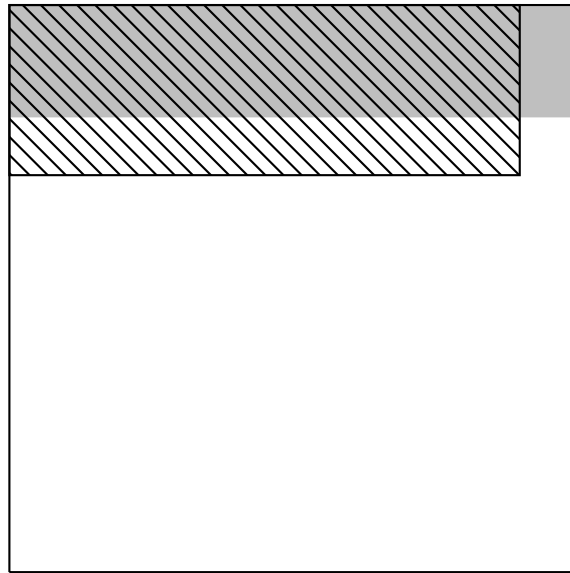


Abbildung 8.3.4: Graue Fläche repräsentiert Anteil der Kranken (0.03), weisse Fläche den Anteil der Gesunden. Der Schnitt der schraffierten Fläche mit den Gesunden repräsentiert den Anteil der Gesunden mit Testresultat Krank (0.02), der Schnitt der schraffierten Fläche mit den Kranken repräsentiert den Anteil der Kranken mit Testresultat Krank (0.97). Der Anteil der Kranken an den Personen mit Testresultat Krank ist nicht so hoch, wie man eventuell erwartet.

$$\Omega = \{kk, gk, kg, gg\}$$

wobei die erste Stelle in kk , gk , kg und gg jeweils den Gesundheitszustand anzeigt und die Zweite Stelle das Testresultat.

Sei K der Umstand, krank zu sein. Damit ist

$$K = \{kk, kg\} \subset \Omega$$

K^c ist dann der Umstand, gesund zu sein, wobei $\{K, K^c\}$ eine Zerlegung von Ω ist. Gemäss Aufgabenstellung gilt

$$P(K) = 0.02,$$

und damit für K^c

$$P(K^c) = 0.98.$$

Sei T_k das Ereignis, dass der Test „krank“ anzeigt. Dann ist

$$T_k = \{kk, gk\} \subset \Omega.$$

Gemäss Aufgabenstellung gilt:

$P(T_k|K) = 0.97$ (Wahrscheinlichkeit von „Krank anzeigen“ unter der Bedingung von „Krank“)
 $P(T_k|K^c) = 0.01$ (Wahrscheinlichkeit von „Krank anzeigen“ unter der Bedingung von „Gesund“).

Gesucht ist $P(K|T_k)$, die Wahrscheinlichkeit krank zu sein, unter der Bedingung, dass der Test krank anzeigt. Gemäss Satz 8.3.20 gilt:

$$\begin{aligned} P(K|T_k) &= \frac{P(T_k|K) \cdot P(K)}{P(T_k|K) \cdot P(K) + P(T_k|K^c) \cdot P(K^c)} \\ &= \frac{0.97 \cdot 0.02}{0.97 \cdot 0.02 + 0.01 \cdot 0.98} \\ &= 0.6648 \end{aligned}$$

Die Wahrscheinlichkeit ist vermutlich nicht so hoch, wie man sie erwartet hat. Dies ist dadurch zu erklären, dass 98% der Personen gesund sind und bei diesen der Test zu 1% auf „krank“ anzeigt. Dies ergibt viele gesunde Personen mit Krank-Testresultat relativ zur Anzahl der Kranken mit Krank-Testresultat (Bemerkung: in der Medizin wird die Wahrscheinlichkeit, dass ein Test bei Kranken das Resultat „krank“ anzeigt als „Sensitivität“ bezeichnet. Die Wahrscheinlichkeit, dass ein Test bei Gesunden nicht „krank“ anzeigt wird als „Spezifizität“ bezeichnet). $\diamond$

Satz 8.3.22. Ist $P(A_2|A_1) = P(A_2)$, dann gilt $P(A_2)P(A_1) = P(A_2 \cap A_1)$.

Beweis. Folgt unmittelbar aus der Definition und der Bedingung. $\square$

Definition 8.3.23. Sei der diskrete Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ gegeben. Für Ereignisse $C, D \subset \Omega$ gilt: Ereignisse C und D sind unabhängig bezüglich der Verteilung P genau dann, wenn

$$P(C) \cdot P(D) = P(C \cap D)$$

Für $P(C) > 0$ ist dies äquivalent mit der Formulierung: Ereignisse C und D sind unabhängig genau dann, wenn

$$P(D) = \frac{P(C \cap D)}{P(C)} = P(D|C).$$

Damit sind für diesen Fall C und D unabhängig genau dann wenn

$$P(D) = P(D|C).$$

Zu beachten ist, dass Unabhängigkeit von Ereignissen nicht dasselbe ist wie das Fehlen gegenseitiger Beeinflussung von Ereignissen. Dies kann man einfach im Falle der Gleichverteilung auf Ω sehen. Unabhängigkeit bedeutet dann, dass

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|} = P(A) = \frac{|A|}{|\Omega|},$$

d.h. die relative Häufigkeit der Ergebnisse von A , die in B sind, ist bezüglich der Menge der Ergebnisse von B identisch mit der relativen Häufigkeit der Ergebnisse von A bezüglich Ω . Die Idee des Fehlens gegenseitiger Beeinflussung fehlt hier. Diese Idee setzt mindestens zwei Ergebnismengen voraus, während die Unabhängigkeit von Ereignissen bezüglich einer einzigen Ergebnismenge

definiert ist. Trotzdem gibt es einen Zusammenhang zwischen der Unabhängigkeit von Ereignissen und der fehlenden gegenseitigen Beeinflussung von Ereignissen: Spezifische Ereignisse in Produkträumen sind genau dann unabhängig, wenn die Ereignisse der Ereignisräume, aus denen die Produkträume gebildet werden, sich gegenseitig nicht beeinflussen. Dies wird im folgenden Abschnitt bewiesen.

Zusammenhang von Produktemass und Unabhängigkeit von Ereignissen (fakultativer Stoff)

Bevor der Zusammenhang im Satz 8.3.27 formell formuliert wird, drei vorbereitende Sätze:

Satz 8.3.24. Für beliebige Mengen A, B, Ω_1, Ω_2 mit $A \subset \Omega_1$ und $B \subset \Omega_2$ gilt :

$$A \times B = A \times \Omega_2 \cap \Omega_1 \times B$$

Beweis. (1) Sei $(\omega_1, \omega_2) \in A \times B$. Damit gilt $(\omega_1, \omega_2) \in A \times \Omega_2$ (denn $B \subset \Omega_2$) und $(\omega_1, \omega_2) \in \Omega_1 \times B$, denn $A \subset \Omega_1$. Damit gilt $(\omega_1, \omega_2) \in A \times \Omega_2 \cap \Omega_1 \times B$.

(2) Sei $(\omega_1, \omega_2) \in A \times \Omega_2 \cap \Omega_1 \times B$. Damit gilt: $\omega_1 \in A, \omega_2 \in \Omega_2, \omega_1 \in \Omega_1, \omega_2 \in B$. Daraus folgt: $(\omega_1, \omega_2) \in A \times B$. $\square$

Daraus folgt unmittelbar für beliebige diskrete Wahrscheinlichkeitsräume $(\Omega_1 \times \Omega_2, \mathcal{P}(\Omega_1 \times \Omega_2), P_{\Omega_1 \times \Omega_2})$:

Satz 8.3.25. Für $A \subset \Omega_1, B \subset \Omega_2$ gilt:

$$P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2 \cap \Omega_1 \times B)$$

Satz 8.3.26. Seien $(\Omega_i, \mathcal{P}(\Omega_i), P_{\Omega_i})$ diskrete Wahrscheinlichkeitsräume und $P_{\Omega_1 \times \Omega_2}$ das Produktemass auf $\Omega_1 \times \Omega_2$ dann gilt

$$\begin{aligned} P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) &= P_{\Omega_1}(A) \\ P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B) &= P_{\Omega_2}(B). \end{aligned}$$

Beweis. $P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) \stackrel{\text{Produktemass}}{=} P_{\Omega_1}(A) \cdot P_{\Omega_1}(\Omega_2) = P_{\Omega_1}(A) \cdot 1 = P_{\Omega_1}(A)$.

$P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B) = P_{\Omega_1}(\Omega_1) \cdot P_{\Omega_2}(B) = 1 \cdot P_{\Omega_2}(B) = P_{\Omega_2}(B)$. $\square$

Satz 8.3.27. Seien die diskreten Wahrscheinlichkeitsräume $(\Omega_i, \mathcal{P}(\Omega_i), P_{\Omega_i})$ für $i \in \{1, 2\}$ gegeben und $A \subset \Omega_1$ sowie $B \subset \Omega_2$, dann gilt:

$A \times \Omega_2$ und $\Omega_1 \times B$ sind unabhängig bezüglich des Produktemasses $P_{\Omega_1 \times \Omega_2}$ genau dann, wenn sich A und B gegenseitig nicht beeinflussen.

Beweis. (1) Seien $A \times \Omega_2$ und $\Omega_1 \times B$ unabhängig bezüglich des Produktemasses $P_{\Omega_1 \times \Omega_2}$ auf $\Omega_1 \times \Omega_2$. Dann gilt gemäss Definition der Unabhängigkeit

$$P_{\Omega_1 \times \Omega_2}(A \times \Omega_2 \cap \Omega_1 \times B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) \cdot P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B). \quad (8.3.1)$$

Da gemäss Satz 8.3.25

$$P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2 \cap \Omega_1 \times B) \quad (8.3.2)$$

und gemäss Satz 8.3.26

$$P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) \cdot P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B) = P_{\Omega_1}(A) \cdot P_{\Omega_2}(B) \quad (8.3.3)$$

folgt aus (8.3.2) und (8.3.3)

$$P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1}(A) \cdot P_{\Omega_2}(B)$$

d.h. A und B beeinflussen sich gegenseitig nicht.

(2) Wir nehmen umgekehrt an, A und B beeinflussen sich gegenseitig nicht. Damit gilt:

$$P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1}(A) \cdot P_{\Omega_2}(B), \quad (8.3.4)$$

wobei $P_{\Omega_1 \times \Omega_2}$ das Produktemass auf $\Omega_1 \times \Omega_2$ ist. Gemäss Satz 8.3.26 gilt

$$P_{\Omega_1}(A) \cdot P_{\Omega_2}(B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) \cdot P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B). \quad (8.3.5)$$

Aus (8.3.4) und (8.3.5) folgt damit

$$P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) \cdot P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B). \quad (8.3.6)$$

Gemäss Satz 8.3.25 gilt:

$$P_{\Omega_1 \times \Omega_2}(A \times B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2 \cap \Omega_1 \times B). \quad (8.3.7)$$

Aus (8.3.6) und (8.3.7) folgt

$$P_{\Omega_1 \times \Omega_2}(A \times \Omega_2) \cdot P_{\Omega_1 \times \Omega_2}(\Omega_1 \times B) = P_{\Omega_1 \times \Omega_2}(A \times \Omega_2 \cap \Omega_1 \times B),$$

d.h. $A \times \Omega_2$ und $\Omega_1 \times B$ sind unabhängig bezüglich des Produktemasses $P_{\Omega_1 \times \Omega_2}$ auf $\Omega_1 \times \Omega_2$. $\square$

8.3.3 Übungen

1. Sei $\Omega = \{1, 2, 3, 4, 5, 8, 9, 10\}$ mit $P(\{1, 2\}) = 0.5$; $P(\{4, 5\}) = 0.05$. Wieviel beträgt die Wahrscheinlichkeit $P(\{1, 2, 4, 5\})$?
2. Sei $\Omega = \{1, 2, 3, 4, 5, 8, 9, 10\}$ mit $P(\{1, 2, 4\}) = 0.5$; $P(\{4, 5\}) = 0.05$. Wieviel beträgt die Wahrscheinlichkeit $P(\{1, 2, 4, 5\})$?
3. Sei $\Omega = \{1, 2, 3, 4, 5, 8, 9, 10\}$ mit $P(\{1, 2, 4\}) = 0.5$; $P(\{4, 5\}) = 0.05$ und $P(\{4\}) = P(\{5\})$. Wieviel beträgt die Wahrscheinlichkeit $P(\{1, 2, 4, 5\})$?
4. Wie gross ist die Wahrscheinlichkeit, mit einem Würfel dreimal nacheinander „1“ oder dreimal „6“ zu werfen?
5. Ein Werkstück besteht aus zwei Hauptteilen und einem Zwischenstück. Diese werden von drei verschiedenen Spezialfirmen bezogen, welche für die Hauptteile einen Ausschuss von 4 %, für das Zwischenstück von 10 % garantieren. Wie gross ist die Wahrscheinlichkeit, dass drei zufällig ausgewählte Teile ein einwandfreies Werkstück ergeben?
6. Zwei Schützen sehen eine Tontauben und schiessen gleichzeitig auf sie. Normalerweise trifft der Schütze A in 30 % der Fälle und der Schütze B in 40 % der Fälle.
Wie gross ist die Wahrscheinlichkeit, dass die Tontauben
(a) getroffen wird?
(b) genau einmal getroffen wird?
7. Ein Maturand billigt sich für die Zulassung zum Studium an der Universität A eine Chance von 65 % zu. Er hat sich auch an der Uni B beworben, die ihn aber mit 45 % Wahrscheinlichkeit zurückweisen wird. Welche Aussicht hat er, bei mindestens einer der Universitäten zugelassen zu werden?

8. Eine Pensionskasse beabsichtigt, von einem Ehepaar ein Grundstück zu erwerben gegen Gewährung einer Leibrente auf 25 Jahre.
Gemäss Sterbetafel beträgt die Wahrscheinlichkeit, noch 25 Jahre zu leben, für den 60jährigen Mann 0.18 und für die 65jährige Frau 0.13 .
Berechnen Sie die Wahrscheinlichkeit, dass nach 25 Jahren
 - (a) nur der Mann noch lebt.
 - (b) nur die Frau noch lebt.
 - (c) genau eine der beiden Personen noch lebt.
 - (d) beide Personen noch leben.
 - (e) keine der beiden Personen mehr lebt
 - (f) mindestens eine der beiden Personen noch lebt.
9. Wie gross ist die Wahrscheinlichkeit, beim Ziehen einer Karte aus einem Jassspiel keinen König zu erhalten?
10. Wie gross ist die Wahrscheinlichkeit, beim zweimaligen Ziehen und jeweiligem Zurücklegen von Kugeln keine weisse zu erhalten (Ziehen aus 20 weissen und 10 schwarzen Kugeln)?
11. Beim Würfeln gewinnt der mit der höheren Augenzahl. Der erste würfelt 3. Wieviel beträgt die Wahrscheinlichkeit, dass der zweite gewinnt? Der erste würfelt 5. Wieviel beträgt die Wahrscheinlichkeit, dass der zweite gewinnt?
12. Beim Überqueren eines Landes muss ein Flugkörper fünf Radarstationen überfliegen. Von jeder Station werde er ohne Beeinflussung durch die anderen mit Wahrscheinlichkeit 0.02 entdeckt.
 - a) Mit welcher Wahrscheinlichkeit wird er während eines Einzelfluges entdeckt?
 - b) Mit welcher Wahrscheinlichkeit wird er bei zehn Flügen nicht entdeckt?
13. Mit welcher Wahrscheinlichkeit kann mit drei Würfeln in einem Wurf eine Summe grösser-gleich 10 geworfen werden.
14. Eine Würfel sei so verfälscht, dass die Augenzahlen 2, 3, 4 und 5 die gleiche Wahrscheinlichkeit besitzen, während die Zahl 6 die doppelte Wahrscheinlichkeit der 2 und die dreifache Wahrscheinlichkeit der 1 besitzt. Berechnen Sie die Wahrscheinlichkeit der Ereignisse G (gerade Augenzahl), U (ungerade Augenzahl) und A (die Augenzahl beträgt mindestens 5 beim einmaligen Würfeln).
15. Mit einem Glücksrad sollen die Zahlen 1, 2, 3, 4, 5, 6, erzeugt werden. Dabei besitzen die zugehörigen Kreissektoren die folgenden Innenwinkel

1	45°
2	60°
3	90°
4	25°
5	40°
6	100°

Tabelle 8.3.1: Winkel des Glücksrades

- a) bestimmen Sie die Wahrscheinlichkeit, dass das Rad bei einer spezifischen Zahl stehen bleibt.
 - b) wie hoch ist die Wahrscheinlichkeit, dass das Rad bei einer geraden Zahl stehen bleibt.
 - c) wie hoch ist die Wahrscheinlichkeit, dass das Rad bei einer Zahl grösser-gleich 5 oder kleiner-gleich 3 stehen bleibt.
16. In einer Urne befinden sich 20 rote, 30 blaue und 50 schwarze Kugeln. Wie gross ist die Wahrscheinlichkeit, in einem Zug

- eine rote
 - eine rote oder eine blaue
 - keine rote zu ziehen?
17. Sei $\Omega = \{1, 2, 3\}$ und $f = \{(1, \frac{1}{3}), (2, \frac{1}{6}), (3, \frac{1}{2})\}$; $A = \{1\}, B = \{1, 2\}, C = \{2, 3\}, D = \{1, 3\}$
Berechnen Sie $P(A|B); P(A|A); P(B|A)$
18. Mit wie vielen Würfeln muss mindestens geworfen werden, damit das Ereignis „es erscheint mindestens eine Sechs“ eine Wahrscheinlichkeit von mehr als 0.99 besitzt?
19. Eine Hausfrau kauft einen Karton mit 10 Eiern, wovon zwei verdorben sind. Zum Backen eines Kuchens wählt sie drei Eier zufällig aus, ohne die Eier zu kontrollieren. Mit welcher Wahrscheinlichkeit gelangt keines der verdorbenen Eier in den Kuchen?
20. Seien die Graphen zweier W-Funktionen auf $\Omega_1 = \{a, b, c\}$ und $\Omega_2 = \{d, e, f\}$ gegeben:
 $f_1 = \{(a, 0.2) (b, 0.1) (c, 0.7)\}$
 $f_2 = \{(d, 0.1) (e, 0.3) (f, 0.6)\}$.
 a) Berechnen Sie die Produkte-W-Funktion auf $\Omega_1 \times \Omega_2$.
 b) Überprüfen Sie, ob der folgende Graph einer W-Funktion auf $\Omega_1 \times \Omega_2$ eine Gemeinsame W-Funktion bezüglich f_1 und f_2 ist. $\{(a, d, 0.01), (a, e, 0.02), (a, f, 0.001), (b, d, 0.15), (b, e, 0.2), (b, f, 0.33), (c, d, 0.15), (c, e, 0.029), (c, f, 0.11)\}$
 c) Überprüfen Sie, ob der folgende Graph einer W-Funktion auf $\Omega_1 \times \Omega_2$ eine Gemeinsame W-Funktion bezüglich f_1 und f_2 ist.
 $\{(a, d, 0.02), (a, e, 0.1), (a, f, 0.08), (b, d, 0.05), (b, e, 0.01), (b, f, 0.04), (c, d, 0.03), (c, e, 0.19), (c, f, 0.48)\}$
21. a) Finden Sie eine allgemeine Formulierung für Satz 8.3.6
 b) Finden Sie eine allgemeine Formulierung der Definition 8.3.8
 c) Finden Sie eine allgemeine Formulierung für den Satz 8.3.10
 d) Suchen Sie eine allgemeine Formulierung der Definition 8.3.11.
 e) Suchen Sie eine allgemeine Formulierung der Definition 8.3.4
22. Vor einer Prüfung erhält ein Kandidat einen Katalog mit 20 verschiedenen Fragen, von denen in der Prüfung drei zu beantworten sind. Der Prüfling wählt zehn Fragen zufällig aus und bereitet sich nur auf diese vor. Mit welcher Wahrscheinlichkeit hat er sich auf k der in der Prüfung gestellten Fragen vorbereitet (für $k = 0, 1, 2, 3$). Wir nehmen an, der Kandidat beantworte vorbereitete Fragen mit dem Maximum der Punkte und bei nicht vorbereiteten Fragen 0 Punkte. Bei 2 richtig beantworteten Fragen hat er bestanden. Wie gross ist die Wahrscheinlichkeit, dass er besteht? Wie gross ist die Wahrscheinlichkeit, dass er besteht, wenn nur eine richtig beantwortete Frage verlangt wird?
23. An einem Spiel nehmen zehn Personen teil. Bei jedem Durchgang wird eine Person ausgelost, die ausscheiden muss. Berechnen Sie die Wahrscheinlichkeiten für folgende Ereignisse:
 a) das Ehepaar Müller bleibt vor der letzten Auslosung im Spiel,
 b) Herr Müller bleibt als letzter übrig.
24. Beim Lotto: 99 Steine von 1 bis 99 durchnummeriert. Wie gross ist die Wahrscheinlichkeit, die Reihe $[5, 11, 28, 55, 82]$ auf Grund der ersten fünf ausgerufenen Zahlen decken zu können.
25. Eine Firma fertigt Elektrogeräte in drei Produktionsstätten und zwar 25% im Werk A, 35% im Werk B und den Rest im Werk C. Von den im Werk A produzierten Geräten sind 2%, von den in B produzierten 2.5% und den in C produzierten 3% fehlerhaft.
 a) Mit welcher Wahrscheinlichkeit ist ein aus der Gesamtproduktion zufällig ausgewähltes Gerät fehlerhaft?
 b) Aus der Gesamtproduktion wird ein fehlerhaftes Gerät ausgewählt. Mit welcher Wahrscheinlichkeit wurde es in A, B bzw. C produziert?

26. Aus Erfahrung sei bekannt, dass 92% der in einem Werk produzierten Maschinen einwandfrei sind. In der Endkontrolle werden 5% der einwandfreien und 98% der fehlerhaften als fehlerhaft aussortiert. Berechnen Sie die Wahrscheinlichkeiten für folgende Ereignisse:
 - a) ein nicht aussortiertes Gerät ist fehlerfrei;
 - b) ein aussortiertes Gerät ist tatsächlich fehlerhaft.
27. Eine Krankheit kommt bei ca. 0.5% der Bevölkerung vor. Ein Test zur Auffindung der Krankheit führt bei 99% der Kranken zu einer Reaktion, aber auch bei 2% der Gesunden. Wie gross ist die Wahrscheinlichkeit, dass eine Person, bei der die Reaktion eintritt, die Krankheit wirklich hat?
28. Drei Manager verlassen vorzeitig die Aktionärsversammlung. Sie greifen nacheinander in der verdunkelten Garderobe auf gut Glück je einmal in den Schirmständer, in welchem sich 30 Schirme befinden, darunter auch ihre eigenen. Wie gross ist die Wahrscheinlichkeit, dass jeder durch Zufall gerade seinen eigenen Schirm erwischt?
29. Aus den Buchstaben STUTTGART werden dreimal hintereinander je ein Buchstabe a) mit und b) ohne Zurücklegen ausgewählt. Berechnen Sie die Wahrscheinlichkeiten für folgende Ereignisse: A) in der Reihenfolge der Ziehung entsteht das Wort RAT. B) aus den gezogenen Buchstaben lässt sich das Wort RAT bilden.
30. Die Geburtstage der Bewohnerinnen und Bewohner der Schweiz seien gleichmässig auf die zwölf Monate verteilt. Zehn Personen werden zufällig ausgewählt. Mit welcher Wahrscheinlichkeit haben
 - a) genau drei im Januar,
 - b) mindestens drei im Januar,
 - c) alle zehn Personen in verschiedenen Monaten Geburtstag.

8.3.4 Lösungen

Bemerkung: Es gibt manchmal verschiedene Möglichkeiten, eine Aufgabe anzugehen. Die gelieferten Lösungen sind als Lösungsvorschläge zu betrachten.

1. Da $\{1, 2\} \cap \{4, 5\} = \emptyset$, gilt $P(\{1, 2\}) + P(\{4, 5\}) = 0.5 + 0.05 = 0.55$.
2. Da $\{1, 2, 4\} \cap \{4, 5\} \neq \emptyset$, können wir ohne zusätzliche Information $P(\{1, 2, 4, 5\})$ nicht berechnen.
3. Da $P(\{4\}) = P(\{5\})$, gilt $P(\{4\}) = \frac{0.05}{2} = 0.025$ (denn $\{4\} \cap \{5\} = \emptyset$, somit gilt $P(\{4\}) + P(\{5\}) = P(\{4, 5\}) = 0.05$). Da $\{1, 2, 4, 5\} = \{1, 2, 4\} \cup \{4, 5\} = \{1, 2, 4\} \cup \{5\}$ und $\{1, 2, 4\} \cap \{5\} = \emptyset$ gilt $P(\{1, 2, 4, 5\}) = P(\{1, 2, 4\}) + P(\{5\}) = 0.5 + 0.025 = 0.525$.
4. Ergebnismenge: $\Omega = \Omega_1 \times \Omega_1 \times \Omega_1$ mit $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$. Das Eintreten der einzelnen Ergebnisse beeinflusst die Wahrscheinlichkeit des Eintretens der anderen Ergebnisse nicht. Wir setzen die gleiche Wahrscheinlichkeit der Ergebnisse von Ω_1 voraus. Damit gilt dann: $A = \{(1, 1, 1), (6, 6, 6)\}$. $P_\Omega(A) = f_\Omega((1, 1, 1)) + f_\Omega((6, 6, 6)) = (f_{\Omega_1}(1))^3 + (f_{\Omega_1}(6))^3 = \left(\frac{1}{6}\right)^3 + \left(\frac{1}{6}\right)^3 = 2 \left(\frac{1}{6}\right)^3 = \frac{1}{108} = 0.0092593$
5. Ergebnismenge: $\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$ Menge der Tripel, die aus einem Hauptteil der ersten Art, einem Hauptteil der zweiten Art und einem Zwischenstück bestehen ($\Omega_1 = \{W_1, A_1\}$; $\Omega_2 = \{W_2, A_2\}$; $\Omega_3 = \{W_3, A_3\}$; W_i für „einwandfrei in Ergebnismenge Ω_i “, „ A_i “ für „Ausschuss in Ergebnismenge Ω_i “). Die Ziehungen erfolgen ohne gegenseitige Beeinflussung. Damit können die Wahrscheinlichkeiten multipliziert werden. Die Wahrscheinlichkeit, dass ein Stück aus Ω_1 einwandfrei ist: $\frac{96}{100}$. Die Wahrscheinlichkeit, dass ein Zwischenstück aus Ω_2 einwandfrei ist: $\frac{90}{100}$. Die Wahrscheinlichkeit, dass ein Hauptstück aus Ω_3 einwandfrei ist: $\frac{96}{100}$. $A = \{(W_1, W_2, W_3)\}$. $P_\Omega(A) = \omega_\Omega((W_1, W_2, W_3)) = f_{\Omega_1}(W_1)f_{\Omega_2}(W_2)f_{\Omega_3}(W_3) = \frac{96}{100} \frac{90}{100} \frac{96}{100} = \frac{2592}{3125} = 0.82944$. Damit müsste die Firma mit einem Ausschuss von ca 17% rechnen.

6. Ergebnismenge: $\Omega = \Omega_1 \times \Omega_2 = \{(T, T), (NT, T), (T, NT), (NT, NT)\}$. ($\Omega_1 = \{T, NT\}$; $\Omega_2 = \{T, NT\}$ mit T für „Treffer“, NT für „Nichttreffer“). Das Treffen erfolge ohne gegenseitige Beeinflussung (ist hier nicht unbedingt selbstverständlich. Ein absolut gleichzeitiges Schiessen ist eher unwahrscheinlich. Der Knall des ersten Schützen könnte eine Auswirkung auf den zweiten haben. Andererseits kann man das Schiessen so organisieren, dass sich die Schützen nicht hören). Somit können wir die Wahrscheinlichkeiten multiplizieren.

a) Für beide treffen: $A = \{(T, T)\}$; $P_\Omega(A) = f_\Omega((T, T)) = f_{\Omega_1}(T)f_{\Omega_2}(T) = \frac{3}{10} \cdot \frac{4}{10} = \frac{3}{25}$

Für der erste trifft, der zweite nicht:

$$A = \{(T, NT)\}; P_\Omega(A) = f_\Omega((T, NT)) = f_{\Omega_1}(T)f_{\Omega_2}(NT) = \frac{3}{10} \cdot \frac{6}{10} = \frac{9}{50}$$

Für der erste trifft nicht, der zweite trifft:

$$A = \{(NT, T)\}; P_\Omega(A) = f_\Omega((NT, T)) = f_{\Omega_1}(NT)f_{\Omega_2}(T) = \frac{7}{10} \cdot \frac{4}{10} = \frac{7}{25}$$

Für mindestens einer trifft: $A = \{(T, NT), (NT, T), (T, T)\}$;

$$P_\Omega(A) = f_\Omega((T, NT)) + f_\Omega((NT, T)) + f_\Omega((T, T)) =$$

$$f_{\Omega_1}(T)f_{\Omega_2}(NT) + f_{\Omega_1}(NT)f_{\Omega_2}(T) + f_{\Omega_1}(T)f_{\Omega_2}(T) = \frac{3}{10} \cdot \frac{6}{10} + \frac{7}{10} \cdot \frac{4}{10} + \frac{3}{10} \cdot \frac{4}{10} = \frac{29}{50} = 0.58$$

Man könnte effizienter mit der Gegenwahrscheinlichkeit von A rechnen.

b) Für genau einmal: $A = \{(T, NT), (NT, T)\} = f_\Omega((T, NT)) + f_\Omega((NT, T)) =$

$$f_{\Omega_1}(T)f_{\Omega_2}(NT) + f_{\Omega_1}(NT)f_{\Omega_2}(T) = \frac{3}{10} \cdot \frac{6}{10} + \frac{7}{10} \cdot \frac{4}{10} = \frac{23}{50} = 0.46$$

7. Ergebnismenge:

$$\Omega = \Omega_1 \times \Omega_2 = \{Z, NZ\} \times \{Z, NZ\} = \{(Z, Z), (Z, NZ), (NZ, Z), (NZ, NZ)\}.$$

($\Omega_1 = \{Z, NZ\}$; $\Omega_2 = \{Z, NZ\}$); Z für „Zulassung“, NZ für „Nichtzulassung“. Die Zulassungschancen beeinflussen sich gegenseitig nicht. Wir können somit die Wahrscheinlichkeiten der Ergebnisse aus Ω_1 und Ω_2 multiplizieren. Die Wahrscheinlichkeit, in keiner Uni zugelassen zu werden: $A = \{(NZ, NZ)\}$; $P_\Omega(A) = f_\Omega((NZ, NZ)) = f_{\Omega_1}(NZ)f_{\Omega_2}(NZ) = \frac{35}{100} \cdot \frac{45}{100} = \frac{63}{400}$

Die Wahrscheinlichkeit, mindestens an einer Uni zugelassen zu werden ist damit $1 - \frac{63}{400} = \frac{337}{400} = 0.8425$ (Gegenwahrscheinlichkeit).

8. Ergebnismenge

$$\Omega = \Omega_1 \times \Omega_2 = \{L, LN\} \times \{L, LN\} = \{(L, L), (L, LN), (LN, L), (LN, LN)\}$$

(Ω_1 Ergebnismenge für Mann, Ω_2 für Frau, L für „lebt nach 25 Jahren“, LN für „lebt nicht mehr nach 25 Jahren“). Wir postulieren die fehlende gegenseitige Beeinflussung der Ereignisse (was hier keineswegs selbstverständlich ist. Der Tod eines Ehepartners kann sehr wohl Auswirkungen auf die Wahrscheinlichkeit der Lebensdauer des anderen haben).

(a) $A = \{(L, LN)\}$; $P_\Omega(A) = f_\Omega((L, LN)) = f_{\Omega_1}(L)f_{\Omega_2}(LN) = \frac{18}{100} \cdot \frac{87}{100} = \frac{783}{5000} = 0.1566$

(b) $A = \{(LN, L)\}$; $P_\Omega(A) = f_\Omega((LN, L)) = f_{\Omega_1}(LN)f_{\Omega_2}(L) = \frac{82}{100} \cdot \frac{13}{100} = \frac{533}{5000} = 0.1066$

(c) $A = \{(L, LN), (LN, L)\}$; $P_\Omega(A) = f_\Omega((L, LN)) + f_\Omega((LN, L)) = f_{\Omega_1}(L)f_{\Omega_2}(LN) + f_{\Omega_1}(LN)f_{\Omega_2}(L) = \frac{18}{100} \cdot \frac{87}{100} + \frac{82}{100} \cdot \frac{13}{100} = 0.2632$

(d) $A = \{(L, L)\}$; $P_\Omega(A) = f_\Omega((L, L)) = f_{\Omega_1}(L)f_{\Omega_2}(L) = \frac{18}{100} \cdot \frac{13}{100} = \frac{117}{5000} = 0.0234$

(e) $A = \{(LN, LN)\}$; $P_\Omega(A) = f_\Omega((LN, LN)) = f_{\Omega_1}(LN)f_{\Omega_2}(LN) = \frac{82}{100} \cdot \frac{87}{100} = \frac{3567}{5000} = 0.7134$

(f) $A = \{(L, LN), (LN, L), (L, L)\}$; $A^c = \{(LN, LN)\}$; $P_\Omega(A) = 1 - A^c = 1 - \frac{3567}{5000} = 0.2866$

9. Ω besteht aus den 36 Kartenzügen, die möglich sind. Gleiche Wahrscheinlichkeit der Ergebnisse. Die Wahrscheinlichkeit, einen König zu ziehen, beträgt $\frac{4}{36}$, da es vier Könige im Spiel hat. Die Wahrscheinlichkeit, keinen König zu ziehen, ist dann das Gegenereignis: $1 - \frac{4}{36} = \frac{32}{36} = \frac{8}{9}$

10. $\Omega = \Omega_1 \times \Omega_1$, wobei Ω_1 die Menge der Ergebnisse des einmaligen Ziehens aus den 30 Kugeln ist (mit 10 schwarzen und 20 weißen). Wir setzen gleiche Wahrscheinlichkeit der Ergebnisse in Ω_1 voraus. Die Ergebnisse beeinflussen sich gegenseitig nicht. Damit können wir die Wahrscheinlichkeiten multiplizieren. Wir erhalten keine weiße Kugel, wenn wir beide Male eine schwarze Kugel erhalten. $A = \{(\omega, \omega)\}$; $P_\Omega(A) = f_\Omega((\omega, \omega)) = f_{\Omega_1}(\omega)f_{\Omega_1}(\omega) = \frac{10}{30} \cdot \frac{10}{30} = 0.11111$.

11. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Gleiche Wahrscheinlichkeit der Ergebnisse.
 Höher als 3 sind 4, 5, 6. Somit beträgt die Wahrscheinlichkeit: $\frac{3}{6}$
 Höher als 5 sind: 6. Somit beträgt die Wahrscheinlichkeit: $\frac{1}{6}$.
12. $\Omega = \Omega_1 \times \Omega_1 \times \Omega_1 \times \Omega_1 \times \Omega_1$ mit $\Omega_1 = \{A, NA\}$ mit A für „entdeckt werden“ und NA für „nicht entdeckt werden“. Wir nehmen an, die Ereignisse der verschiedenen Ergebnismengen seien ohne gegenseitige Beeinflussung. Wir können damit die Wahrscheinlichkeiten multiplizieren.
 a) $A = \{(NA, NA, NA, NA, NA)\}; P_\Omega(A) = f_\Omega((NA, NA, NA, NA, NA)) = (f_{\Omega_1}(NA))^5 = 0.98^5 = 0.903\,92 \Rightarrow 1 - 0.903\,92 = 0.096\,08$.
 b) $\Omega' = \Omega^{10}; A' = \{NA\}^{10} = (NA, NA, NA, NA, NA, NA, NA, NA, NA, NA); P_{\Omega'}(A') = 0.903\,92^{10} = 0.364\,17$
13. Ergebnismenge: $\Omega = \Omega_1 \times \Omega_1 \times \Omega_1$ mit $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$. Gleichwahrscheinliche Ergebnisse von Ω_1 . Keine Beeinflussung der Ereignisse zwischen den Ergebnismengen.
 $A = \{(\omega, y, z) : \omega + y + z \geq 10\}$
 Wir rechnen mit der Gegenwahrscheinlichkeit:
 Kleinere Summen als 10 ergeben alle Kombinationen von 1, 2 und 3, da $3 + 3 + 3 = 9 < 10$. Dies ergibt $3^3 = 27$ Fälle.
 Zudem ergeben die Kombinationen folgender Fälle Summen < 10 :
 $(4, 4, 1), (4, 3, 2), (4, 3, 1), (4, 2, 2), (4, 2, 1), (4, 1, 1) :$
 $\binom{3}{1} + 3! + 3! + 3! + \binom{3}{1} + \binom{3}{1} = 27$ Fälle.
 $(5, 3, 1), (5, 2, 2), (5, 2, 1), (5, 1, 1) : 3! + \binom{3}{1} + 3! + \binom{3}{1} = 18$ Fälle.
 $(6, 2, 1), (6, 1, 1) : 3! + \binom{3}{1} = 9$ Fälle.
 Dies ergibt: $27 + 27 + 18 + 9 = 81$
 Somit beträgt die Wahrscheinlichkeit, eine Summe kleiner als 10 zu erhalten: $P(A^c) = \frac{81}{6^3} = \frac{3}{8} = 0.375$
 $P(A) = (1 - \frac{3}{8}) = \frac{5}{8} = 0.625$
14. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Von der Wahrscheinlichkeitsverteilung wissen wir, dass gilt:
 $P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\})$
 $P(\{6\}) = 2P(\{2\})$
 $P(\{6\}) = 3P(\{1\}) \Rightarrow P(\{1\}) = \frac{P(\{6\})}{3} = \frac{2P(\{2\})}{3}$
 $P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 1$
 Wir einsetzen:
 $\frac{2P(\{2\})}{3} + P(\{2\}) + P(\{2\}) + P(\{2\}) + P(\{2\}) + 2P(\{2\}) = \frac{20}{3}P(\{2\})$
 Somit beträgt $P(\{2\}) = \frac{1}{\frac{20}{3}} = \frac{3}{20} = 0.15 = P(\{3\}) = P(\{4\}) = P(\{5\});$
 $P(\{6\}) = 0.3; \quad P(\{1\}) = 0.1$.
 Die Wahrscheinlichkeit „gerade Augenzahl“:
 $P(\{2\}) + P(\{4\}) + P(\{6\}) = 0.15 + 0.15 + 0.3 = 0.6$
 Die Wahrscheinlichkeit „ungerade Augenzahl“:
 $P(\{1\}) + P(\{3\}) + P(\{5\}) = 1 - 0.6 = 0.4$
 Die Wahrscheinlichkeit „mindestens 5“:
 $P(\{5\}) + P(\{6\}) = 0.15 + 0.3 = 0.45$
15. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Von der Wahrscheinlichkeitsfunktion wissen wir:
 a) $f(1) = \frac{45}{360}; f(2) = \frac{60}{360}; f(3) = \frac{90}{360}; f(4) = \frac{25}{360}; f(5) = \frac{40}{360}; f(6) = \frac{100}{360}$
 Wobei gilt: $P(\{\omega\}) = f(\omega)$
 b) $A = \{2, 4, 6\}; P(A) = f(2) + f(4) + f(6) = \frac{60}{360} + \frac{25}{360} + \frac{100}{360} = \frac{37}{72} = 0.513\,89$
 c) $1 - \frac{25}{360} = \frac{67}{72} = 0.930\,56$ (Gegenereignis von $A = \{4\}$).
16. Ω ist die Menge der Kugeln (oder der Ereignisse, eine Kugel zu ziehen). R für das Ereignis, eine rote Kugel zu ziehen, B für das Ereignis eine blaue Kugel zu ziehen, NR für das Ereignis, nicht eine rote Kugel zu ziehen,
 $P(R) = \frac{20}{100}$

$$P(R \cup B) = \frac{20}{100} + \frac{30}{100} = \frac{1}{2} \quad (\text{da } R \cap B = \emptyset)$$

$$P(NR) = P(R^c) = 1 - \frac{20}{100} = \frac{80}{100} = \frac{4}{5}$$

$$17. \text{ a) } P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{1\})}{P(\{1,2\})} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6}} = \frac{2}{3}$$

$$P(A|A) = \frac{P(A \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(\{1\})}{P(\{1\})} = 1$$

18. Die Wahrscheinlichkeit, dass keine 6 auftritt, muss 0.01 betragen (Gegenereignis). Die Wahrscheinlichkeit, dass in n Würfeln keine 6 auftaucht, ist $(\frac{5}{6})^n$. Wir müssen somit $(\frac{5}{6})^n = 0.01$ berechnen. Wir logarithmieren und lösen nach n auf: $n = \frac{\log 0.01}{\log \frac{5}{6}} = 25.259$.

Somit müssen wir 26 mal würfeln, damit wir mit 0.99 Wahrscheinlichkeit mindestens eine 6 erhalten.

(wir haben hier keine Ergebnismenge angegeben. In der Tat geht es darum zu wissen, wie viele Stichprobenräume n mit $\Omega = \Omega_1^n$ wir haben müssen, damit die entsprechende Wahrscheinlichkeit für das vorgegebene Ereignis gilt - wobei sich die Ereignisse aus den Ω_1 gegenseitig nicht beeinflussen).

19. Da wir die Eier nicht zurücklegen, liegt hier nicht eine Serie von unabhängigen Zufallsexperimenten vor. Je nach dem Ergebnis der i -ten Ziehung ergeben sich bei den folgenden Ziehungen andere Möglichkeiten und andere Wahrscheinlichkeiten.

Wir definieren deshalb direkt die Ergebnismenge $\Omega = \{\{e_i, e_j, e_k\} : i \neq j; i \neq k; k \neq j \text{ und } e_i, e_j, e_k \text{ sind Eier}\}$. Dieser repräsentiert die möglichen Ergebnisse des Ziehens von drei Eiern - ohne Reihenfolge und ohne Zurücklegen. Es gilt: $|\Omega| = \binom{10}{3}$ (Ziehen ohne Zurücklegen ohne Reihenfolge). Jede dieser Möglichkeiten hat dieselbe Wahrscheinlichkeit.

Es gibt $\binom{8}{3}$ Möglichkeiten, aus den 8 guten Eiern drei auszuwählen (= günstige Ergebnisse).

Dies ergibt: $\frac{\binom{8}{3}}{\binom{10}{3}} = \frac{7}{15} = 0.46667$.

20. a) Als Graph: $\{(a, d, 0.02), (a, e, 0.06), (a, f, 0.12), (b, d, 0.01), (b, e, 0.03), (b, f, 0.06), (c, d, 0.07), (c, e, 0.21), (c, f, 0.42)\}$

In Tabellenform:

	d	e	f	Summen
a	0.02	0.06	0.12	0.2
b	0.01	0.03	0.06	0.1
c	0.07	0.21	0.42	0.7
Summen	0.1	0.3	0.6	1

Tabelle 8.3.2:

- b) nein. Die Rand-W-Funktionen stimmen nicht mit f_1 und f_2 überein:

	d	e	f	Summen
a	0.01	0.02	0.001	0.031
b	0.15	0.2	0.33	0.68
c	0.15	0.029	0.11	0.289
Summen	0.31	0.249	0.441	1

Tabelle 8.3.3:

- c) ja, da die Rand-W-Funktionen mit f_1 und f_2 übereinstimmen

	d	e	f	Summen
a	0.02	0.1	0.08	0.2
b	0.05	0.01	0.04	0.1
c	0.03	0.19	0.48	0.7
Summen	0.1	0.3	0.6	1

Tabelle 8.3.4:

21. a) Allgemeine Formulierung von Satz 8.3.6: Durch $f_{\times_{i=1}^n \Omega_i} : \times_{i=1}^n \Omega_i \rightarrow [0, 1]$ mit

$$f_{\times_{i=1}^n \Omega_i}((\omega_j^{(1)}, \dots, \omega_k^{(n)})) = \prod_{i=1}^n f_i(\omega_j^{(i)})$$

mit $\omega_j^{(i)} \in \Omega_i$ ist eine Wahrscheinlichkeitsfunktion auf $\times_{i=1}^n \Omega_i$ definiert.

- b) Die allgemeine Formulierung der Definition 8.3.8 ist: seien n diskrete Stichprobenräume Ω_i gegeben mit n Wahrscheinlichkeitsfunktion f_i , so nennen wir das durch die Produktwahrscheinlichkeitsfunktion $f_{\times_{i=1}^n \Omega_i}$ definiert Wahrscheinlichkeitsmass

$$P_{\times_{i=1}^n \Omega_i} : \mathcal{P}(\times_{i=1}^n \Omega_i) \rightarrow [0, 1]$$

$$P_{\times_{i=1}^n \Omega_i}(A) = \sum_{(\omega_j^{(1)}, \dots, \omega_k^{(n)}) \in A} f_{\times_{i=1}^n \Omega_i}((\omega_j^{(1)}, \dots, \omega_k^{(n)}))$$

mit $A \subset \times_{i=1}^n \Omega_i$

Produktmass auf $\times_{i=1}^n \Omega_i$.

- c) eine allgemeine Formulierung von Satz ist 8.3.10:

Sei das Produktmass $P_{\times_{i=1}^n \Omega_i}$ auf die diskrete Ergebnismenge $\times_{i=1}^n \Omega_i$ gegeben mit $A_i \subset \Omega_i$,

dann gilt: $P_{\times_{i=1}^n \Omega_i}(A) = \prod_{i=1}^n P_i(A_i)$ mit $A = \times_{i=1}^n A_i \subset \times_{i=1}^n \Omega_i$.

- d) Die allgemeine Formulierung der Definition 8.3.11 lautet: Ereignisse A_i beeinflussen sich gegenseitig nicht genau dann, wenn $P_{\times_{i=1}^n \Omega_i}(A) = \prod_{i=1}^n P_i(A_i)$ (für $A = \times_{i=1}^n A_i \subset \times_{i=1}^n \Omega_i$).

- e) Seien n diskrete Ergebnismengen Ω_i gegeben. Dann nennen wir eine Wahrscheinlichkeitsfunktion $f_{\times_{i=1}^n \Omega_i} : \times_{i=1}^n \Omega_i \rightarrow [0, 1]$ „gemeinsame Wahrscheinlichkeitsfunktion von den f_{Ω_i} auf $\times_{i=1}^n \Omega_i$ “ genau dann, wenn

$$f_{\Omega_i}(\omega_i) = \sum_{\omega_1 \in \Omega_1} \dots \sum_{\omega_{i-1} \in \Omega_{i-1}} \sum_{\omega_{i+1} \in \Omega_{i+1}} \dots \sum_{\omega_n \in \Omega_n} f_{\times_{i=1}^n \Omega_i}(\omega_1, \dots, \omega_i, \dots, \omega_n) \quad (\text{für alle } \omega_i \in \Omega_i)$$

und für alle f_{Ω_i} mit $1 \leq i \leq n$.

Wir nennen in diesem Zusammenhang f_{Ω_i} die Rand-W-Funktionen von $f_{\times_{i=1}^n \Omega_i}$. Diese bestimmen sogenannte Randverteilungen, nämlich die Wahrscheinlichkeitsmasse P_{Ω_i} , die durch f_{Ω_i} bestimmt sind. Die durch $f_{\times_{i=1}^n \Omega_i}$ definierte Wahrscheinlichkeitsverteilung $\mathcal{P}(\times_{i=1}^n \Omega_i) \rightarrow [0, 1]$ nennen wir „gemeinsame Wahrscheinlichkeitsverteilung“ oder „gemeinsames Wahrscheinlichkeitsmass von den P_{Ω_i} auf $\times_{i=1}^n \Omega_i$ “.

22. Sie H die Menge der Fragen. Wir numerieren diese von 1 bis 20 durch. Es liegt beim Ziehen von Fragen nicht die Situation einer Serie von unabhängigen Zufallsexperimenten vor

(eine Frage kann nicht mehr als einmal gezogen werden und je nach Zug ergeben sich für die weiteren Züge andere Möglichkeiten und Wahrscheinlichkeiten). Wir definieren entsprechend unmittelbar die folgende Ergebnismenge: $\Omega = \{\{F_i, F_j, F_k\} : F_i, F_j, F_k \in H \text{ und } i \neq j \neq k \neq i \text{ und } i, j, k \in \mathbb{N}_{20}^*\}$. Es gilt: $|\Omega| = \binom{20}{3}$ (Ziehen ohne Zurücklegen und ohne Reihenfolge). Wir setzen die gleiche Wahrscheinlichkeit der Ergebnisse voraus.

k = 3: Wir nehmen an, die drei Fragen befinden sich unter den 10. Es ist möglich, auf $\binom{10}{3}$ Arten drei Fragen aus 10 zu ziehen (= „günstige“ Ergebnisse). Somit ist die Wahr-

scheinlichkeit, dass die drei aus den 10 gezogen werden: $\frac{\binom{10}{3}}{\binom{20}{3}} = \frac{120}{1140} = \frac{2}{19} = 0.10526$

k = 2: Es gibt $\binom{10}{2}$ Möglichkeiten, aus den 10 Fragen zwei Fragen auszuwählen. Für jede gibt es 10 Möglichkeiten, eine dritte Frage aus den nicht vorbereiteten Fragen auszuwählen. Dies ergibt:

$$10 \cdot \frac{\binom{10}{2}}{\binom{20}{3}} = \frac{10 \cdot 45}{1140} = \frac{15}{38} = 0.39474$$

k = 1: Es gibt 10 Möglichkeiten, aus den 10 Fragen eine Frage auszuwählen. Für jede gibt es $\binom{10}{2}$ Möglichkeiten, zwei aus den nicht vorbereiteten zu ziehen. Dies ergibt wiederum

$$10 \cdot \frac{\binom{10}{2}}{\binom{20}{3}} = \frac{10 \cdot 45}{1140} = \frac{15}{38} = 0.39474$$

k = 0: entweder werden 3 vorbereitete Fragen, 2 Fragen, eine oder keine gezogen (= sicheres Ereignis. Somit ergibt die Summe dieser Wahrscheinlichkeiten 1). Die paarweisen Schnittmengen dieser Ereignisse ist zudem leer. Entsprechend gilt für $k = 0$: $1 - (\frac{15}{38} + \frac{15}{38} + \frac{2}{19}) = \frac{2}{19} = 0.10526$

Die Wahrscheinlichkeit zu bestehen (2 Fragen sind verlangt) ist: $\frac{2}{19} + \frac{15}{38} = \frac{1}{2}$

Die Wahrscheinlichkeit zu bestehen (1 Fragen ist verlangt) ist: $\frac{2}{19} + \frac{15}{38} + \frac{15}{38} = \frac{17}{19} = 0.89474$

23. Sei H die Menge der 10 Personen. Auch hier beeinflussen vorangehende Aufstellungen die Möglichkeiten und Wahrscheinlichkeiten weiterer Aufstellungen, etc. Die Problemstellung ist Ziehen ohne Zurücklegen in Reihenfolge. Wir wählen deshalb die Ergebnismenge $\Omega = \{(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}) : \omega_i \neq \omega_j \text{ für } i \neq j \text{ und } i, j \in \mathbb{N}_{10}^*\}$. Es gilt $|\Omega| = 10!$. Wir können die gleiche Wahrscheinlichkeit der Ergebnisse von Ω voraussetzen. Eine günstige Reihenfolge ist eine mit Herr Müller und Frau Müller am Schluss oder Frau Müller und Herr Müller am Schluss. Damit gibt es zwei Möglichkeiten, Herr und Frau Müller in diesem Fall anzuordnen. Für jede dieser zwei Möglichkeiten gibt es $8!$ Möglichkeiten, die übrigen Spielteilnehmer anzuordnen. Somit ergeben sich $2 \cdot 8!$ „günstige Ergebnisse“. Die Wahrscheinlichkeit, Herr oder Frau Müller am Schluss zu haben ist damit: $\frac{2 \cdot 8!}{10!} = 0.022222$
- b) Ω wie oben. Es gibt $9!$ Möglichkeiten, die übrigen 9 Personen anzuordnen. Damit gilt für die Wahrscheinlichkeit, Herrn Müller am Schluss zu haben: $\frac{9!}{10!} = \frac{1}{10}$
24. Zahlen können nicht mehrmals gezogen werden. Entsprechend beeinflusst das Resultat vorangehender Ziehung die Möglichkeiten bei den folgenden Ziehungen. Wir definieren deshalb $\Omega = \{\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\} : \omega_i \neq \omega_j \text{ für } i \neq j \text{ und } \omega_i, \omega_j \in \mathbb{N}_{99}^*\}$. $|\Omega| = \binom{99}{5} = 71\,523\,144$ (Ziehen ohne Zurücklegen, ohne Reihenfolge). Alle Ergebnisse von Ω sind gleich wahrscheinlich. Damit gilt für die Wahrscheinlichkeit, diese Zahlen zu erhalten: $\frac{1}{\binom{99}{5}} = \frac{1}{71\,523\,144} = 1$.

$$3981 \times 10^{-8}$$

25. Ω Menge der produzierten Objekte. A = Menge der Objekte der Produktion A, B = Menge der Objekte der Produktion B, C = Menge der Objekte der Produktion C. F = Menge der fehlerhafte Objekte. $P(A) = 0.25$; $P(B) = 0.35$; $P(C) = 0.4$.

a) 2% der 25% sind fehlerhaft, 2.5% der 35% und 3% der 40%. Dies ergibt: $P(F|A) = 0.02$; $P(F|B) = 0.025$; $P(F|C) = 0.03$. $\implies$

$$(i) 0.02 = \frac{P(A \cap F)}{P(A)} \implies P(A \cap F) = P(A)0.02 = 0.25 \cdot 0.02 = 0.005$$

$$(ii) 0.025 = \frac{P(B \cap F)}{P(B)} \implies P(B \cap F) = P(B)0.025 = 0.35 \cdot 0.025 = 0.00875$$

$$(iii) 0.03 = \frac{P(C \cap F)}{P(C)} \implies P(C \cap F) = P(C)0.03 = 0.4 \cdot 0.03 = 0.012$$

Da die paarweisen Schnittmengen von A , C und B leer sind, gilt: Die paarweisen Schnitte von $A \cap F$, $B \cap F$ und $C \cap F$ sind leer. Da zudem $A \cup C \cup B = \Omega$, gilt: $(A \cap F) \cup (B \cap F) \cup (C \cap F) = F$, da $F \subset \Omega$ (Satz 8.1.13). Damit gilt dann $P(F) = P(A \cap F) + P(B \cap F) + P(C \cap F) = 0.005 + 0.00875 + 0.012 = 0.02575$

$$b) P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{0.005}{0.02575} = 0.19417$$

$$P(B|F) = \frac{P(B \cap F)}{P(F)} = \frac{0.00875}{0.02575} = 0.33981$$

$$P(C|F) = \frac{P(C \cap F)}{P(F)} = \frac{0.012}{0.02575} = 0.46602$$

$$\text{Kontrolle: } 0.19417 + 0.33981 + 0.46602 = 1$$

26. (Ω Menge der produzierten Objekte. F für die Menge der fehlerhaften, A für die Menge der aussortierten Gegenstände)

Angaben: $P(A|F^c) = 0.05$; $P(A|F) = 0.98$; $P(F^c) = 0.92$; $P(F) = 0.08$

$$(i) \text{ Aus } P(A|F^c) = \frac{P(A \cap F^c)}{P(F^c)} = 0.05 \text{ folgt damit:}$$

$$P(A \cap F^c) = P(F^c) \cdot 0.05 = 0.92 \cdot 0.05 = 0.046$$

$$(ii) \text{ Aus } P(A|F) = \frac{P(A \cap F)}{P(F)} = 0.98 \text{ und } P(F) = 0.08 \text{ folgt:}$$

$$P(A \cap F) = 0.98 \cdot 0.08 = 0.0784$$

$$(iii) P(A) = P(A \cap F^c) + P(A \cap F) = 0.046 + 0.0784 = 0.1244$$

$$\text{und } P(A^c) = 1 - 0.1244 = 0.8756.$$

$$(iv) \text{ Gemäss 8.1.14 gilt: } F^c \cap A^c = (F \cup A)^c.$$

Wir erhalten also: $P(F^c \cap A^c) = P((F \cup A)^c) = 1 - P(F \cup A)$. Da $P(F \cup A) = P(F) + P(A) - P(F \cap A)$, erhalten wir: $P(F^c \cap A^c) = 1 - (0.08 + 0.1244 - 0.0784) = 0.874$

Damit gilt dann:

$$a) P(F^c|A^c) = \frac{P(F^c \cap A^c)}{P(A^c)} = \frac{0.874}{0.8756} = 0.99817$$

$$b) P(F|A) = \frac{P(F \cap A)}{P(A)} = \frac{0.08 \cdot 0.98}{0.1244} = 0.63023$$

Aufgaben dieser Art werden auch mit der Bayesschen Formel berechnet: $P(F|A) = \frac{P(F \cap A)}{P(A)} = \frac{P(A|F) \cdot P(F)}{P(A)}$

$$\text{In unserem Beispiel: } P(F|A) = \frac{P(F \cap A)}{P(A)} = \frac{P(A|F) \cdot P(F)}{P(A)} = \frac{0.98 \cdot 0.08}{0.1244} = 0.63023.$$

27. a) *Variante I:* Sei R das Ereignis, zu reagieren, K das Ereignis, krank zu sein, G das Ereignis, Gesund zu sein. Aus der Aufgabenstellung erhalten wir:
 $P(K) = 0.005$. (Wahrscheinlichkeit, krank zu sein) $\implies P(G) = 0.995$ (Wahrscheinlichkeit, gesund zu sein = Gegenwahrscheinlichkeit, krank zu sein)
 $P(R|K) = 0.99$ (Wahrscheinlichkeit zu reagieren unter der Bedingung, dass man krank ist)
 $P(R|G) = 0.02$. (Wahrscheinlichkeit zu reagieren unter der Bedingung, dass man gesund ist)
 Gesucht ist $P(K|R) = \frac{P(R \cap K)}{P(R)}$ (Wahrscheinlichkeit krank zu sein unter der Bedingung, dass man reagiert ist).

Mit Hilfe der obigen Definition erhalten wir:

$$P(R|K) = \frac{P(R \cap K)}{P(K)} = 0.99 \implies P(R \cap K) = P(K) \cdot 0.99 = 0.005 \cdot 0.99 = 0.00495$$

$$P(R|G) = \frac{P(R \cap G)}{P(G)} = 0.02 \implies P(R \cap G) = P(G) \cdot 0.02 = 0.995 \cdot 0.02 = 0.0199$$

Da K und G eine Zerlegung der Grundgesamtheit darstellen, gilt mit den Sätzen 8.1.12 und 8.1.13: $(R \cap K) \cup (R \cap G) = R$ und $(R \cap K) \cap (R \cap G) = \emptyset$. Damit gilt dann: $P(R) = P((R \cap K) \cup (R \cap G)) = P(R \cap K) + P(R \cap G) = 0.00495 + 0.0199 = 0.02485$.

$$\text{Wir erhalten damit } P(K|R) = \frac{P(R \cap K)}{P(R)} = \frac{0.00495}{0.02485} = 0.19920$$

Variante II (intuitive Überlegungen): 2% der 99.5% zeigen eine Reaktion, obwohl sie gesund sind. Es sind dies

$$\frac{99.5 \cdot 2}{100} = 1.99\% \text{ der Gesamtbevölkerung.}$$

99% der 0.5% zeigen eine Reaktion, wenn sie krank sind. Dies sind:

$$\frac{0.5 \cdot 99}{100} = 0.495\% \text{ der Gesamtbevölkerung.}$$

Die Wahrscheinlichkeit, krank zu sein, wenn man reagiert beträgt (der Anteil kranken Reagierenden an der Gesamtheit der Reagierenden): $\frac{0.495}{0.495 + 1.99} = 0.1992$

28. $\Omega = \{(\omega_i, \omega_j, \omega_k) : \omega_i, \omega_j, \omega_k \in \text{Menge der Regenschirme und } \omega_i \neq \omega_j \neq \omega_k \neq \omega_i\}$ (Ziehen ohne Zurücklegen, in Reihenfolge). $|\Omega| = 30 \cdot 29 \cdot 28$. Alle Ergebnisse von Ω sind gleichwahrscheinlich. Somit ergibt sich eine Wahrscheinlichkeit von $\frac{1}{30 \cdot 29 \cdot 28} = 4.1051 \times 10^{-5}$

29. Wir können das Experiment wie folgt auffassen. Aus drei identischen Ergebnismengen Ω_i wird je ein Buchstabe gezogen ($\Omega = \Omega_1 \times \Omega_2 \times \Omega_3$): Die Wahrscheinlichkeit im ersten Zug aus Ω_1 ein R zu ziehen beträgt $\frac{1}{9}$.

Die Wahrscheinlichkeit im zweiten Zug aus Ω_2 ein A zu ziehen beträgt $\frac{1}{9}$.

Die Wahrscheinlichkeit im dritten Zug aus Ω_3 ein T zu ziehen beträgt $\frac{4}{9}$.

Die Ergebnisse der Züge sind ohne gegenseitige Beeinflussung. Somit erhalten wir

$$P(\{(R, A, T)\}) = f_{\Omega_1}(R) f_{\Omega_2}(A) f_{\Omega_3}(T) = \frac{1}{9} \frac{1}{9} \frac{4}{9} = \frac{4}{729} = 0.005487$$

B) Die Wahrscheinlichkeit, im ersten Zug ein R , im zweiten ein A und im dritten ein T zu ziehen ist - wie gezeigt - $\frac{1}{9} \frac{1}{9} \frac{4}{9} = \frac{4}{729}$. Es gibt dabei $3!$ Reihenfolgen, ein R , ein A und ein T zu ziehen (diese „Reihenfolgen“ sind Tripel von R , A und T und Elemente von Ω). Jede hat dieselbe Wahrscheinlichkeit: Dies ergibt: $\frac{4 \cdot 3!}{729} = \frac{8}{243}$

b) ohne Zurücklegen:

A) Es geht um Ziehen in Reihenfolge ohne Zurücklegen:

$$\Omega = \{(\omega, y, z) : \omega, y, z \in \{S, T_1, U, T_2, T_3, G, A, R, T_4\} \text{ und } \omega \neq y \neq z \neq \omega\}$$

Das günstige Ereignis ist $\{(R, A, T_1), (R, A, T_2), (R, A, T_3), (R, A, T_4)\}$ mit einer Anzahl von 4 gleichwahrscheinlichen Ergebnisse. $|\Omega| = \frac{9!}{6!} = 504$.

Damit erhalten wir $\frac{4}{504} = 7.9365 \times 10^{-3}$

B) Es gibt für jede der vier Elemente von $\{(R, A, T_1), (R, A, T_2), (R, A, T_3), (R, A, T_4)\}$ $3! = 6$ Möglichkeiten, diese anzuordnen. Wir erhalten 24 „günstige“ Ergebnisse. Insgesamt gibt es $\binom{9}{3} = 84$ Möglichkeiten, Teilmengen mit drei verschiedenen Elementen aus Ω zu ziehen. Damit gilt: $\frac{24}{84} = 0.28571$

30. $\Omega = \Omega_1^{10}$, $\Omega_1 = \{Jan, Feb, \dots, Dez\}$. Wir setzen die gleiche Wahrscheinlichkeit der Ergebnisse von Ω_1 voraus (ist kaum haltbar). Die Wahrscheinlichkeit für eine Person, in einem bestimmten Monat Geburtstag zu haben, beträgt damit $\frac{1}{12}$. Die Wahrscheinlichkeit, nicht in diesem Monat Geburtstag zu haben, beträgt dann: $\frac{11}{12}$ (Gegenwahrscheinlichkeit von $\frac{1}{12}$). Die Wahrscheinlichkeiten für die verschiedenen Personen sind ohne gegenseitige Beeinflussung (auch das ist vermutlich z.B. bei Geschwistern nicht zutreffend). Die Wahrscheinlichkeit, dass drei spezifische Personen Geburtstag im Januar haben und die anderen nicht, beträgt dann: $(\frac{1}{12})^3 \cdot (\frac{11}{12})^7$. Nun gibt es $\binom{10}{3} = 120$ Möglichkeiten, aus 10 Personen 3 zu ziehen.

Jedes dieser Tripel hat dieselbe Wahrscheinlichkeiten, im Januar Geburtstag zu haben. Dies ergibt:

$\left(\frac{1}{12}\right)^3 \cdot \left(\frac{11}{12}\right)^7 \cdot 120 = 0.03\,776\,7$ als Wahrscheinlichkeit, dass genau 3 Personen im Januar Geburtstag haben.

b) Wir bilden die Summe darüber, dass keine Person, genau eine und genau zwei Personen im Januar Geburtstag haben. Die Gegenwahrscheinlichkeit ist dann die gesuchte Wahrscheinlichkeit: Für keine Person hat im Januar Geburtstag = alle haben an einem anderen Monat Geburtstag;

$$\left(\frac{11}{12}\right)^{10} = 0.418\,9$$

Für 1: Die Wahrscheinlichkeit für eine Person, im Januar Geburtstag zu haben, beträgt $\frac{1}{12}$.

Die Wahrscheinlichkeit, die übrigen nicht im Januar Geburtstag hat, beträgt dann $\left(\frac{11}{12}\right)^9$.

Die Wahrscheinlichkeit dafür, dass genau eine Person im Januar Geburtstag hat, beträgt also: $\left(\frac{11}{12}\right)^9 \cdot \frac{1}{12}$. Dies gilt nun für 10 Personen, die im Januar Geburtstag haben können:

Dies ergibt: $10 \cdot \left(\frac{11}{12}\right)^9 \cdot \frac{1}{12}$.

Für 2: Für die Wahrscheinlichkeit, dass genau zwei Personen Geburtstag haben, ergibt sich

$$\binom{10}{2} \cdot \left(\frac{1}{12}\right)^2 \cdot \left(\frac{11}{12}\right)^8 = 0.155\,79$$

Somit ergibt sich die Summe:

$$\left(\frac{11}{12}\right)^{10} + 10 \cdot \left(\frac{11}{12}\right)^9 \cdot \frac{1}{12} + \binom{10}{2} \cdot \left(\frac{1}{12}\right)^2 \cdot \left(\frac{11}{12}\right)^8 = 0.955\,52$$

Die Gegenwahrscheinlichkeit beträgt somit: $1 - 0.955\,52 = 0.0444\,8$

Eine weitere Berechnungsmethode würde darin bestehen, die Wahrscheinlichkeiten davon zu berechnen, dass genau 3, 4, 5, 6, 7, 8, 9, 10 Personen im Januar Geburtstag haben und die Summe davon zu bilden.

$$\left(\frac{1}{12}\right)^4 \cdot \left(\frac{11}{12}\right)^6 \cdot \binom{10}{4} = 6.008\,5 \times 10^{-3}$$

$$\left(\frac{1}{12}\right)^5 \cdot \left(\frac{11}{12}\right)^5 \cdot \binom{10}{5} = 6.554\,7 \times 10^{-4}$$

$$\left(\frac{1}{12}\right)^6 \cdot \left(\frac{11}{12}\right)^4 \cdot \binom{10}{6} = 4.965\,7 \times 10^{-5}$$

$$\left(\frac{1}{12}\right)^7 \cdot \left(\frac{11}{12}\right)^3 \cdot \binom{10}{7} = 2.579\,6 \times 10^{-6}$$

$$\left(\frac{1}{12}\right)^8 \cdot \left(\frac{11}{12}\right)^2 \cdot \binom{10}{8} = 8.794 \times 10^{-8}$$

$$\left(\frac{1}{12}\right)^9 \cdot \left(\frac{11}{12}\right)^1 \cdot \binom{10}{9} = 1.776\,6 \times 10^{-9}$$

$$\left(\frac{1}{12}\right)^{10} = 1.615\,1 \times 10^{-11}$$

$$\begin{aligned} &0.03\,776\,7 + 6.008\,5 \times 10^{-3} + 6.554\,7 \times 10^{-4} + 4.965\,7 \times 10^{-5} + \\ &2.579\,6 \times 10^{-6} + 8.794 \times 10^{-8} + 1.776\,6 \times 10^{-9} + 1.615\,1 \times 10^{-11} \\ &= 4.448\,3 \times 10^{-2} \quad (\text{Bei dieser Methode multiplizieren sich Rundungsfehler!}). \end{aligned}$$

c) Die Wahrscheinlichkeit dafür, dass genau je eine spezifische Person in den ersten 10 Monaten Geburtstag hat, beträgt: $\left(\frac{1}{12}\right)^{10}$. Da die Reihenfolge der Personen keine Rolle spielt, ergeben sich 10! Möglichkeiten, dass die zehn Personen in den ersten 10 Monaten Geburtstag haben. Die Wahrscheinlichkeit, dass je eine Person in den ersten 10 Monaten Geburtstage zu hat ist damit $10! \cdot \left(\frac{1}{12}\right)^{10}$.

Dies gilt für jede mögliche Kombination von 10 Monaten (aus 12 Monaten). Deren gibt es $\binom{12}{10} = 66$. Somit beträgt die Wahrscheinlichkeit, dass die 10 Personen an verschiedenen Monaten Geburtstag haben:

$$10! \cdot \left(\frac{1}{12}\right)^{10} \cdot \binom{12}{10} = 3.8681 \times 10^{-3}.$$

8.4 Lernziele

- Den Additionssatz für die Wahrscheinlichkeit von Ereignissen kennen und anwenden können.
- Den Begriff des Produktes von Ergebnismengen erklären können (Zusammenhang mit einer Serie von Zufallsexperimenten). Die Definition des Produktemasses und der Produktewahrscheinlichkeitsfunktion hinschreiben können. Den Begriff des Fehlens gegenseitiger Beeinflussung, der durch das Produktemass definiert wird, erläutern können. Berechnungen von Wahrscheinlichkeiten bezüglich Serien von Zufallsexperimenten bei fehlender gegenseitiger Beeinflussung durchführen können.
- Die Definition der bedingten Wahrscheinlichkeit kennen.
- Übungen von der Art der aufgelisteten Übungen [8.3.3 1](#) - [8.3.3 20](#) (inkl.) lösen können.

Kapitel 9

Zufallsvariablen

9.1 Zufallsvariablen und Bildmasse

Zufallsvariablen sind Abbildungen von einer Ergebnismenge Ω in die reellen Zahlen, die wir unter Anderem aus folgenden zwei Gründen einführen:

(i) Ergebnisse sind im allgemeinen keine Zahlen. Deshalb kann man mit den Ergebnissen nicht rechnen. Wollen wir z.B. berechnen, welchen Ausgang eines Zufallsexperimentes wir im Mittel zu erwarten haben, so müssen wir statt Ergebnissen Zahlen zur Verfügung haben. Zufallsvariablen sind Funktionen, die u.a. diesem Missstand abhelfen. Sie bilden Ergebnismengen in die reellen Zahlen ab.

Beispiel 9.1.1. *Zufallsexperiment Würfeln (Würfel mit sechs Farben); Ergebnismenge $\Omega = \{\text{rot, gelb, schwarz, blau, violett, weiss}\}$. Graph einer möglichen Zufallsvariable $X = \{(\text{rot}, 1), (\text{gelb}, 2), (\text{schwarz}, 3), (\text{blau}, 4), (\text{violett}, 5), (\text{weiss}, 6)\}$. $\text{Bild}(X) = \{1, 2, 3, 4, 5, 6\}$. Das Bild von X enthält nun Zahlen.* $\diamond$

Beispiel 9.1.2. *Zufallsexperiment Würfeln (Würfel mit sechs Augen); Ergebnismenge $\Omega = \{1 \text{ Auge, } 2 \text{ Augen, } 3 \text{ Augen, } 4 \text{ Augen, } 5 \text{ Augen, } 6 \text{ Augen}\}$. Graph einer anderen möglichen Zufallsvariable $Y = \{(1 \text{ Auge}, 1), (2 \text{ Augen}, 1), (3 \text{ Augen}, 1), (4 \text{ Augen}, 1), (5 \text{ Augen}, 10), (6 \text{ Augen}, 1)\}$. Auch das Bild dieser Zufallsvariable enthält Zahlen statt Ergebnisse. Die Zufallsvariable Y könnte z.B. die Auszahlungen bei einem Glücksspiel mit einem Würfel festlegen.* $\diamond$

(ii) Das letzte Beispiel führt uns zu einem weiteren Vorteil von Zufallsvariablen. Wir können mit Zufallsvariablen flexibel Situationen wie z.B. Glücksspiele modellieren.

Wir führen Zufallsvariablen noch formell durch eine Definition ein:

Definition 9.1.3. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum. X ist eine Zufallsvariable, wenn X eine Abbildung ist mit:

$$X : \Omega \rightarrow \mathbb{R}$$

Definition 9.1.4. Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable. Dann schreiben wir für das Bild von X

$$\mathfrak{X} := \text{Bild}(X)$$

und das Bild von X_i

$$\mathfrak{X}_i := \text{Bild}(X_i)$$

Es gibt überabzählbar unendlich viele mögliche Zufallsvariablen auf Ω . Da wir diskrete Ergebnismengen betrachten, sind die Bilder der Zufallsvariablen ebenfalls diskret.

Man beachte: Zufallsvariablen sind keine Variablen, sondern Funktionen. Als Funktionen sind Zufallsvariablen Tripel $(\Omega, \mathbb{R}, g)$ mit einem Definitionsbereich Ω (= Ergebnismenge), dem Wertebereich $\mathbb{R}$ und einem Graphen (= Menge von geordneten Paaren). Mit Zufall haben sie diesbezüglich nichts zu tun. Sie existieren, wenn sie definiert sind, als mathematische Objekte unabhängig von Zeit und Raum und sind nicht irgendwie aktiv. Die Begriffswahl „Zufallsvariable“ ist vermutlich nicht sehr geschickt, da sie zu vielen Missverständnissen Anlass gibt. Der Begriff hat sich aber eingebürgert. Zufallsvariablen werden in manchen Schriften auch „Zufallsgrößen“ genannt. Der „Zufall“ kommt bei Zufallsvariablen allerdings dann ins Spiel, wenn wir die konkreten Ausgänge von Zufallsexperimenten betrachten. Wir definieren:

Definition 9.1.5. - Der Funktionswert $X(\omega)$ einer Zufallsvariable X realisiert sich (= tritt ein), wenn ω eintritt.

- Sei $\mathfrak{X}$ das Bild der Zufallsvariable X und seien Ereignisse aus $\mathfrak{X}$ als Teilmengen von $\mathfrak{X}$ definiert: Ein Ereignis A aus $\mathfrak{X}$ tritt ein, wenn es ein eintretendes $\omega \in \Omega$ gibt, so dass $X(\omega) \in A$.

Die Wahrscheinlichkeiten, die durch die Wahrscheinlichkeitsfunktion auf Ω und durch die Wahrscheinlichkeitsverteilung auf $\mathcal{P}(\Omega)$ definiert sind, werden durch die Zufallsvariablen auf $\mathfrak{X}$ „transportiert“, wenn man folgende Definitionen vornimmt:

Definition 9.1.6. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum.

Sei X eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ und $\mathfrak{X}$ das Bild von X . $f : \Omega \rightarrow [0, 1]$ sei die durch P festgelegte Wahrscheinlichkeitsfunktion auf Ω . Dann definieren wir:

$$f_X : \mathfrak{X} \rightarrow [0, 1]$$

$$f_X(x) := \sum_{X(\omega)=x} f(\omega).$$

(für $\omega \in \Omega$).

Wir nennen f_X „die Wahrscheinlichkeitsfunktion auf das Bild von X “ oder „Bildwahrscheinlichkeitsfunktion von X “. $\sum_{X(\omega)=x} f(\omega)$ ist die Summe aller Wahrscheinlichkeiten $f(\omega)$, so dass X den ω die Zahl x zuordnet (s. für ein Beispiel die folgende Abbildung 9.1.1).

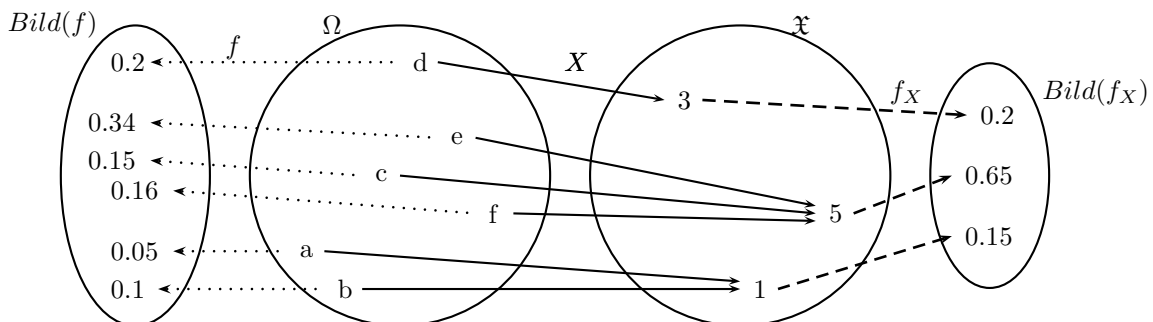


Abbildung 9.1.1: „Transport“ der Wahrscheinlichkeitsfunktion auf Ω durch die Zufallsvariable X auf das Bild $\mathfrak{X}$ von X . Graph von $X = \{(a, 1), (b, 1), (c, 5), (e, 5), (f, 5), (d, 3)\}$. Graph von $f = \{(a, 0.05), (b, 0.1), (c, 0.15), (e, 0.34), (f, 0.16), (d, 0.2)\}$. Durch X auf $\mathfrak{X}$ erzeugter Graph der Bildwahrscheinlichkeitsfunktion $f_X = \{(1, 0.15), (3, 0.2), (5, 0.65)\}$

Die folgenden drei Beispiele sollten sehr genau studiert werden:

Beispiel 9.1.7. (s. Beispiel 9.1.1): Zufallsexperiment Würfeln (Würfel mit sechs Farben); $\Omega = \{\text{rot, gelb, schwarz, blau, violett, weiss}\}$. Es gelte $f(\omega) = \frac{1}{6}$ für $\omega \in \Omega$ (Gleichverteilung). Mit der Zufallsvariable $X = \{(\text{rot}, 1), (\text{gelb}, 2), (\text{schwarz}, 3), (\text{blau}, 4), (\text{violett}, 5), (\text{weiss}, 6)\}$ gilt damit: 1 tritt genau dann ein, wenn „rot“ eintritt. Die Wahrscheinlichkeit, dass 1 eintritt, ist damit identisch mit der Wahrscheinlichkeit, dass „rot“ eintritt. Formal: $f_X(1) = \sum_{X(\omega)=1} f(\omega) = f(\text{rot}) = \frac{1}{6}$.

Allgemein: $f_X(x) = \sum_{X(\omega)=x} f(\omega) = f(\text{rot}) = f(\text{gelb}) = f(\text{schwarz}) =$

$f(\text{blau}) = f(\text{violett}) = f(\text{weiss}) = \frac{1}{6}$. (Spezialfall einer bijektiven Zufallsvariable mit $f_X(x) = f(\omega)$ für $X(\omega) = x$).

$\mathfrak{X} = \{1, 2, 3, 4, 5, 6\}$

Graph von $f_X : \{(1, \frac{1}{6}), (2, \frac{1}{6}), (3, \frac{1}{6}), (4, \frac{1}{6}), (5, \frac{1}{6}), (6, \frac{1}{6})\}$

◇

Beispiel 9.1.8. (s. Beispiel 9.1.2): Zufallsexperiment Würfeln (Würfel mit sechs Augen); Ergebnismenge $\Omega = \{1 \text{ Auge, } 2 \text{ Augen, } 3 \text{ Augen, } 4 \text{ Augen, } 5 \text{ Augen, } 6 \text{ Augen}\}$. Es gelte $f(\omega) = \frac{1}{6}$ für $\omega \in \Omega$. Mit der möglichen Zufallsvariable $Y = \{(1 \text{ Auge, } 1), (2 \text{ Augen, } 1), (3 \text{ Augen, } 1), (4 \text{ Augen, } 1), (5 \text{ Augen, } 10), (6 \text{ Augen, } 1)\}$ gilt dann: Die 1 tritt ein genau dann, wenn 1 Auge, 2 Augen, 3 Augen, 4 Augen oder 6 Augen gewürfelt werden. Damit beträgt die Wahrscheinlichkeit, dass 1 eintritt (d.h. dass die Zufallsvariable Y den Wert 1 annimmt) $\frac{5}{6}$. Die Wahrscheinlichkeit, dass 10 eintritt beträgt $\frac{1}{6}$. Formal: $f_Y(1) = \sum_{Y(\omega)=1} f(\omega) = f(1 \text{ Auge}) + f(2 \text{ Augen}) + f(3 \text{ Augen}) +$

$f(4 \text{ Augen}) + f(6 \text{ Augen}) = \frac{5}{6}$.

$f_Y(10) = \sum_{Y(\omega)=10} f(\omega) = f(5 \text{ Augen}) = \frac{1}{6}$.

Das Bild von $Y : \mathfrak{X} = \{1, 10\}$.

Graph von $f_Y = \{(1, \frac{5}{6}), (10, \frac{1}{6})\}$.

◇

Beispiel 9.1.9. Zufallsexperiment Würfeln (Würfel mit sechs Augen); Ergebnismenge $\Omega = \{1 \text{ Auge, } 2 \text{ Augen, } 3 \text{ Augen, } 4 \text{ Augen, } 5 \text{ Augen, } 6 \text{ Augen}\}$. Auf Grund von mehreren Tausend Würfeln ist man zum Schluss gekommen, dass die folgende Wahrscheinlichkeitsfunktion zu diesem Würfel passt: Graph von $f = \{(1 \text{ Auge, } \frac{1}{7}), (2 \text{ Augen, } \frac{3}{16}), (3 \text{ Augen, } \frac{19}{112}), (4 \text{ Augen, } \frac{1}{6}), (5 \text{ Augen, } \frac{4}{17}), (6 \text{ Augen, } \frac{5}{51})\}$

Wir überprüfen, ob es sich um eine Wahrscheinlichkeitsfunktion handelt:

$$\frac{1}{7} + \frac{3}{16} + \frac{19}{112} + \frac{1}{6} + \frac{4}{17} + \frac{5}{51} = 1$$

Wir spielen mit diesem Würfel um folgendes Spiel: bei 1, 2 und 3 Augen erfolgt keine Auszahlung, bei 4, 5 und 6 Augen erfolgt eine Auszahlung von 10 Sfr. Dies entspricht der folgenden Zufallsvariable: Graph von $X = \{(1 \text{ Auge, } 0), (2 \text{ Augen, } 0), (3 \text{ Augen, } 0), (4 \text{ Augen, } 10), (5 \text{ Augen, } 10), (6 \text{ Augen, } 10)\}$

$$f_X(0) = \sum_{X(\omega)=0} f(\omega) = \frac{1}{7} + \frac{3}{16} + \frac{19}{112} = \frac{1}{2}$$

$$f_X(10) = \sum_{X(\omega)=10} f(\omega) = \frac{1}{6} + \frac{4}{17} + \frac{5}{51} = \frac{1}{2}.$$

◇

Beachten Sie: f_X hängt von P und von X ab!

Mit Hilfe der Wahrscheinlichkeitsfunktion auf das Bild von X können wir nun das von der Zufallsvariable von Ω auf das Bild von X transportierte Wahrscheinlichkeitsmass definieren. Das Vorgehen entspricht dabei der Definition des Wahrscheinlichkeitsmasses auf Ω mit Hilfe der Wahrscheinlichkeitsfunktion auf Ω .

Definition 9.1.10. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum.

Sei X eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ und $\mathfrak{X}$ das Bild von X . $f_X : \mathfrak{X} \rightarrow [0, 1]$ die von P bestimmte Wahrscheinlichkeitsfunktion auf das Bild von X .

Dann definieren wir:

$$P_X : \mathcal{P}(\mathfrak{X}) \rightarrow [0, 1]$$

$$P_X(A) := \sum_{x \in A} f_X(x)$$

($\mathcal{P}(\mathfrak{X})$ ist die Potenzmenge des Bildes von X ; $A \in \mathcal{P}(\mathfrak{X})$)

P_X ist die Wahrscheinlichkeitsverteilung auf das Bild von X unter P ($P_X = \text{Bildmass}$). Wir nennen P_X auch die Verteilung der Zufallsvariable X (unter P). P_X hängt sowohl von P als auch von X ab!

Beispiel 9.1.11. (s. Beispiel 9.1.7) Zufallsexperiment Würfeln (Würfel mit sechs Farben); $\Omega = \{\text{rot, gelb, schwarz, blau, violett, weiss}\}$. Es gelte $f(\omega) = \frac{1}{6}$ für $\omega \in \Omega$ (Gleichverteilung). Zufallsvariable $X = \{(\text{rot}, 1), (\text{gelb}, 2), (\text{schwarz}, 3), (\text{blau}, 4), (\text{violett}, 5), (\text{weiss}, 6)\}$. Es gilt damit: $P_X(\{x\}) = \sum_{\omega \in \{x\}} f_X(\omega) = f_X(x) = f(\omega)$ mit $X(\omega) = x$ für $x \in \mathfrak{X}$. (Spezialfall einer bijektiven Zufallsvariable X).

Der Graph von P_X enthält 64 Elemente. Wir verzichten auf ein Ausschreiben, geben aber dafür zwei Elemente des Graphen an: $(\{1\}, \frac{1}{6})$; $(\{2, 5, 6\}, \frac{1}{2})$.

Im Beispiel ist das Bildmass einer Gleichverteilung P wiederum gleichverteilt.

Wir sagen, die Zufallsvariable X sei gleichverteilt. $\diamond$

Beispiel 9.1.12. (s. Beispiel 9.1.8) Zufallsexperiment Würfeln (Würfel mit sechs Augen); Ergebnismenge $\Omega = \{1 \text{ Auge, } 2 \text{ Augen, } 3 \text{ Augen, } 4 \text{ Augen, } 5 \text{ Augen, } 6 \text{ Augen}\}$. Es gelte $f(\omega) = \frac{1}{6}$ für $\omega \in \Omega$. Mit der möglichen Zufallsvariable $Y = \{(1 \text{ Auge, } 1), (2 \text{ Augen, } 1), (3 \text{ Augen, } 1), (4 \text{ Augen, } 1), (5 \text{ Augen, } 10), (6 \text{ Augen, } 1)\}$ gilt dann: Das Bild von Y ist $\{1, 10\}$.

Die Potenzmenge von $\{1, 10\}$ ist $\{\emptyset, \{1\}, \{10\}, \{1, 10\}\}$.

Das Bildmass P_X hat als Graphen: $(\{\emptyset, 0\}, (\{1\}, \frac{5}{6}), (\{10\}, \frac{1}{6}), (\{1, 10\}, 1))$ $\diamond$

Beispiel 9.1.13. (s. Beispiel 9.1.9) Zufallsexperiment Würfeln (Würfel mit sechs Augen); Ergebnismenge $\Omega = \{1 \text{ Auge, } 2 \text{ Augen, } 3 \text{ Augen, } 4 \text{ Augen, } 5 \text{ Augen, } 6 \text{ Augen}\}$. Auf Grund von mehreren Tausend Würfen ist man zum Schluss gekommen, dass die folgende Wahrscheinlichkeitsfunktion zu diesem Würfel passt: Graph von $f = \{(1 \text{ Auge, } \frac{1}{7}), (2 \text{ Augen, } \frac{3}{16}), (3 \text{ Augen, } \frac{19}{112}), (4 \text{ Augen, } \frac{1}{6}), (5 \text{ Augen, } \frac{4}{17}), (6 \text{ Augen, } \frac{5}{51})\}$. Graph von $X = \{(1 \text{ Auge, } 0), (2 \text{ Augen, } 0), (3 \text{ Augen, } 0), (4 \text{ Augen, } 10), (5 \text{ Augen, } 10), (6 \text{ Augen, } 10)\}$

$\mathfrak{X} = \{0, 10\}$

$\mathcal{P}(\mathfrak{X}) = \{\emptyset, \{0\}, \{10\}, \{0, 10\}\}$. Da $f_X(0) = \frac{1}{2}$ erhält man für das Bildmass P_X den Graphen: $(\{\emptyset, 0\}, (\{0\}, \frac{1}{2}), (\{10\}, \frac{1}{2}), (\{0, 10\}, 1))$ $\diamond$

Wir führen eine oft gebrauchte und die Intuition stützende Schreibweise ein:

Definition 9.1.14. Sei $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum.

Sei X eine Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ und $A \subset \mathfrak{X}$ (= Bild von X) und $a, b \in \mathbb{R}$.

$P(X \in A) := P_X(A)$ (= Die Wahrscheinlichkeit, dass die Zufallsvariable einen Wert aus A annimmt).

$P(a \leq X \leq b) := P(X \in [a, b] \cap \mathfrak{X}) = P_X(\{x | a \leq x \leq b\} \cap \mathfrak{X})$

$P(X \leq b) := P(X \in]-\infty, b] \cap \mathfrak{X})$

$P(X \geq b) := P(X \in [b, \infty[\cap \mathfrak{X})$

$P(X = a) := P(X \in \{a\})$ (etc. d.h. analog z.B. für $<, >$)

Beispiel 9.1.15. Für $P_X(A)$ mit $A = \{5\}$ schreiben wir: $P(X = 5)$, gelesen als „die Wahrscheinlichkeit, dass der Wert der Zufallsvariable 5 ist“.

Für $P_X(A)$ mit $A = \{x : x \geq 3\} \cap \mathfrak{X}$ schreiben wir: $P(X \geq 3)$, gelesen als „die Wahrscheinlichkeit, dass der Wert der Zufallsvariable grösser gleich 3 ist“. $\diamond$

Definition 9.1.16. (Verteilungsfunktion): Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathfrak{X}$ das Bild einer Zufallsvariable X auf Ω . Die Funktion

$$F: \mathfrak{X} \rightarrow [0, 1]$$

$$F(x) := P(X \leq x)$$

wird „Verteilungsfunktion der Zufallsvariable X unter P “ genannt.

$F(x)$ gibt die Wahrscheinlichkeit an, dass die Zufallsvariable X einen Wert kleiner-gleich x annimmt. F hängt von X und von P ab. Um die Abhängigkeit von X auszudrücken, verwenden wir manchmal F_X für F . Verteilungsfunktionen entsprechen den empirischen Verteilungsfunktionen und können mit denselben graphischen Mitteln dargestellt werden. Man beachte: Verteilungsfunktionen sind etwas anderes als Verteilungen.

Beispiel 9.1.17. (1) Beim Graphen $f_Z = \{(1, \frac{1}{6}), (2, \frac{1}{6}), (3, \frac{1}{6}), (4, \frac{1}{6}), (5, \frac{1}{6}), (6, \frac{1}{6})\}$ ist der Graph von $F = \{(1, \frac{1}{6}), (2, \frac{2}{6}), (3, \frac{3}{6}), (4, \frac{4}{6}), (5, \frac{5}{6}), (6, \frac{6}{6})\}$
 (2) Würfelspiel mit Augenzahlen (Gleichverteilung) und Auszahlungen (0 für die Augenzahlen 1, 2, 3 oder 4; 5 für die Augenzahl 5; 6 für die Augenzahl 6): Beim Graphen $f_Z = \{(0, \frac{4}{6}), (5, \frac{1}{6}), (6, \frac{1}{6})\}$ ist der Graph von $F = \{(0, \frac{4}{6}), (5, \frac{5}{6}), (6, 1)\}$
 (3) Beim Graphen $f_X = \{(4, \frac{1}{2}), (2, \frac{1}{2})\}$ ist der Graph von $F = \{(2, \frac{1}{2}), (4, 1)\}$. $\diamond$

Die Verteilungsfunktion wurde mit Hilfe des Wahrscheinlichkeitsmasses definiert. Umgekehrt kann mit Hilfe der Verteilungsfunktion das Wahrscheinlichkeitsmass und damit auch die Wahrscheinlichkeitsfunktion berechnet werden. Sei y der zu x nächstkleinere Wert der Zufallsvariable. Es gilt: $P(X = x) = P(X \leq x) - P(X < x) = F(x) - F(y)$. Wir betrachten das Beispiel 9.1.17 (1): $P(X = 3) = P(X \leq 3) - P(X < 3) = F(3) - F(2) = \frac{3}{6} - \frac{2}{6} = \frac{1}{6}$. Allgemein gilt für $x < y$: $P(x < X \leq y) = P(X \leq y) - P(X \leq x) = F(y) - F(x)$. Wir betrachten wiederum das Beispiel 9.1.17 (1):

So ist z.B. $P(2 < X \leq 5) = P(X \leq 5) - P(X \leq 2) = F(5) - F(2) = \frac{5}{6} - \frac{2}{6} = \frac{3}{6}$

Diesen Umstand werden wir uns zunutze machen, sobald wir für stetige Wahrscheinlichkeitsräume das Wahrscheinlichkeitsmass nicht mehr mit Hilfe von Wahrscheinlichkeitsfunktionen auf Ergebnisse definieren können.

Bemerkung 9.1.18. Beachten Sie, dass die Verteilungsfunktion auf das Bild einer Zufallsvariable definiert ist, nicht jedoch auf der Ergebnismenge selber. Verteilungsfunktionen sind nur sinnvoll konstruierbar, wenn der Definitionsbereich geordnet ist. Dies ist bei Zahlen der Fall, nicht jedoch bei den Elementen der Ergebnismenge (wie z.B. Farben bei einem Farbenwürfel). $\diamond$

Bemerkung 9.1.19. Bei diskreten Zufallsvariablen gilt $P(X = x) = f_X(x)$. $\diamond$

Eine Übersicht über die eingeführten Funktionen finden Sie in der Abbildung 9.1.2.

9.2 Erwartungswert einer Zufallsvariable

Mit dem Begriff des Erwartungswertes wird klarer, wieso Zufallsvariablen eingeführt werden. Der Erwartungswert einer Zufallsvariable wird nach dem Vorbild des Mittelwertes empirischer Daten konstruiert. Der Erwartungswert entspricht bei Glücksspielen dem zu erwartenden durchschnittlichen Gewinn (Verlust) bei häufigem Spielen des gleichen Spiels. Wir können ein spezifisches Würfelspiel durch eine Wahrscheinlichkeitsfunktion f_X charakterisieren (die Gewinne an erster Stelle und die Wahrscheinlichkeiten an zweiter Stelle):

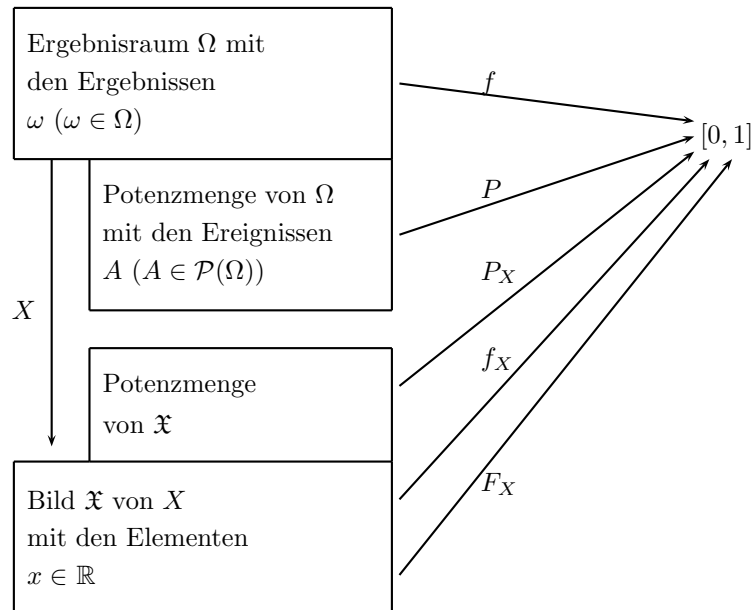


Abbildung 9.1.2: Überblick über die eingeführten Funktionen

- f = Wahrscheinlichkeitsfunktion auf Ergebnismenge Ω
- P = Wahrscheinlichkeitsverteilung auf Potenzmenge der Ergebnismenge Ω
- X = Zufallsvariable auf Ergebnismenge Ω
- f_X = Wahrscheinlichkeitsfunktion auf Bild $\mathfrak{X}$ von X
- P_X = Bildmass = Wahrscheinlichkeitsverteilung auf Potenzmenge von $\mathfrak{X}$.
- F_X = Verteilungsfunktion auf $\mathfrak{X}$

Beispiel 9.2.1. Ein bestimmtes Glücksspiel sei durch den folgenden Graphen einer Bildwahrscheinlichkeitsfunktion f_X

$$\text{Graph von } f_X = \left\{ \left(10, \frac{1}{6} \right), \left(2, \frac{1}{6} \right), \left(0, \frac{2}{3} \right) \right\}$$

bestimmt, wobei eine Auszahlung von 10 beim Würfeln von 6 Augen, eine Auszahlung von 2 beim Würfeln von einem Auge und keine Auszahlung bei den übrigen Augen erfolgt. Es stellt sich die Frage, welchen durchschnittlichen Gewinn wir zu erwarten haben, wenn wir das Spiel ein paar Tausend Mal spielen. Der Erwartungswert ist der Wert, von dem der durchschnittliche Gewinn mit angegebener Wahrscheinlichkeit immer weniger abweicht, je mehr Spiele wir machen. $\diamond$

Da wir den Erwartungswert nach dem Vorbild des Mittelwertes konstruieren und Wahrscheinlichkeiten „idealisierte“ relative Häufigkeiten sind, wiederholen wir die Formel für die Berechnung des arithmetischen Mittels mit Hilfe relativer Häufigkeiten:

Für klassierte Daten gilt (siehe S. 44, beschreibende Statistik):

$$\bar{x} = \sum_{i=1}^m (h(x_i) \cdot x_i)$$

wobei $h(x_i)$ die relative Häufigkeit des Auftretens von x_i ist und m die Anzahl der Klassen (Ausprägungen). Da Wahrscheinlichkeiten ein Modell für relative Häufigkeiten sind, ist es naheliegend, den Erwartungswert einer Zufallsvariable X wie folgt zu definieren (mit $P(X = x)$ = Wahrscheinlichkeit, dass die Zufallsvariable X den Wert x annimmt).

Definition 9.2.2. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable auf Ω und $\mathfrak{X}$ das Bild von Ω unter X . Der Erwartungswert $E(X)$ der Zufallsvariable X unter der diskreten Wahrscheinlichkeitsverteilung P ist definiert als:

$$E(X) = \sum_{x \in \mathfrak{X}} [P(X = x) \cdot x].$$

(sofern $\sum_{x \in \mathfrak{X}} [P(X = x) \cdot x]$ definiert ist).

Der Erwartungswert einer Zufallsvariable ist damit die Summe der - mit der Wahrscheinlichkeit ihres Eintreffens gewichteten - Werte der Zufallsvariablen.

Beachten Sie, dass der Erwartungswert einer Zufallsvariable bezüglich einer spezifischen Verteilung P definiert ist. Betrachten wir künftig verschiedene Verteilungen, werden wir diese Abhängigkeit durch die Abkürzung $E_P(X)$ berücksichtigen.

Satz 9.2.3. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable auf Ω . Dann gilt, sofern die definierende Summe definiert ist:

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\})$$

Beweis. Laut Definitionen gilt: $P(X = x) = \sum_{X(\omega)=x} P(\{\omega\})$ und $E(X) = \sum_{x \in \mathfrak{X}} xP(X = x)$. Damit

$$\text{gilt: } \sum_{x \in \mathfrak{X}} xP(X = x) = \sum_{x \in \mathfrak{X}} \left(\sum_{X(\omega)=x} X(\omega)P(\{\omega\}) \right) = \sum_{\omega \in \Omega} X(\omega)P(\{\omega\}).$$

(Da wir über alle x des Bildes von X summieren und über alle ω , denen die Zufallsvariable X ein x zuordnet, entspricht dies der Summation über Ω). $\square$

Beispiel 9.2.4. Eine Münze wird geworfen. Zufallsvariable: Bei Kopf wird 5 Fr. ausgezahlt, bei Zahl 0.

Ergebnismenge $\Omega = \{K, Z\}$

Wahrscheinlichkeitsfunktion auf Ω : Graph von $f = \{(K, \frac{1}{2}), (Z, \frac{1}{2})\}$

(gleiche Wahrscheinlichkeit der Ergebnisse)

Wahrscheinlichkeitsverteilung auf Ω : Graph von $P = \{(\{K\}, \frac{1}{2}), (\{Z\}, \frac{1}{2}), (\{K, Z\}, 1), (\emptyset, 0)\}$

Graph der Zufallsvariable $X = \{(K, 5), (Z, 0)\}$

Bild $\mathfrak{X}$ von $X = \{5, 0\}$

Graph der Wahrscheinlichkeitsfunktion auf $\mathfrak{X}$: $f_X = \{(5, \frac{1}{2}), (0, \frac{1}{2})\}$

Graph der Wahrscheinlichkeitsverteilung auf $\mathfrak{X}$:

$P_X = \{(\{5\}, \frac{1}{2}), (\{0\}, \frac{1}{2}), (\{0, 5\}, 1), (\emptyset, 0)\}$

Erwartungswert $E(X)$ der Zufallsvariable X unter P .

$$E(X) = \sum_{x \in \{5, 0\}} P(X = x) \cdot x = \frac{1}{2} \cdot 5 + \frac{1}{2} \cdot 0 = \frac{5}{2} = 2.5$$

Im Beispiel entspricht der Erwartungswert dem zu erwartenden durchschnittlichen Gewinn bei vielfachem Münzwurf. $\diamond$

Beispiel 9.2.5. R verdient sein Leben mit Glücksspielen auf der Strasse. Dazu verwendet er eine Münze, von der er aufgrund der Erfahrung weiss, dass Kopf mit einer Wahrscheinlichkeit von 0.6 auftaucht. Er bietet Passanten folgendes Spiel an: Bei Kopf wird 0 Fr. ausgezahlt, bei Zahl 5 Fr. Der Einsatz des spielwilligen Passanten beträgt 2.5. Wieviel gewinnt R durchschnittlich pro Spiel?

$\Omega = \{K, Z\}$;

Graph von $f = \{(K, 0.6), (Z, 0.4)\}$

Graph der Wahrscheinlichkeitsverteilung auf Ω : $P = \{(\{K\}, 0.6), (\{Z\}, 0.4), (\{K, Z\}, 1), (\emptyset, 0)\}$

Graph von $X = \{(K, 0), (Z, 5)\}$;

Bild $\mathfrak{X}$ von $X = \{0, 5\}$;

Graph von $f_X = \{(0, 0.6), (5, 0.4)\}$

$$\text{Graph von } P_X = \{(\{0\}, 0.6), (\{5\}, 0.4), (\{0, 5\}, 1), (\emptyset, 0)\}$$

$$E(X) = \sum_{x \in \{5, 0\}} x \cdot P(X = x) = 0 \cdot 0.6 + 5 \cdot 0.4 = 2 \text{ Fr.}$$

Soviel gewinnt ein spielwilliger Passant im Durchschnitt (Soviel muss R im Durchschnitt auszahlen). Somit gewinnt R pro Spiel durchschnittlich 0.5 Fr (= Einsatz - Auszahlung) In 100 Spielen ergibt dies 50 Fr. (Für den Passanten ist das Spiel kein „faires Spiel“, da der Einsatz und der Erwartungswert nicht identisch sind. Es ist zudem zu vermuten, dass der spielwillige Passant davon ausgeht, dass der Graph von $f_X = \{(0, 0.5), (5, 0.5)\}$ und damit $E(X) = 2.5$. Damit würden die Passanten von R getäuscht. (Andererseits kann kein vernünftiger Passant glauben, dass R gratis arbeitet?!?). $\diamond$

Beispiel 9.2.6. Zwei Münzen werden geworfen. Bei zweimal Kopf wird 5 Fr. ausbezahlt, sonst 1 Fr. Welches ist ein Einsatz, der das Spiel zu einem fairen Spiel macht?

$$\Omega = \{(K, K), (K, Z), (Z, K), (Z, Z)\}$$

$$\text{Graph von } f = \{((K, K), 0.25), ((K, Z), 0.25), ((Z, K), 0.25), ((Z, Z), 0.25)\}$$

(gleiche Wahrscheinlichkeit der Ergebnisse)

Die Angabe des Graphen von P wird mühsam. Beispielshalber liefern wir sie:

$$\begin{aligned} \text{Graph von } P = & \{(\emptyset, 0), \\ & (\{(K, K)\}, 0.25), (\{(K, Z)\}, 0.25), (\{(Z, K)\}, 0.25), (\{(Z, Z)\}, 0.25) \\ & (\{(K, K), (K, Z)\}, 0.5), (\{(K, K), (Z, K)\}, 0.5), (\{(K, K), (Z, Z)\}, 0.5), \\ & (\{(Z, K), (K, Z)\}, 0.5), (\{(Z, K), (Z, Z)\}, 0.5), \\ & (\{(K, Z), (Z, Z)\}, 0.5), \\ & (\{(K, K), (K, Z), (Z, K)\}, 0.75), (\{(K, K), (K, Z), (Z, Z)\}, 0.75), \\ & (\{(K, K), (Z, K), (Z, Z)\}, 0.75), (\{(K, Z), (Z, K), (Z, Z)\}, 0.75), (\{(K, K), (Z, K), (K, Z), (Z, Z)\}, 1)\} \end{aligned}$$

$$\text{Graph von } X = \{((K, K), 5), ((K, Z), 1), ((Z, K), 1), ((Z, Z), 1)\}$$

$$\text{Bild von } X = \{5, 1\}$$

$$\text{Graph von } f_X = \{(5, 0.25), (1, 0.75)\}$$

$$\text{Graph von } P_X = \{(\emptyset, 0), (\{1\}, 0.75), (\{5\}, 0.25), (\{1, 5\}, 1)\}$$

$$E(X) = 5 \cdot 0.25 + 1 \cdot 0.75 = 2 \text{ Fr.}$$

Ein Einsatz von 2 Fr. würde das Spiel zu einem „fairen“ machen. $\diamond$

Beispiel 9.2.7. Um Geld für die Glückskette zu sammeln, ruft ein Verein seine Mitglieder auf, Wein zur GV zu bringen. Es soll ein Glücksspiel veranstaltet werden, um den Wein in Geld umzusetzen. Dazu wurde ein Glücksrad ausgeliehen, das mit einer Wahrscheinlichkeit von 0.1 auf der Zahl 1, mit einer Wahrscheinlichkeit von 0.2 auf der Zahl 2, mit einer Wahrscheinlichkeit von 0.3 auf der Zahl 3 und mit einer Wahrscheinlichkeit von 0.4 auf der Zahl 4 stehen bleibt (die Flächen des Kreises sind entsprechend gross gewählt). Für einen Einsatz von 20 Fr. kann man folgende Gewinne machen: bleibt der Zeiger auf 1 stehen, werden 3 Flaschen abgegeben, bleibt der Zeiger auf 2 stehen, wird eine Flasche abgegeben. In den übrigen Fällen wird nichts abgegeben. Man berechne den Erwartungswert.

$$\Omega = \{1, 2, 3, 4\}$$

$$\text{Graph von } f = \{(1, 0.1), (2, 0.2), (3, 0.3), (4, 0.4)\}$$

$$\text{Graph von } X = \{(1, 3), (2, 1), (3, 0), (4, 0)\}$$

$$\text{Bild von } X = \{3, 1, 0\}$$

$$\text{Graph von } f_X = \{(3, 0.1), (1, 0.2), (0, 0.7)\}$$

$$E(X) = 3 \cdot 0.1 + 1 \cdot 0.2 + 0 \cdot 0.7 = 0.5 \text{ Flaschen.}$$

Durchschnittlich gewinnt man somit 0.5 Flaschen pro Spiel. Damit würde eine Flasche durchschnittlich 40 Fr. kosten. $\diamond$

Es folgen ein paar Sätze, die eine wichtige Rolle spielen. Ordnet X allen $\omega \in \Omega$ dieselbe Zahl a zu - d.h. $X(\omega) = a$ für $\omega \in \Omega$, so schreiben wir statt $E(X)$ kurz $E(a)$.

Satz 9.2.8. $E(a) = a$ (a ist eine beliebige reelle Zahl).

Beweis. Offensichtlich gilt für das Bild $\mathfrak{X}$ von $X : P_X(\mathfrak{X}) = 1$, denn

$$P_X(\mathfrak{X}) \stackrel{9.1.10}{=} \sum_{x \in \mathfrak{X}} f_X(x) \stackrel{9.1.6}{=} \sum_{x \in \mathfrak{X}} \sum_{X(\omega)=x} f(\omega) = \sum_{\omega \in \Omega} f(\omega) = 1.$$

Die Zufallsvariable X ordnet allen Werten des Ergebnisraums Ω dieselbe Zahl a zu. Damit ist $\mathfrak{X} = \{a\}$ und $P(X = a) = P_X(\{a\}) = 1$. Entsprechend gilt in die Definition des Erwartungswertes $E(X) = \sum_{x \in \mathfrak{X}} xP(X = x)$ eingesetzt:

$$E(X) = \sum_{x \in \{a\}} xP(X = x) = aP(X = a) = a \cdot 1 = a.$$

□

Intuitive Begründung: Ist die Zufallsvariable eine konstante Funktion - sie ordnet jedem ω der Ergebnismenge die Zahl a zu - so ist zu erwarten, dass „im Mittel“ a auftaucht.

Rechnerisches Beispiel: $\Omega = \{1, 3, 4\}$. $X(\omega) = a$ für alle $\omega \in \Omega$.

Graph von $f_X = \{(a, 1)\}$.

Damit gilt: $E(X) =: E(a) = a \cdot 1 = a$.

Satz 9.2.9. $E(E(X)) = E(X)$

Beweis. $E(X)$ ist eine Konstante. Laut Satz 9.2.8 gilt damit der Satz. □

Satz 9.2.10. $E(bX) = bE(X)$ ($b \in \mathbb{R}$)

Beweis. bX bedeutet, dass wir die Werte der Zufallsvariable X mit b multiplizieren. Die Wahrscheinlichkeiten, die den Werten von bX zugeordnet sind, sind dabei dieselben, wie die welche den Werten von X zugeordnet sind. Entsprechend sind die Werte der Zufallsvariable bX mit $P(X = x)$ zu gewichten. Auf Grund dieser Überlegung gilt:

$$\begin{aligned} E(bX) &= \sum_{x \in \mathfrak{X}} bXP(X = x) \\ &= b \sum_{x \in \mathfrak{X}} xP(X = x) \\ &= bE(X) \end{aligned}$$

□

Intuitive Erläuterung: Multiplizieren wir die Werte der Zufallsvariable mit einer Konstanten b , so hat die neue Zufallsvariable bX den Erwartungswert $bE(X)$. Die Operation des Multiplizierens der Werte einer Zufallsvariable entspricht z.B. der Währungsumrechnung. Sei der Erwartungswert des Gewinns einer Kapitalanlage in sFr. 50 sFr. Rechnen wir die Anlage in britische Pfund um (Kurs 2.2 sFr = 1 Pfund), so muss der Erwartungswert des Gewinns der Anlage $\frac{1}{2.2} \cdot 50 = 22.727$ Pfund betragen.

Rechnerisches Beispiel. Ein Spiel in Schweizer Franken werde beschrieben durch $\mathfrak{X} = \{-2, 0, 3\}$ mit dem folgenden Graphen der Wahrscheinlichkeitsfunktion

$$f_X = \left\{ \left(-2, \frac{5}{6}\right), \left(0, \frac{1}{12}\right), \left(3, \frac{1}{12}\right) \right\}.$$

Dann gilt $E(X) = -2 \cdot \frac{5}{6} + 0 \cdot \frac{1}{12} + 3 \cdot \frac{1}{12} = -\frac{17}{12} = -1.4167$. Wir nehmen nun an, die Auszahlungen würden in Euro vorgenommen - wobei gelte: 1 Franken = $\frac{2}{3}$ Euro. Dann gilt $Bild(bX) = \{-\frac{4}{3}, 0, 2\}$. Damit gilt: $E(bX) = -\frac{4}{3} \cdot \frac{5}{6} + 0 \cdot \frac{1}{12} + 2 \cdot \frac{1}{12} = -\frac{17}{18} = -0.94444$. Andererseits gilt: $bE(X) = \frac{2}{3}(-\frac{17}{12}) = -\frac{17}{18} = -0.94444$, wodurch der Satz im Beispiel bestätigt wird.

Satz 9.2.11. $E(XE(Y)) = E(Y)E(X)$

Beweis. Da $E(Y)$ eine Konstante ist, gilt der Satz unmittelbar mit Satz 9.2.8 und der Kommutativität der Multiplikation. $\square$

9.3 Varianz und Standardabweichung einer Zufallsvariable

Für Zufallsvariablen X kann bezüglich eines Wahrscheinlichkeitsmasses P die Varianz und die Standardabweichung berechnet werden. Auch hier wird die Begriffsbildung durch das Vorbild der relativen Häufigkeiten geleitet.

Um die folgende Definition plausibel zu machen, betrachten wir eine Variante der empirischen Varianz mit dem Faktor $\frac{1}{n}$ statt $\frac{1}{n-1}$, definiert als:

$$s_v^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Bei klassierten Daten wird daraus:

$$s_v^2 = \frac{1}{n} \sum_{i=1}^m k(x_i)(x_i - \bar{x})^2$$

wobei $k(x_i)$ die absolute Häufigkeit der Objekte mit dem Skalenwert x_i ist und m die Anzahl der Klassen.

Dabei gilt nun:

$$\begin{aligned} s_v^2 &= \frac{1}{n} \sum_{i=1}^m k(x_i)(x_i - \bar{x})^2 \\ &= \sum_{i=1}^m \frac{k(x_i)}{n} (x_i - \bar{x})^2 \\ &= \sum_{i=1}^m h(x_i)(x_i - \bar{x})^2, \end{aligned}$$

wobei $h(x_i)$ die relative Häufigkeit Objekte mit dem Skalenwert x_i ist.

Ersetzen wir den Mittelwert durch den Erwartungswert, x_i durch die Werte der Zufallsvariablen und die relativen Häufigkeiten durch Wahrscheinlichkeiten, so erhalten wir:

Definition 9.3.1. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und X eine Zufallsvariable auf Ω und $\mathfrak{X}$ das Bild von Ω unter X .

(a) Die Varianz der Zufallsvariable X unter der diskreten Wahrscheinlichkeitsverteilung P ist definiert als:

$$V(X) = \sum_{x \in \mathfrak{X}} P(X = x)(x - E(X))^2.$$

(Sofern die Summe $\sum_{x \in \mathfrak{X}} P(X = x)(x - E(X))^2$ definiert ist).

(b) Die Standardabweichung einer Zufallsvariable X unter der diskreten Wahrscheinlichkeitsverteilung P ist definiert als $\sqrt{V(X)}$.

Die Varianz einer Zufallsvariable ist die Summe der - mit der Wahrscheinlichkeit des Eintreffens der Werte der Zufallsvariable gewichteten - quadrierten Abweichungen der Werte der Zufallsvariablen vom Erwartungswert der Zufallsvariable.

Die Varianz und die Standardabweichung einer Zufallsvariable können als „Risiko“ interpretiert werden, dass Werte der Zufallsvariable im Schnitt weit weg vom Erwartungswert zu liegen kommen.

Bei kleiner Standardabweichung ist die Wahrscheinlichkeit grösser, näher beim Erwartungswert zu liegen als bei grosser Standardabweichung. Entsprechend ist z.B. ein Glücksspiel mit grösserer Standardabweichung riskanter als ein solches mit kleinerer Standardabweichung.

Um den Bezug zur Verteilung P anzugeben, werden wir auch $V_P(X)$ verwenden.

Beispiel 9.3.2. (siehe Beispiel 9.2.4):

Sei gegeben der Graph von $f_X = \{(5, \frac{1}{2}), (0, \frac{1}{2})\}$

Die Varianz $V(X)$ der Zufallsvariable X unter P ist dann:

$$V(X) = \sum_{i=1}^2 P(X = x_i) \cdot (x_i - 2.5)^2 = \frac{1}{2}(5 - 2.5)^2 + \frac{1}{2}(0 - 2.5)^2 = 6.25$$

Standardabweichung: $V(X) = \sqrt{6.25} = 2.5$

Betrachten wir hingegen $f_Z = \{(10, \frac{1}{2}), (-5, \frac{1}{2})\}$, so beträgt der Erwartungswert wiederum $E(Z) = 10 \cdot \frac{1}{2} + -5 \cdot \frac{1}{2} = \frac{5}{2} = 2.5$. Das Risiko hat sich aber erhöht: Wenn wir gewinnen, erhalten wir zwar mehr. Verlieren wir aber, haben wir einen grösseren Verlust. Rechnerisch:

$$V(Z) = \sum_{i=1}^2 P(Z = z_i) \cdot (z_i - 2.5)^2 = \frac{1}{2}(10 - 2.5)^2 + \frac{1}{2}(-5 - 2.5)^2 = 56.25. \quad \diamond$$

Beispiel 9.3.3. (siehe Beispiel 9.2.5):

Graph von $f_X = \{(0, 0.6), (5, 0.4)\}$ mit $E(X) = 2$

$$V(X) = 0.6(0 - 2)^2 + 0.4(5 - 2)^2 = 6$$

$$\sqrt{V(X)} = \sqrt{6} = 2.4495. \quad \diamond$$

Beispiel 9.3.4. (siehe Beispiel 9.2.6):

Graph von $f_X = \{(5, 0.25), (1, 0.75)\}$ mit $E(X) = 2$

$$V(X) = 0.25(5 - 2)^2 + 0.75(1 - 2)^2 = 3$$

$$\sqrt{V(X)} = \sqrt{3} = 1.7321. \quad \diamond$$

Beispiel 9.3.5. (siehe Beispiel 9.2.7):

Graph von $f_X = \{(3, 0.1), (1, 0.2), (0, 0.7)\}$ mit $E(X) = 0.5$

$$V(X) = (3 - 0.5)^2 \cdot 0.1 + (1 - 0.5)^2 \cdot 0.2 + (0 - 0.5)^2 \cdot 0.7 = 0.85$$

$$\sqrt{V(X)} = \sqrt{0.85} = 0.92195. \quad \diamond$$

Der nächste Satz führt eine häufig verwendete zusätzliche Formel für die Berechnung der Varianz ein.

Satz 9.3.6. $V(X) = E[(X - E(X))^2]$

Beweis. (i) Gemäss Definition der Varianz gilt:

$$V(X) = \sum_{x \in \mathfrak{X}} (x - E(X))^2 P(X = x)$$

(ii) $(X - E(X))^2$ bedeutet, dass von den Werten der Zufallsvariable X der Erwartungswert der Zufallsvariable X abgezählt wird und diese Differenz quadriert wird. Die Wahrscheinlichkeiten, die den Werten von $(X - E(X))^2$ zugeordnet sind, sind dabei dieselben, wie die welche den Werten von X zugeordnet sind. Entsprechend sind die Werte der Zufallsvariable $(X - E(X))^2$ mit $P(X = x)$ zu gewichten. Auf Grund dieser Überlegung gilt gemäss Definition des Erwartungswertes:

$$E[(X - E(X))^2] = \sum_{x \in \mathfrak{X}} (x - E(X))^2 P(X = x)$$

□

Der Satz ist leicht zu interpretieren: die Varianz ist der Erwartungswert der quadrierten Abweichungen der Werte der Zufallsvariable vom Erwartungswert der Zufallsvariable. Damit ist sie die zu erwartende durchschnittliche quadrierte Abweichung der Werte der Zufallsvariablen von ihrem Erwartungswert.

Satz 9.3.7. $V(X) \geq 0$

Beweis. $V(X) = \sum_{x \in \mathfrak{X}} P(X = x)(x - E(X))^2$. Da $(x - E(X))^2 \geq 0$ für alle x und $E(X)$ sowie $P(X = x) \geq 0$, ist das Produkt $P(X = x)(x - E(X))^2 \geq 0$. Die Summe von Zahlen, die grösser-gleich 0 sind, ist grösser-gleich 0. $\square$

Wie bereits bemerkt, kann die Varianz als ein Risikomass betrachtet werden. Dies wird nun noch an Beispielen von Aktieninvestitionen und weiteren Spielen diskutiert. Wir nehmen an, jemand stehe vor der Alternative, in zwei Aktienpakete zu investieren. Mit einer spezifischen Wahrscheinlichkeit erfolgt ein spezifischer Gewinn (in Prozenten, s. Tabelle 9.3.1). Die Zufallsvariablen nehmen somit die Gewinnprozentzahlen als Werte an.

Investition 1		Investition 2	
Wahrscheinlichkeit	Gewinn in %	Wahrscheinlichkeit	Gewinn in %
0.05	-3	0.05	-20
0.04	-1	0.04	-14.5
0.32	0	0.32	0
0.4	5	0.4	8
0.19	16	0.19	17

Tabelle 9.3.1: Vergleich zweier Investitionen (gleicher Erwartungswert, unterschiedliche Varianz)

Da die Wahrscheinlichkeiten für beide Investitionen gleich gewählt sind, sieht man ohne Rechnung, dass die Varianz der Investition 2 grösser ist als die der Investition 1. Sollten beide denselben Erwartungswert haben, so müsste man bei „rationaler“ Investition die Investition 1 wählen. Wir berechnen die Erwartungswerte und die Varianzen der beiden Investitionen:

$$\begin{aligned}
 E(I_1) &= 0.05 \cdot -3 + 0.04 \cdot -1 + 0.32 \cdot 0 + 0.4 \cdot 5 + 0.19 \cdot 16 = 4.85 \\
 E(I_2) &= 0.05 \cdot -20 + 0.04 \cdot -14.5 + 0.32 \cdot 0 + 0.4 \cdot 8 + 0.19 \cdot 17 = 4.85 \\
 V(I_1) &= 0.05(-3 - 4.85)^2 + 0.04 \cdot (-1 - 4.85)^2 + 0.32 \cdot (0 - 4.85)^2 + \\
 &\quad 0.4 \cdot (5 - 4.85)^2 + 0.19 \cdot (16 - 4.85)^2 = 35.608 \\
 V(I_2) &= 0.05(-20 - 4.85)^2 + 0.04 \cdot (-14.5 - 4.85)^2 + 0.32 \cdot (0 - 4.85)^2 + \\
 &\quad 0.4 \cdot (8 - 4.85)^2 + 0.19 \cdot (17 - 4.85)^2 = 85.398
 \end{aligned}$$

Damit ist die zweite Investition riskanter als die erste. Da beide denselben Erwartungswert aufweisen, ist die erste vorzuziehen.

Im faktischen Wirtschaftsleben ist ein höherer Erwartungswert (höherer durchschnittlicher Gewinn) im Allgemeinen mit einem höheren Risiko verknüpft, d.h. es gilt: wenn $E(I_1) > E(I_2)$, dann $V(I_1) > V(I_2)$. So gilt z.B. für das folgende Beispiel (s. Tabelle 9.3.2):

Investition 1		Investition	
Wahrscheinlichkeit	Gewinn in %	Wahrscheinlichkeit	Gewinn in %
0.06	-4	0.1	-3
0.15	-2	0.3	-2.3
0.2	-1	0.12	1.5
0.12	1	0.44	3.8
0.4	5	0.04	7
0.07	10		

Tabelle 9.3.2: Vergleich zweier Investitionen (ungleicher Erwartungswert, unterschiedliche Varianz)

$$\begin{aligned}
E(I_1) &= 0.06 \cdot -4 + 0.15 \cdot -2 + 0.2 \cdot -1 + 0.12 \cdot 1 + 0.4 \cdot 5 + 0.07 \cdot 10 = 2.08 \\
E(I_2) &= 0.1 \cdot -3 + 0.3 \cdot -2.3 + 0.12 \cdot 1.5 + 0.44 \cdot 3.8 + 0.04 \cdot 7 = 1.142 \\
V(I_1) &= 0.06 \cdot (-4 - 2.08)^2 + 0.15 \cdot (-2 - 2.08)^2 + 0.2 \cdot (-1 - 2.08)^2 + \\
&\quad 0.12 \cdot (1 - 2.08)^2 + 0.4 \cdot (5 - 2.08)^2 + 0.07 \cdot (10 - 2.08)^2 = 14.554 \\
V(I_2) &= 0.1 \cdot (-3 - 1.142)^2 + 0.3 \cdot (-2.3 - 1.142)^2 + 0.12 \cdot (1.5 - 1.142)^2 + \\
&\quad 0.44 \cdot (3.8 - 1.142)^2 + 0.04 \cdot (7 - 1.142)^2 = 9.7664
\end{aligned}$$

Wie soll man sich da verhalten? Hat man so viel Geld, dass man viele riskante Investitionen mit hohem Erwartungswert machen kann, so kann man das riskantere Portfolio wählen, sofern die Entwicklung der Kurse der verschiedenen Aktienpakete unabhängig verläuft. Im Durchschnitt gewinnt man dann mehr als bei vorsichtigerer Investition. Dies zeigt: wer viel Geld hat, wird zu noch mehr Geld kommen, da er seine Anlagen breit streuen kann, deshalb grössere Risiken eingehen kann und damit im Durchschnitt mehr gewinnen wird. Hat man aber wenig Geld und kann z.B. nur eine Investition tätigen, so gibt es keine allgemeine Regel mehr. Die Entscheidung hängt von der Einschätzung des Risikounterschiedes ab. Die Menschen verhalten sich je nach der Einstellung zum Risiko anders. Die meisten Menschen sind bei tiefem Einsatz risikofreudig. So ist z.B. bei der Lotterie die Varianz sehr gross. Zudem ist sogar der Erwartungswert kleiner als 0 und trotzdem spielen viele Menschen. Geht es um grössere Beträge sind die meisten Menschen risikoscheu (= risikoavers). Dies kann das folgende Beispiel klar machen: So entscheiden sich bei einem Spiel, wo man sich zwischen den beiden folgenden Varianten entscheiden muss:

Variante I: man erhält mit Sicherheit 30'000 Sfr.

Variante II: man erhält mit einer Wahrscheinlichkeit von 20% 1 Million Sfr. und mit einer Wahrscheinlichkeit von 80% 0 Sfr., viele für die 30'000 Sfr, obwohl der Erwartungswert der zweiten Alternative 200'000 Sfr. beträgt. (Varianz der Variante I: 0; Varianz der Variante II: $0.8(0 - 200000)^2 + 0.2(1000000 - 200000)^2 = 1.6 \times 10^{11}$. Die sehr hohe Varianz = sehr hohes Risiko und den für die meisten Personen relativ hohen sicheren Gewinn bei der Wahl für Variante I erklärt das Verhalten vieler Spieler in dieser Situation. Für eine Person, die 5 Milliarden besitzt, sieht die Situation vermutlich anders aus. 30'000 ist für sie nicht viel. Sie kann getrost die Variante II wählen).

Es ist durchaus rational, die risikoärmere Variante zu wählen, auch wenn der Erwartungswert tiefer liegt, sofern man das Spiel nur einmal spielen kann. Könnte man das Spiel ein paar Hundert Mal spielen, dann müsste man allerdings die Variante mit dem höheren Erwartungswert wählen - unabhängig von seinem Vermögen und unabhängig von der Varianz.

9.4 Produkte von Bildmassen und Unabhängigkeit von Zufallsvariablen

Betrachten wir n Zufallsexperimente, die alle die selben Ergebnismenge Ω aufweisen, so stellt sich die Frage nach den Wahrscheinlichkeiten der möglichen Ergebnisse solcher Serien oder Folgen von Zufallsexperimenten, wenn auf Ω für jedes Zufallsexperiment eine Zufallsvariable definiert ist. Diese Zufallsvariablen können verschieden oder identisch sein. Im Folgenden beschränken wir uns im Allgemeinen auf den Fall der identischen Zufallsvariablen. Wir führen ihn mit Hilfe eines Beispiels ein.

Wie wahrscheinlich ist es, bei 5-fachem Werfen einer Münze mit den Auszahlungen 1 bei Kopf (K) und 10 bei Zahl (Z) genau die Auszahlungsserie (1, 1, 10, 10, 10) zu erhalten, wobei auf $\Omega = \{K, Z\}$ z.B. eine Gleichverteilung vorausgesetzt sei. Wir haben eine ähnliche Fragestellung bereits beim Produkt von Ergebnismengen behandelt. Dort haben wir die möglichen Ergebnisse beim 5-fache Werfen einer Münze durch die Ergebnismenge Ω^5 ausgedrückt. Offenbar würden wir eine Serie (1, 1, 10, 10, 10) erhalten, wenn sich das Ergebnis $(K, K, Z, Z, Z) \in \Omega^5$ einstellen würde. Von (K, K, Z, Z, Z) zu (1, 1, 10, 10, 10) können wir mit den fünf Zufallsvariablen $X_i = \{(K, 1), (Z, 10)\}$ wie folgt gelangen: $(X_1(K), X_2(K), X_3(Z), X_4(Z), X_5(Z)) = (1, 1, 10, 10, 10)$.

Dabei ist $(1, 1, 10, 10, 10) \in \mathfrak{X}^5 = \{1, 10\}^5$.

Damit bildet $(X_1(\omega_1), X_2(\omega_2), X_3(\omega_3), X_4(\omega_4), X_5(\omega_5))$ ein Element $(\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)$ von Ω^5 nach $\times_{i=1}^5 \mathfrak{X}_i = \mathfrak{X}^5$ ab ($\omega_i \in \Omega$ für $i \in \mathbb{N}_5^*$).

Wir halten dieses Vorgehen allgemein fest, indem wir einen Zufallsvektor X auf Ω^n definieren, wobei ein Zufallsvektor ein n -Tupel von Zufallsvariablen ist:

Definition 9.4.1. (i) Seien n Zufallsvariablen X_i auf n identische Ergebnismengen Ω gegeben und $\mathfrak{X}$ das Bild der X_i . Dann definieren wir einen Zufallsvektor $(X_1, \dots, X_n)$ auf Ω^n wie folgt:

$$\begin{aligned} X &:= (X_1, \dots, X_n) : \Omega^n \rightarrow \mathfrak{X}^n \\ (X_1, \dots, X_n)(\omega) &:= (X_1(\omega_1), \dots, X_n(\omega_n)) \\ \text{mit } \omega &:= (\omega_1, \dots, \omega_n) \in \Omega^n \text{ und } \omega_i \in \Omega \text{ für } i \in \mathbb{N}_n^*. \end{aligned}$$

Als nächstes müssen wir uns nun der Wahrscheinlichkeit des Eintretens von $(1, 1, 10, 10, 10)$ zuwenden. Die Wahrscheinlichkeit des Eintretens von $(1, 1, 10, 10, 10)$ ist im Beispiel offensichtlich mit der Wahrscheinlichkeit des Eintretens von (K, K, Z, Z, Z) identisch. Bei einer Gleichverteilung auf Ω (d.h. Graph der Wahrscheinlichkeitsfunktion f_0 auf $\Omega = \{(K, \frac{1}{2}), (Z, \frac{1}{2})\}$) und ohne gegenseitige Beeinflussung gilt für die Wahrscheinlichkeitsfunktion f_{Ω^5} auf Ω^5 :

$$\begin{aligned} f_{\Omega^5}((K, K, Z, Z, Z)) &= f(K) \cdot f(K) \cdot f(Z) \cdot f(Z) \cdot f(Z) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{32}. \end{aligned}$$

Damit würde für die auf $\times_{i=1}^5 \mathfrak{X}_i$ definierte Wahrscheinlichkeitsfunktion $f_{(X_1, \dots, X_5)}$ gelten:

$$f_{(X_1, \dots, X_5)}((1, 1, 10, 10, 10)) = \frac{1}{32}.$$

Dasselbe Ergebnis erhalten wir auch auf einem anderen Weg. Offenbar gilt

$$f_{X_i} = \{(1, \frac{1}{2}), (10, \frac{1}{2})\}.$$

Da nun Ergebnisse der einzelnen Münzwürfe die Wahrscheinlichkeiten des Eintretens der Ergebnisse der anderen Münzwürfe nicht beeinflussen, können wir die Wahrscheinlichkeit $f_{(X_1, \dots, X_5)}((1, 1, 10, 10, 10))$ offenbar wie folgt berechnen:

$$\begin{aligned} f_{(X_1, \dots, X_5)}((1, 1, 10, 10, 10)) &= f_{X_1}(1) \cdot f_{X_2}(1) \cdot f_{X_3}(10) \cdot f_{X_4}(10) \cdot f_{X_5}(10) \\ &= \left(\frac{1}{2}\right)^5 = \frac{1}{32}. \end{aligned}$$

Dies zeigt, dass man den Begriff des Fehlens gegenseitiger Beeinflussung von Ereignissen aus Ω auf das Fehlen gegenseitiger Beeinflussung von Ereignissen aus $\mathfrak{X}_i$ verallgemeinern kann. Wir liefern die folgenden Definitionen für zwei Zufallsvariablen, wobei analoge Definitionen auch für n Zufallsvariablen geliefert werden könnten.

Definition 9.4.2. $f_{(X_1, X_2)}$ ist die Wahrscheinlichkeitsfunktion einer Gemeinsamen Wahrscheinlichkeitsverteilung von X_1 und X_2 auf $\mathfrak{X}_1 \times \mathfrak{X}_2$ genau dann, wenn

$$\begin{aligned} f_{X_1}(x_i) &= \sum_{x_j \in \mathfrak{X}_2} f_{(X_1, X_2)}(x_i, x_j) \text{ für alle } x_i \in \mathfrak{X}_1 \\ f_{X_2}(x_j) &= \sum_{x_i \in \mathfrak{X}_1} f_{(X_1, X_2)}(x_i, x_j) \text{ für alle } x_j \in \mathfrak{X}_2 \end{aligned}$$

Die durch $f_{(X_1, X_2)}$ definierte Wahrscheinlichkeitsverteilung $P_{(X_1, X_2)}$ wird „gemeinsame Verteilung von X_1 und X_2 “ genannt. Die Abkürzungen $f_{(X_1, X_2)}$ und $P_{(X_1, X_2)}$ werden hier nur für Gemeinsame W-Funktionen und W-Masse verwendet.

Zu beachten ist, dass mit der Angabe des Graphen von X_1 und X_2 jeweils auch die Verteilungen dieser Variablen bestimmt sind. Wir könnten für zwei Zufallsvariablen X_1 und X_2 auf Ω wiederum zeigen, dass für diskrete Wahrscheinlichkeitsräume durch $f_{(X_1, X_2)}((x_i, x_j)) := f_{X_1}(x_i) \cdot f_{X_2}(x_j)$ eine Wahrscheinlichkeitsfunktion auf $\mathfrak{X}_1 \times \mathfrak{X}_2$ definiert wird. Durch diese Wahrscheinlichkeitsfunktion ist eine spezifische Gemeinsame Wahrscheinlichkeitsverteilung auf $\mathcal{P}(\mathfrak{X}_1 \times \mathfrak{X}_2)$ definiert, die wir Produktemass der Zufallsvariablen X_1 und X_2 nennen.

Definition 9.4.3. Seien $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und X_i Zufallsvariablen auf Ω und $\mathfrak{X}_i$ die Bilder von X_i ($i \in \mathbb{N}_2^*$).

Sei $f_{(X_1, X_2)}$ die W-Funktion einer Gemeinsamen Wahrscheinlichkeitsfunktion auf $\mathfrak{X}_1 \times \mathfrak{X}_2$, so dass

$$f_{(X_1, X_2)}((x_i, x_j)) = f_{X_1}(x_i) \cdot f_{X_2}(x_j)$$

(für alle $(x_i, x_j) \in \mathfrak{X}_1 \times \mathfrak{X}_2$). Dann ist $f_{(X_1, X_2)}$ die W-Funktion des Produkte-W-Masses der Zufallsvariablen X_1 und X_2 . Das durch $f_{(X_1, X_2)}$ bestimmte W-Mass $P_{(X_1, X_2)}$ wird in diesem Fall „Produkte-W-Mass von X_1 und X_2 “ genannt (beachten Sie, dass „Verteilung“ und „Mass“ immer gleichbedeutend verwendet wird).

Gilt $f_{(X_1, X_2)}((x_i, x_j)) = f_{X_1}(x_i) \cdot f_{X_2}(x_j)$ für alle $(x_i, x_j) \in \mathfrak{X}_1 \times \mathfrak{X}_2$, so sagen wir, dass die Zufallsvariablen X_1 und X_2 unabhängig sind. Intuitiv besagt die Unabhängigkeit der Zufallsvariablen, dass das Eintreffen eines Objektes aus $\mathfrak{X}_1$ die Wahrscheinlichkeit des Eintreffens eines Objektes aus $\mathfrak{X}_2$ nicht beeinflusst. Für das durch $f_{(X_1, X_2)}$ bestimmte Bildwahrscheinlichkeitsmass $P_X = P_{(X_1, X_2)}$ auf $\mathcal{P}(\mathfrak{X}_1 \times \mathfrak{X}_2)$ können wir dann definieren:

Definition 9.4.4. Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein diskreter Wahrscheinlichkeitsraum und X_i Zufallsvariablen auf Ω und $\mathfrak{X}_i$ die Bilder von X_i ($i \in \mathbb{N}_2^*$). Dann definieren wir:

Die Zufallsvariablen X_1 und X_2 sind unabhängig bezüglich P genau dann, wenn für alle $A \subset \mathfrak{X}_1$ und $B \subset \mathfrak{X}_2$ gilt: $P_{(X_1, X_2)}(A \times B) = P_{X_1}(A) \cdot P_{X_2}(B)$.

($A \times B \subset \mathfrak{X}_1 \times \mathfrak{X}_2$)

Damit sind zwei Zufallsvariablen unabhängig, wenn ihre Gemeinsame Verteilung dem Produktemass der Zufallsvariablen entspricht.

Wie gesagt, stellt das Produktemass eine spezifische *Gemeinsame Verteilung* der zwei Zufallsvariablen X_1 und X_2 dar, die für die Unabhängigkeit von X_1 und X_2 charakteristisch ist. Haben zwei Zufallsvariablen eine Gemeinsame Verteilung, die nicht mit dem Produktemass von X_1 und X_2 zusammenfällt, sind sie nicht unabhängig. Dazu noch ein

Beispiel 9.4.5. Sei $\Omega = \{\text{Kopf}, \text{Zahl}\}$, und P auf Ω gegeben durch den Graphen von $f = \{(\text{Kopf}, 0.4), (\text{Zahl}, 0.6)\}$

Seien der Graph von $X_i = \{(\text{Kopf}, -15), (\text{Zahl}, 6)\}$ ($i \in \mathbb{N}_2^*$).

und damit $f_X = \{(-15, 0.4), (6, 0.6)\}$.

Eine Verteilung auf $\{-15, 6\} \times \{-15, 6\}$ sei durch den folgenden Graphen einer Wahrscheinlichkeitsfunktion gegeben: $\{((-15, -15), 0.25), ((-15, 6), 0.25), ((6, -15), 0.2), ((6, 6), 0.3)\}$.

	-15	6	Summen
-15	0.25	0.25	0.5
6	0.2	0.3	0.5
Summen	0.45	0.55	1

Entsprechend wird dadurch keine Gemeinsame Verteilung der Zufallsvariablen X_1 und X_2 festgelegt (untere Randsumme von -15 ist 0.45, während $f(-15) = 0.4$).

Die folgende W-Funktion legt eine Gemeinsame W-Verteilung von X_1 und X_2 fest, wobei die Verteilungen von X_1 und X_2 gegeben sind durch den Graphen von $f_{X_i} = \{(-15, 0.4), (6, 0.6)\}$.

	-15	6	Summen
-15	0.1	0.3	0.4
6	0.3	0.3	0.6
Summen	0.4	0.6	1

Da $0.4 \cdot 0.4 = 0.16 \neq 0.1$, ist diese Gemeinsame W -Verteilung nicht das Produktmass und die Zufallsvariablen X_1 und X_2 sind nicht unabhängig. Bei Unabhängigkeit müsste nämlich gelten: $\{((-15, -15), 0.16), ((-15, 6), 0.24), ((6, -15), 0.24), ((6, 6), 0.36)\}$ oder anders dargestellt

	-15	6	Summen
-15	0.16	0.24	0.4
6	0.24	0.36	0.6
Summen	0.4	0.6	1

◇

Bemerkung 9.4.6. (1) In der schliessenden Statistik werden wir gewöhnlich nicht zur zwei Zufallsvariablen auf einmal betrachten, sondern n -Tupel $(X_1, \dots, X_n)$ von unabhängigen Zufallsvariablen X_i auf Ω .

(2) Oft werden die Variablen X_i bei Folgen von Zufallsvariablen auf $\times_{i=1}^n \Omega_i$ statt auf die einzelnen Ergebnismengen Ω_i definiert, wobei ihnen die selben Werte zugeordnet werden. Man kann sich überlegen, dass bis anhin sich an den grundlegenden Überlegungen dadurch nichts ändert. In diesem Skript wird die zweite Definition nur bei der (fakultativen) Diskussion des Zusammenhangs der Unabhängigkeit von Ereignissen mit der Unabhängigkeit von Zufallsvariablen verwendet. Statt auf $\times_{i=1}^n \Omega_i$ werden die X_i oft auch auf $\times_{i=1}^n \mathfrak{X}_i$ definiert, wobei ihnen dieselben Werte zugeordnet werden. Die X_i werden dann „Projektionen“ genannt. Eine solche Definition führt etwa zu identischen Berechnungen des Erwartungswertes und der Varianz. Diese letzte Definition hat den Vorteil, dass man die Ergebnismengen von der Mathematik her gesehen weglassen und sich in rein mathematischen Sphären bewegen kann. Diese Definition wird hier bei einigen Beweisen im Unterkapitel „Statistiken“ verwendet. ◇

9.4.1 Unabhängigkeit von Ereignissen und von Zufallsvariablen (fakultativer Stoff)

In vielen statistischen Lehrbüchern wird Unabhängigkeit von Zufallsvariablen mittels der Unabhängigkeit spezifischer Ereignisse definiert. In der Folge werden die entsprechenden Zusammenhänge skizziert:

Definition 9.4.7. Sei $f : D \rightarrow D'$, D' eine Menge von Mengen, die Teilmengen von D' enthält (bei abzählbarem D' wählen wir $\mathcal{D}' = \mathcal{P}(D')$ (= Potenzmenge von D')).

$$f^{-1} : \mathcal{D}' \rightarrow \mathcal{P}(D)$$

$$f^{-1}(A) := \{x \mid f(x) \in A\} \quad (\text{für } A \in \mathcal{D}')$$

Wir nennen f^{-1} die Urabbildung von f (Achtung: f^{-1} ist im Gegensatz zur Umkehrfunktion nicht auf das Bild von f sondern auf eine Menge von Mengen definiert. Verwechslungen sind auf Grund der unterschiedlichen Symbolik für die Argumente nicht zu befürchten. Während die Umkehrfunktion nur für bijektive Funktionen definiert ist, kann die Urabbildung für nicht bijektive f definiert sein).

Satz 9.4.8. Sei $\Omega := \times_{i=1}^n \Omega_i$

$$f_i : \Omega \rightarrow \mathfrak{X}_i$$

$$f : \Omega \rightarrow \times_{i=1}^n \mathfrak{X}_i$$

$$f(x) =: (f_1(x), \dots, f_n(x)),$$

dann gilt für $A = \times_{i=1}^n A_i \subset \times_{i=1}^n \mathfrak{X}_i$:

$$f^{-1}(A) = \bigcap_{i=1}^n f_i^{-1}(A_i)$$

Beweis.

$$\begin{aligned}
 f^{-1}(A) &= f^{-1}(\times_{i=1}^n A_i) \\
 &= \{\omega \mid f(\omega) \in \times_{i=1}^n A_i\} \\
 &= \{\omega \mid (f_1(\omega), \dots, f_n(\omega)) \in \times_{i=1}^n A_i\} \\
 &= \{\omega \mid f_i(\omega) \in A_i \text{ für alle } 1 \leq i \leq n\} \\
 &= \bigcap_{i=1}^n \{\omega \mid f_i(\omega) \in A_i\} \\
 &= \bigcap_{i=1}^n f_i^{-1}(A_i)
 \end{aligned}$$

□

Es folgt noch ein illustratives Beispiel zum Satz:

Beispiel 9.4.9. Sei $\Omega_i = \{1, 2, 3\}$ und $\Omega = \Omega_1 \times \Omega_2$.

$f_1 : \Omega \rightarrow \mathfrak{X}_1$; $f_1 := \{(1, 1, 2), (1, 2, 3), (1, 3, 1), (2, 1, 1), (2, 2, 2), (2, 3, 3), (3, 1, 1), (3, 2, 1), (3, 3, 3)\}$

$f_2 : \Omega \rightarrow \mathfrak{X}_2$; $f_2 := \{(1, 1, 3), (1, 2, 1), (1, 3, 2), (2, 1, 3), (2, 2, 1), (2, 3, 3), (3, 1, 3), (3, 2, 1), (3, 3, 2)\}$

$f : \Omega \rightarrow \mathfrak{X}_1 \times \mathfrak{X}_2$

$f := (f_1, f_2)$ mit $f((x_1, x_2)) = (f_1((x_1, x_2)), f_2((x_1, x_2)))$

$A = \{(1, 2)\} = A_1 \times A_2 = \{1\} \times \{2\} \subset \mathfrak{X}_1 \times \mathfrak{X}_2$.

Dann ist

$f^{-1}(A) = \{(1, 3)\}$

denn $f_1((1, 3)) = 1$ und $f_2((1, 3)) = 2$ und damit $f((1, 3)) = (1, 2)$.

Es gilt $f_1^{-1}(\{1\}) = \{(1, 3), (2, 1), (3, 1), (3, 2)\}$

$f_2^{-1}(\{2\}) = \{(1, 3), (3, 3)\}$

und damit $f^{-1}(A) = f_1^{-1}(A_1) \cap f_2^{-1}(A_2) = \{(1, 3)\}$.

◇

Satz 9.4.10. Seien die Wahrscheinlichkeitsräume $(\Omega, \mathcal{P}(\Omega), P)$ und $(\Omega_i, \mathcal{P}(\Omega_i), P_i)$ mit $\Omega := \times_{i=1}^n \Omega_i$ gegeben, sowie die Zufallsvariablen

$$\begin{aligned}
 X_i &:= \Omega \rightarrow \mathfrak{X}_i, \\
 X &:= \Omega \rightarrow \times_{i=1}^n \mathfrak{X}_i
 \end{aligned}$$

mit $X(\omega) := (X_1(\omega), \dots, X_n(\omega))$, $A := \times_{i=1}^n A_i$, dann gilt:

Die Zufallsvariablen X_i sind unabhängig genau dann, wenn die Ereignisse $X_i^{-1}(A_i)$ für $A_i \subset \mathfrak{X}_i$ unabhängig sind.

Beweis. (1) Seien die Zufallsvariablen unabhängig. Dann gilt definitionsgemäss

$$P_{(X_1, \dots, X_n)}(\times_{i=1}^n A_i) = \prod_{i=1}^n P_{X_i}(A_i) \quad (9.4.1)$$

Es gilt gemäss Definition von P_{X_i}

$$\prod_{i=1}^n P_{X_i}(A_i) = \prod_{i=1}^n P(X_i^{-1}(A_i)). \quad (9.4.2)$$

Zudem gilt gemäss Definition von X und A :

$$\begin{aligned}
 P_{(X_1, \dots, X_n)}(\times_{i=1}^n A_i) &= P_X(A) \\
 &= P(X^{-1}(A)).
 \end{aligned} \quad (9.4.3)$$

Gemäss Satz 9.4.8 gilt:

$$P(X^{-1}(A)) = P\left(\bigcap_{i=1}^n X_i^{-1}(A_i)\right) \quad (9.4.4)$$

Aus den Gleichungsketten (9.4.1 – 9.4.2) und (9.4.3 – 9.4.4) folgt:

$$\prod_{i=1}^n P(X_i^{-1}(A_i)) = P\left(\bigcap_{i=1}^n X_i^{-1}(A_i)\right),$$

d.h. die Unabhängigkeit der Ereignisse $X_i^{-1}(A_i)$.

(2) Sei die Unabhängigkeit der Ereignisse $X_i^{-1}(A_i)$ gegeben. Damit gilt definitionsgemäss:

$$\prod_{i=1}^n P(X_i^{-1}(A_i)) = P\left(\bigcap_{i=1}^n X_i^{-1}(A_i)\right), \quad (9.4.5)$$

Mit der Gleichungskette (9.4.3 – 9.4.4) folgt:

$$P\left(\bigcap_{i=1}^n X_i^{-1}(A_i)\right) = P_{(X_1, \dots, X_n)}(\times_{i=1}^n A_i) \quad (9.4.6)$$

Wegen der Definition von X_i gilt:

$$\prod_{i=1}^n P(X_i^{-1}(A_i)) = \prod_{i=1}^n P_{X_i}(A_i) \quad (9.4.7)$$

Aus (9.4.5), (9.4.6) und (9.4.7) folgt dann

$$P_{(X_1, \dots, X_n)}(\times_{i=1}^n A_i) = \prod_{i=1}^n P_{X_i}(A_i),$$

d.h. die Unabhängigkeit der Zufallsvariablen X_i . □

9.5 Statistiken

Oft interessieren wir uns nur für spezielle Aspekte der möglichen Ausgänge von wiederholten Zufallsexperimenten: Werfen wir zwei Würfel, so können wir uns statt für $\mathfrak{X}_1 \times \mathfrak{X}_2 = \{(x, y) : 1 \leq x, y \leq 6\}$ z.B. für die Summe, das Produkt oder den Mittelwert der Anzahl Augen der Würfel interessieren. Formal sind diese Zufallsvariablen, die wir „Statistiken“ nennen, auf das kartesische Produkt der Bilder von Zufallsvariablen definiert.

Im

Beispiel 9.5.1. betrachten wir die Ergebnismenge

$\Omega = \{1 \text{ Auge}, 2 \text{ Augen}, 3 \text{ Augen}, 4 \text{ Augen}, 5 \text{ Augen}, 6 \text{ Augen}\}$

und die Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$ ($i \in \mathbb{N}_2^*$) mit dem Graphen

$\{(1 \text{ Auge}, 1), (2 \text{ Augen}, 2), (3 \text{ Augen}, 3), (4 \text{ Augen}, 4), (5 \text{ Augen}, 5), (6 \text{ Augen}, 6)\}$.

Das kartesische Produkt der Bilder der Zufallsvariablen X_i ist dann:

$$\mathfrak{X}_1 \times \mathfrak{X}_2 := \{(x, y) : x, y \in \mathbb{N}_6^*\}$$

Auf diese Menge können wir wieder eine Zufallsvariable definieren. Interessieren wir uns für die

Summe der Augenzahlen, so können wir definieren:

$$T: \mathfrak{X}_1 \times \mathfrak{X}_2 \rightarrow \mathbb{R}$$

$T((x, y) = x + y$. Es handelt sich um die Zufallsvariable „Summe“, die jedem Element (x, y) von $\mathfrak{X}_1 \times \mathfrak{X}_2$ die Summe $x + y$ der Komponenten dieser Elemente zuordnet. Das Bild dieser Zufallsvariable ist: $\mathfrak{D}_T := \text{Bild}(T) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Auch hier wird durch die Zufallsvariable das Wahrscheinlichkeitsmass von $\mathfrak{X}_1 \times \mathfrak{X}_2$ nach $\mathfrak{D}_T$ „transportiert“. Wir nehmen an, auf Ω sei eine Gleichverteilung definiert. Im Falle der Unabhängigkeit der Zufallsvariablen X_i gilt damit: $f_{(X_1, X_2)}(x_i, x_j) = f_{X_1}(x_i) \cdot f_{X_2}(x_j) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \quad (i, j \in \mathbb{N}_6^*)$.

Es gilt: $T((1, 1)) = 2$. Sonst gibt es keine geordnete Paare, die eine Summe von 2 aufweisen. Damit gilt dann: $f_T(2) = \frac{1}{36}$.

$$T((1, 2)) = T((2, 1)) = 3. \text{ Damit ist } f_T(3) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$$

$$T((1, 3)) = T((2, 2)) = T((3, 1)) = 4. \text{ Damit ist } f_T(4) = \frac{3}{36}$$

$$T((1, 4)) = T((2, 3)) = T((3, 2)) = T((4, 1)) = 5. \text{ Damit ist } f_T(5) = \frac{4}{36}. \quad \diamond$$

Übung 9.5.2. Vervollständigen Sie diese Berechnung für $f_T(t)$, $t \in \mathfrak{D}_T$. Berechnen Sie den Erwartungswert und die Varianz von T des Beispiels 9.5.1

Beispiel 9.5.3. Gleiche Ausgangslage wie Beispiel 9.5.1. Wir definieren jedoch eine andere Zufallsvariable auf $\mathfrak{X}_1 \times \mathfrak{X}_2$:

$$M: \mathfrak{X}_1 \times \mathfrak{X}_2 \rightarrow \mathbb{R}$$

$$M((x, y) = x \cdot y.$$

Es handelt sich um die Zufallsvariable „Produkt“, die jedem Element (x, y) von $\mathfrak{X}_1 \times \mathfrak{X}_2$ das Produkt $x \cdot y$ der Komponenten dieser Elemente zuordnet. Das Bild dieser Zufallsvariable ist: $\mathfrak{D}_M := \{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$. Als nächstes berechnen wir die Verteilung der Zufallsvariable M , indem wir die Wahrscheinlichkeitsfunktion auf $\mathfrak{D}_M$ berechnen (unter der Voraussetzung, dass auf Ω eine Gleichverteilung definiert ist und die Zufallsvariablen X_i unabhängig sind. Damit ist auch auf $\mathfrak{X}_1 \times \mathfrak{X}_2$ mit $f_{(X_1, X_2)}(x_i, x_j) = \frac{1}{36}$ eine Gleichverteilung definiert).

$$\text{Es gilt: } M((1, 1)) = 1. \text{ Damit gilt: } f_M(1) = \frac{1}{36}.$$

$$\text{Es gilt: } M((1, 2)) = M((2, 1)) = 2. \text{ Damit gilt } f_M(2) = \frac{2}{36}. \quad \diamond$$

Übung 9.5.4. Vervollständigen Sie die Berechnung von f_M . Berechnen Sie den Erwartungswert und die Varianz von M des Beispiels 9.5.3.

Beispiel 9.5.5. Gleiche Ausgangslage wie Beispiel 9.5.1. Wir definieren jedoch eine andere Zufallsvariable auf $\mathfrak{X}_1 \times \mathfrak{X}_2$.

$$\bar{X}: \mathfrak{X}_1 \times \mathfrak{X}_2 \rightarrow \mathbb{R}$$

$\bar{X}((x, y) = \frac{x+y}{2}$. Es handelt sich um die Zufallsvariable „Mittelwert“, die jedem Element (x, y) von $\mathfrak{X}_1 \times \mathfrak{X}_2$ den Mittelwert $\frac{x+y}{2}$ der Komponenten dieser Elemente zuordnet. Das Bild dieser Zufallsvariable ist: $\mathfrak{D}_{\bar{X}} = \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6\}$. Unter der Voraussetzung der Unabhängigkeit der Zufallsvariablen X_i und einer Gleichverteilung auf Ω berechnen wir die Verteilung von $\bar{X}$:

$$\bar{X}((1, 1)) = 1. \text{ Somit ist } f_{\bar{X}}(1) = \frac{1}{36}.$$

$$\bar{X}((1, 2)) = \bar{X}((2, 1)) = 1.5. \text{ Somit ist } f_{\bar{X}}(1.5) = \frac{2}{36}. \quad \diamond$$

Übung 9.5.6. Vervollständigen Sie die Berechnung von $f_{\bar{X}}$ des Beispiels 9.5.5. Berechnen Sie den Erwartungswert und die Varianz von $\bar{X}$ sowie von X_i . Vergleichen Sie!

Beispiel 9.5.7. $\Omega = \{a, b\}$ mit Gleichverteilung. $X_i(a) = 1; X_i(b) = 2$ mit $i \in \{1, 2, 3, 4\}$ (wir haben also vier identische Zufallsvariablen); X_i seien unabhängig. $\mathfrak{X}_i = \{1, 2\}$. $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 \times \mathfrak{X}_4 = \{1, 2\} \times \{1, 2\} \times \{1, 2\} \times \{1, 2\} =$

$$\{(1, 1, 1, 1), (1, 1, 1, 2), (1, 1, 2, 1), (1, 2, 1, 1), (2, 1, 1, 1),$$

$$(1, 1, 2, 2), (1, 2, 1, 2), (2, 1, 1, 2), (1, 2, 2, 1), (2, 1, 2, 1), (2, 2, 1, 1),$$

$$(1, 2, 2, 2), (2, 1, 2, 2), (2, 2, 1, 2), (2, 2, 2, 1), (2, 2, 2, 2)\}$$

Die gemeinsame Verteilung ist wegen der Unabhängigkeit der Zufallsvariablen X_i und deren Bijektivität sowie der Gleichverteilung auf Ω gegeben durch: $f_{(X_1, X_2, X_3, X_4)}((x, y, z, r)) = f_{X_1}(x) \cdot f_{X_2}(y) \cdot f_{X_3}(z) \cdot f_{X_4}(r) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}$.

Wir definieren nun eine Zufallsvariable „Mittelwert“ auf $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 \times \mathfrak{X}_4$:

$$\bar{X} : \mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 \times \mathfrak{X}_4 \rightarrow \mathbb{R}$$

$$\bar{X}((x, y, z, r)) := \frac{x+y+z+r}{4}.$$

Die Zufallsvariable ordnet den den 4-Tupel des Definitionsbereichs jeweils das arithmetische Mittel ihrer Komponenten zu: $\mathfrak{D}_{\bar{X}} = \{1, 1.25, 1.5, 1.75, 2\}$

Die Wahrscheinlichkeitsverteilung der Zufallsvariable „Mittelwert“ können wir Hilfe des Graphen der Wahrscheinlichkeitsfunktion $f_{\bar{X}}(t)$ wie folgt ausdrücken:

$$f_{\bar{X}}(t) = \{(1, \frac{1}{16}), (1.25, \frac{4}{16}), (1.5, \frac{6}{16}), (1.75, \frac{4}{16}), (2, \frac{1}{16})\}.$$

◇

Übung 9.5.8. Kontrollieren Sie dieses Ergebnis. Berechnen Sie den Erwartungswert und die Varianz von $\bar{X}$ und von X_i des Beispiels 9.5.7 Vergleichen Sie!

Schematisch kann man das obige Vorgehen wie folgt darstellen (s. Tabelle 9.5.3):

Ω^n	$\xrightarrow{(X_1, \dots, X_n)} \times_{i=1}^n \mathfrak{X}_i$	$\xrightarrow{T} \mathfrak{D}_T$	Zufallsvariablen mit Definitionsbereichs- oder Bildbereichen
P_{Ω^n}	$P_{(X_1, \dots, X_n)}$	P_T	Wahrscheinlichkeitsmasse auf die jeweiligen Mengen

Erläuterungen:

- Ω = Ergebnismenge der einzelnen n Zufallsexperimente
- $(X_1, \dots, X_n) = n$ -Tupel von Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$
- $\times_{i=1}^n \mathfrak{X}_i = \text{Bild}(X_1, \dots, X_n)$
- $T =$ Zufallsvariable auf $\times_{i=1}^n \mathfrak{X}_i$ (= Statistik auf $\times_{i=1}^n \mathfrak{X}_i$)
- $P_{(X_1, \dots, X_n)} =$ gemeinsame Verteilung der Zufallsvariablen X_i auf $\mathcal{P}\left(\times_{i=1}^n \mathfrak{X}_i\right)$
(= Produktmass bei unabhängigen X_i ; $\mathcal{P}(A)$ = Potenzmenge der Menge A)
- $P_T =$ Bildmass von $P_{(X_1, \dots, X_n)}$ unter T
- $\mathfrak{D}_T = \text{Bild}(T)$

Tabelle 9.5.3: Übersicht über Statistiken

Zufallsvariablen auf das kartesische Produkt der Bilder von Zufallsvariablen spielen in der schliessenden Statistik eine wesentliche Rolle und die obigen Beispiele sollten genau verstanden sein. Deshalb folgt nun noch eine formelle Definition:

Definition 9.5.9. T ist eine Statistik genau dann, wenn T eine Zufallsvariable auf das kartesische Produkt der Bilder von Zufallsvariablen ist.

Wir haben an den obigen Beispielen gesehen, dass $E(\bar{X}) = E(X_i)$ für identisch verteilte Zufallsvariablen und $V(\bar{X}) = \frac{1}{n}V(X_i)$ für unabhängige und identisch verteilte Zufallsvariablen gilt. Diese wesentlichen Ergebnisse wollen wir nun noch herleiten.

9.5.1 Erwartungswert der Statistik Mittelwert

Definition 9.5.10. Seien X_i Zufallsvariablen. $\mathfrak{X}_i$ die Bilder dieser Zufallsvariablen. Dann schreiben wir künftig für

$$T: \times_{i=1}^n \mathfrak{X}_i \rightarrow \mathbb{R}$$

$$T((x_1, \dots, x_n)) = \sum_{i=1}^n x_i :$$

$$\sum_{i=1}^n X_i := T.$$

Mit $x := (x_1, \dots, x_n)$ schreiben wir damit auch: $\sum_{i=1}^n X_i(x) := \sum_{i=1}^n X_i((x_1, \dots, x_n)) = \sum_{i=1}^n x_i$.

Der nächste Satz besagt, dass der Erwartungswert der Summe von n Zufallsvariablen X_i unter der Gemeinsamen Verteilung der X_i identisch ist mit der Summe der Erwartungswerte der Zufallsvariablen X_i unter P , sofern die einzelnen Erwartungswerte reelle Zahlen sind. Für eine intuitive Erläuterung siehe nach dem Beweis.

Satz 9.5.11. *Seien X_i Zufallsvariablen auf Ω mit $-\infty < E_P(X_i) < \infty$ und $P_{(X_1, \dots, X_n)}$ eine Gemeinsame Verteilung der Zufallsvariablen X_i . Dann gilt*

$$E_{P_{(X_1, \dots, X_n)}} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n E_P(X_i).$$

Beweis. Wir beweisen den Satz für zwei Zufallsvariablen X_1, X_2 auf Ω_1 und Ω_2 , wobei $\mathfrak{X}_i$ das Bild von X_i ist. Mit $T := X_1 + X_2$ gilt gemäss Satz 9.2.3 (S. 203)

$$\begin{aligned} E_{P_{(X_1, X_2)}}(T) &= \sum_{t \in \text{Bild}(T)} t \cdot P_{(X_1, X_2)}(T = t) \\ &= \sum_{(x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2} T((x_1, x_2)) \cdot P_{(X_1, X_2)}(\{(x_1, x_2)\}) \\ &= \sum_{(x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2} (x_1 + x_2) \cdot P_{(X_1, X_2)}(\{(x_1, x_2)\}) \\ &= \sum_{x_1 \in \mathfrak{X}_1} \sum_{x_2 \in \mathfrak{X}_2} (x_1 + x_2) \cdot P_{(X_1, X_2)}(\{(x_1, x_2)\}) \\ &= \sum_{x_1 \in \mathfrak{X}_1} \sum_{x_2 \in \mathfrak{X}_2} [x_1 P_{(X_1, X_2)}(\{(x_1, x_2)\}) + x_2 P_{(X_1, X_2)}(\{(x_1, x_2)\})] \\ &= \sum_{x_1 \in \mathfrak{X}_1} \sum_{x_2 \in \mathfrak{X}_2} x_1 P_{(X_1, X_2)}(\{(x_1, x_2)\}) + \sum_{x_2 \in \mathfrak{X}_2} \sum_{x_1 \in \mathfrak{X}_1} x_2 P_{(X_1, X_2)}(\{(x_1, x_2)\}) \\ &= \sum_{x_1 \in \mathfrak{X}_1} x_1 \sum_{x_2 \in \mathfrak{X}_2} P_{(X_1, X_2)}(\{(x_1, x_2)\}) + \sum_{x_2 \in \mathfrak{X}_2} x_2 \sum_{x_1 \in \mathfrak{X}_1} P_{(X_1, X_2)}(\{(x_1, x_2)\}) \end{aligned}$$

Da $P_{(X_1, X_2)}$ eine gemeinsame Verteilung von X_1 und X_2 ist, gilt

$$\begin{aligned} \sum_{x_2 \in \mathfrak{X}_2} P_{(X_1, X_2)}(\{(x_1, x_2)\}) &= P_{X_1}(\{x_1\}) \\ \sum_{x_1 \in \mathfrak{X}_1} P_{(X_1, X_2)}(\{(x_1, x_2)\}) &= P_{X_2}(\{x_2\}) \end{aligned}$$

Damit gilt

$$\begin{aligned} E_{P_{(X_1, X_2)}}(T) &= \sum_{x_1 \in \mathfrak{X}_1} x_1 P_{X_1}(\{x_1\}) + \sum_{x_2 \in \mathfrak{X}_2} x_2 P_{X_2}(\{x_2\}) \\ &= E(X_1) + E(X_2) \end{aligned}$$

Die Verallgemeinerung auf n Zufallsvariablen ergibt sich zwanglos. □

Intuitive Erläuterung für $n = 2$: Wir nehmen an, wir könnten zwei Spiele spielen. Dann besagt der Satz: Der Erwartungswert des Gesamtertrags beim Spielen dieser zwei Spiele ist die Summe der Erwartungswerte der einzelnen Spiele. Rechnerische Beispiele haben wir oben gesehen (s. z.B. Beispiel 9.5.1 auf S. 214).

Bemerkung 9.5.12. Am obigen Beweis sieht man, dass die Unabhängigkeit von X_1 und X_2 für die Gültigkeit des Satzes nicht erforderlich ist. Allerdings muss $P_{(X_1, X_2)}$ eine Gemeinsame Verteilung von X_1 und X_2 sein, damit der Satz gilt. $\diamond$

Satz 9.5.13. $E(a + bX) = a + bE(X)$ ($a, b \in \mathbb{R}, b \neq 0$)

Beweis. $E(a + bX) \stackrel{\text{Satz 9.5.11}}{=} E(a) + E(bX) \stackrel{\text{Sätze 9.2.8; 9.2.10}}{=} a + bE(X).$ $\square$

Satz 9.2.10 ist umgekehrt ein Spezialfall von Satz 9.5.13 mit $a = 0$.

Definition 9.5.14. Die Statistik Mittelwert $\bar{X}_n$ von n Zufallsvariablen X_i wird definiert als:

$$\bar{X}_n : \times_{i=1}^n \mathfrak{X}_i \rightarrow \mathbb{R}$$

$$\bar{X}_n(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{für } x = (x_1, \dots, x_n))$$

Satz 9.5.15. $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

Beweis. $\sum_{i=1}^n X_i(x) \stackrel{\text{Definition 9.5.10}}{=} \sum_{i=1}^n x_i$. Wir multiplizieren beide Seiten mit $\frac{1}{n}$ und erhalten: $\frac{1}{n} \sum_{i=1}^n X_i(x) = \frac{1}{n} \sum_{i=1}^n x_i$. Da $\frac{1}{n} \sum_{i=1}^n x_i \stackrel{\text{Definition 9.5.14}}{=} \bar{X}_n(x)$, gilt $\bar{X}_n(x) = \frac{1}{n} \sum_{i=1}^n X_i(x)$. $\square$

Definition 9.5.16. Für identisch verteilte und unabhängige Zufallsvariablen verwenden wir die Abkürzung „i.i.d.“ (für „identically and independently distributed“), für identische verteilte Zufallsvariablen die Abkürzung „i.d.“

Satz 9.5.17. Seien X_i i.d. und $P_{(X_1, \dots, X_n)}$ eine gemeinsame Verteilung der Zufallsvariablen X_i .

Dann gilt: $E_{P_{(X_1, \dots, X_n)}}\left(\sum_{i=1}^n X_i\right) = nE_P(X_i)$

Beweis. $E_{P_{(X_1, \dots, X_n)}}\left(\sum_{i=1}^n X_i\right) \stackrel{\text{Satz 9.5.11}}{=} \sum_{i=1}^n E_P(X_i)$. Da die X_i identisch verteilt sind, gilt: $E_P(X_i) = E_P(X_j)$ für alle $i, j \in \mathbb{N}^*$. Somit gilt: $\sum_{i=1}^n E_P(X_i) = nE_P(X_i)$. $\square$

Damit können wir schon $E(X_i) = E(\bar{X}_n)$ beweisen:

Satz 9.5.18. Bei n i.d. X_i und einer gemeinsamen Verteilung $P_{(X_1, \dots, X_n)}$ der Zufallsvariablen X_i gilt: $E_{P_{(X_1, \dots, X_n)}}(\bar{X}_n) = E_P(X_i)$

Beweis. $E_{P_{(X_1, \dots, X_n)}}(\bar{X}_n) \stackrel{\text{Definition } \bar{X}_n}{=} E_{P_{(X_1, \dots, X_n)}}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{\text{Satz 9.2.10}}{=} \frac{1}{n} E_{P_{(X_1, \dots, X_n)}}\left(\sum_{i=1}^n X_i\right) \stackrel{\text{Satz 9.5.17}}{=} \frac{1}{n} n E_P(X_i) = E_P(X_i).$ $\square$

9.5.2 Varianz der Statistik Mittelwert

Wir gehen nun zur Herleitung von

$$V(\bar{X}_n) = \frac{1}{n} V(X_i)$$

über. Dabei erwähnen und beweisen wir etliche Sätze, die auch sonst von Bedeutung sind. Zuerst eine weitere, gängige Berechnungsweise für die Varianz.

Satz 9.5.19. $V(X) = E(X^2) - [E(X)]^2$

Beweis. $V(X) = E[(X - E(X))^2] = E[X^2 - 2XE(X) + [E(X)]^2] \stackrel{\text{Satz 9.5.11}}{=} E(X^2) - E(2XE(X)) + E[E(X)]^2 = E(X^2) - 2E(X)E(X) + [E(X)]^2$
(der zweite Ausdruck durch Anwendung der Sätze 9.2.10 und 9.2.11. $[E(X)]^2$ ist eine Konstante. Somit ist $E([E(X)]^2) = E(X)^2$). Somit gilt:
 $E(X^2) - 2E(X)E(X) + [E(X)]^2 = E(X^2) - 2[E(X)]^2 + [E(X)]^2 = E(X^2) - [E(X)]^2$. $\square$

Satz 9.5.20. $V(a) = 0 \quad (a \in \mathbb{R})$

Beweis. $V(a) = E(a^2) - [E(a)]^2 = a^2 - a^2 = 0$. $\square$

Intuitive Bedeutung: Wenn es bei einem Spiel oder einer Kapitalanlage nur einen möglichen Ausgang oder Gewinn gibt, dann ist die Varianz des Spiels oder des Gewinns gleich 0.

Rechnerisches Beispiel: $\Omega = \{1, 3, 4\}$. $X(\omega) = a$ für alle $\omega \in \Omega$. Graph von $f = \{(a, 1)\}$. Damit gilt, da $E(a) = a$:

$$V(a) = (a - a)^2 \cdot 1 = 0 \cdot 1 = 0.$$

Die folgende Gleichung sollten Sie sich einprägen, da sie oft verwendet wird.

Satz 9.5.21. $V(bX) = b^2V(X) \quad (b \in \mathbb{R})$

Beweis. $V(bX) = E((bX)^2) - [E(bX)]^2 = E(b^2X^2) - [bE(X)]^2 = b^2E(X^2) - b^2[E(X)]^2 = b^2(E(X^2) - [E(X)]^2) = b^2V(X)$. $\square$

Intuitive Bedeutung: Wenn wir die Werte einer Zufallsvariable für Kapitalgewinne, deren Werte in Franken gemessen wurden, mit einer Konstanten multiplizieren (wie bei Währungsumrechnungen), so ist die Varianz der Zufallsvariable in der neuen Währung das Produkt des Quadrats der Konstanten und der Varianz der Zufallsvariablen.

Rechnerisches Beispiel: Ein Spiel in Schweizer Franken werde beschrieben durch $\mathfrak{X} = \{-2, 0, 3\}$ mit dem folgenden Graphen der Bildwahrscheinlichkeitsfunktion $f_X = \{(-2, \frac{5}{6}), (0, \frac{1}{12}), (3, \frac{1}{12})\}$.

Dann gilt, wie oben (Satz 9.2.10) gezeigt $E(X) = -\frac{17}{12}$ und
 $V(X) = (-2 + \frac{17}{12})^2 \frac{5}{6} + (0 + \frac{17}{12})^2 \frac{1}{12} + (3 + \frac{17}{12})^2 \frac{1}{12} = \frac{299}{144} = 2.0764$.

Wir nehmen nun an, die Auszahlungen würden in Euro vorgenommen - wobei gelte: 1 Franken = $\frac{2}{3}$ Euro. Dann gilt $\mathfrak{X}_b = \{-\frac{4}{3}, 0, 2\}$ und wie oben gezeigt $E(bX) = -\frac{17}{18}$. Damit gilt: $V(bX) = (-\frac{4}{3} + \frac{17}{18})^2 \frac{5}{6} + (0 + \frac{17}{18})^2 \frac{1}{12} + (2 + \frac{17}{18})^2 \frac{1}{12} = \frac{299}{324} = 0.92284$

Andererseits gilt: $b^2V(X) = (\frac{2}{3})^2 \frac{299}{144} = \frac{299}{324}$, wodurch der Satz im Beispiel bestätigt wird.

Satz 9.5.22. $V(a + bX) = b^2V(X) \quad (b, a \in \mathbb{R}, b \neq 0)$

Beweis. $V(a + bX) = E(E((a + bX)^2) - [E(a + bX)]^2)$

$$(i) \quad E((a + bX)^2) = E(a^2 + 2abX + b^2X^2) = a^2 + 2abE(X) + b^2E(X^2)$$

$$(ii) \quad [E(a + bX)]^2 = (a + bE(X))^2 = a^2 + 2abE(X) + b^2[E(X)]^2$$

Somit gilt: $E((a + bX)^2) - [E(a + bX)]^2 = b^2E(X^2) - b^2[E(X)]^2 =$

$$b^2(E(X^2) - [E(X)]^2) = b^2V(X). \quad \square$$

Daraus folgt, dass die Varianz von der konstanten Verschiebung der Werte der Zufallsvariable unabhängig ist.

In der Folge verwenden wir noch die Statistik „Produkt“:

Definition 9.5.23.

$$\prod_{i=1}^n X_i : \times_{i=1}^n \mathfrak{X}_i \rightarrow \mathbb{R}$$

$$\prod_{i=1}^n X_i(x) := \prod_{i=1}^n x_i$$

(mit $x = (x_1, \dots, x_n)$)

Wie die Statistik „Summe“ ist die Statistik „Produkt“ auf das kartesische Produkt der Bilder von X_i definiert (siehe Beispiel 9.5.3).

Bemerkung 9.5.24. Diese Bemerkung und die Sätze 9.5.25 und 9.5.26 muss man nur lesen, wenn man die folgenden Beweise dieses Unterkapitels durcharbeiten will. In der Folge werden wir auf den Ausdruck $E_{P_{(X_1, X_2)}}(X_1)$ stossen. Dabei stellt sich das Problem, dass X_1 auf Ω definiert ist und nicht auf $\mathfrak{X}_i \times \mathfrak{X}_i$. Definieren wir nun

$$\begin{aligned} X_i : \mathfrak{X}_i \times \mathfrak{X}_i &\rightarrow \mathbb{R} \\ X_i(x) &= x_i \end{aligned}$$

(d.h. dem geordneten Paar (x_1, x_2) wird von X_1 die Zahl x_1 zugeordnet und (x_1, x_2) wird von X_2 die Zahl x_2 zugeordnet), so weisen die so definierten X_i dieselben Bilder auf wie die auf Ω definierten Zufallsvariablen X_i und die alternativ definierten Zufallsvariablen X_i haben dieselben Werte der Verteilung wie die auf Ω definierten Zufallsvariablen. Entsprechend haben die alternativ definierten Zufallsvariablen X_i dieselbe Varianz und denselben Erwartungswert wie die auf Ω definierten Zufallsvariablen. Mit dieser Abmachung gilt nun:

Satz 9.5.25. $E_{P_{(X_1, X_2)}}(X_i) = E_P(X_i) \quad (i \in \{1, 2\})$

Beweis. $X_i = X_i + 0$. Damit gilt dann: $E_{P_{(X_1, X_2)}}(X_i) = E_{P_{(X_1, X_2)}}(X_i + 0) = E_P(X_i) + E_P(0) = E_P(X_i)$. $\square$

Satz 9.5.26. $E_{P_{(X_1, X_2)}}(X_i^2) = E_P(X_i^2)$

Beweis. $E_{P_{(X_1, X_2)}}(X_i^2 + 0) = E_P(X_i^2) + E_P(0) = E_P(X_i^2)$. $\square$
 $\diamond$

Als nächstens führen wir die Kovarianz ein (s. die Definition 5.3.8 der empirischen Kovarianz, S. 109). Es ist eine Zufallsvariable, welche als Werte eine Zahl annimmt, die den Zusammenhang zweier Zufallsvariablen misst. Die Kovarianz hat einen Zusammenhang mit der Unabhängigkeit von Zufallsvariablen. In der Tat ist die Kovarianz 0, wenn die Zufallsvariablen unabhängig sind - wobei die Kovarianz 0 sein kann, ohne dass die Zufallsvariablen unabhängig sind. Diese Zusammenhänge werden nun genauer ausgeführt:

Definition 9.5.27. (Kovarianz):

$$KOV(X_1, X_2) := E_{P_{(X_1, X_2)}}((X_1 - E_P(X_1)) \cdot (X_2 - E_P(X_2))).$$

Beispiel 9.5.28. Im diskreten Fall gilt

$$KOV(X_1, X_2) = \sum_{i=1}^n (x_{1i} - E_P(X_1))(x_{2i} - E_P(X_2))P_{(X_1, X_2)}(\{(x_{1i}, x_{2i})\}).$$

wobei x_{1i} sowie x_{2i} die möglichen Werte der Zufallsvariablen X_1 und X_2 sind und $P_{(X_1, X_2)}$ eine Gemeinsame Verteilung dieser Zufallsvariablen ist. Spiele sind im Allgemeinen unabhängig. Wir wollen untersuchen, ob bei folgenden Spielen die Kovarianz 0 ist.

Seien zwei Spiele durch die Graphen von Wahrscheinlichkeitsfunktionen gegeben:

$$f_{X_1} = \{(2, 0.3), (3, 0.5), (4, 0.2)\}$$

$$f_{X_2} = \{(5, 0.1), (7, 0.25), (1, 0.35), (10, 0.3)\}$$

und eine Gemeinsame Verteilung durch den Graphen einer Wahrscheinlichkeitsfunktion mit Hilfe der ersten drei Spalten der folgenden Tabelle 9.5.4:

$\mathfrak{X}_1 \times \mathfrak{X}_2$	$g_{(X_1, X_2)}$	f_{X_1}	f_{X_2}	$f_{(X_1, X_2)}$	$KOV_g(X_1, X_2)$	$KOV_f(X_1, X_2)$
2 5	0.05	0.3	0.1	0.03	0.027	0.0162
3 5	0.04	0.5	0.1	0.05	-0.0024	-0.003
4 5	0.01	0.2	0.1	0.02	-0.0066	-0.0132
2 7	0.1	0.3	0.25	0.075	-0.126	-0.0945
3 7	0.14	0.5	0.25	0.125	0.0196	0.0175
4 7	0.01	0.2	0.25	0.05	0.0154	0.077
2 1	0.1	0.3	0.35	0.105	0.414	0.4347
3 1	0.14	0.5	0.35	0.175	-0.0644	-0.0805
4 1	0.11	0.2	0.35	0.07	-0.5566	-0.3542
2 10	0.05	0.3	0.3	0.09	-0.198	-0.3564
3 10	0.18	0.5	0.3	0.15	0.0792	0.066
4 10	0.07	0.2	0.3	0.06	0.3388	0.2904
Summen	1			1	-0.06	0

Tabelle 9.5.4: Tabelle zur Berechnung der Kovarianz des Beispiels. Bei $g_{(X_1, X_2)}$ stehen die Werte einer Gemeinsamen Wahrscheinlichkeitsfunktion, die nicht dem Produktmass entspricht. Bei $f_{(X_1, X_2)}$ stehen die Werte der Wahrscheinlichkeitsfunktion des Produktmasses. $Kov_g(X_1, X_2)$ entspricht der Kovarianz bei $g_{(X_1, X_2)}$ und $Kov_f(X_1, X_2)$ der Kovarianz bei Unabhängigkeit (Produktmass)

Damit gilt:

$$E(X_1) = 2 \cdot 0.3 + 3 \cdot 0.5 + 4 \cdot 0.2 = 2.9$$

$$E(X_2) = 5 \cdot 0.1 + 7 \cdot 0.25 + 1 \cdot 0.35 + 10 \cdot 0.3 = 5.6$$

$$\begin{aligned} & \sum_{i=1}^n (x_{1i} - 2.9)(x_{2i} - 5.6)P_{(X_1, X_2)}(\{(x_{1i}, x_{2i})\}) \\ &= (2 - 2.9)(5 - 5.6)0.05 + \dots + (4 - 2.9)(10 - 5.6)0.07 \\ &= -0.06 \end{aligned}$$

(s. Spalte $KOV_g(X_1, X_2)$ in der Tabelle 9.5.4). Damit besteht ein leicht negativer Zusammenhang zwischen den Variablen: hohe Werte von X_1 treten mit tiefen Werten von X_2 ein - mit den entsprechenden Wahrscheinlichkeiten gewichtet (allerdings sehr schwache Tendenz!). Um den Zusammenhang zu interpretieren, ist allerdings die Korrelation besser (s. Bemerkung 9.5.29). Bei unabhängigen Zufallsvariablen erhalten wir eine Kovarianz von 0 (s. Spalten $f_{X_1}, f_{X_2}, f_{(X_1, X_2)}$ und $KOV_f(X_1, X_2)$). $\diamond$

Bemerkung 9.5.29. Man kann auch eine Korrelation für Zufallsvariablen definieren und zwar durch

$$KOR(X_1, X_2) = \frac{KOV(X_1, X_2)}{\sqrt{V(X_1)}\sqrt{V(X_2)}}$$

Im obigen Beispiel (s. Tabelle 9.5.4) würden erhalten:

$$V(X_1) = (2 - 2.9)^2 0.3 + (3 - 2.9)^2 0.5 + (4 - 2.9)^2 0.2 = 0.49$$

$$V(X_2) = (5 - 5.6)^2 0.1 + (7 - 5.6)^2 0.25 + (1 - 5.6)^2 0.35 + (10 - 5.6)^2 0.3 = 13.74$$

Und damit für die gemeinsame Verteilung des Beispiels der Tabelle 9.5.4 $g_{(X_1, X_2)}$:

$$KOR_g(X_1, X_2) = \frac{-0.06}{\sqrt{0.49}\sqrt{13.74}} = -0.023124$$

Die negative Korrelation der Zufallsvariablen ist damit sehr schwach. Bei unabhängigen Zufallsvariablen ist die Korrelation immer 0, da die Kovarianz 0 ist. $\diamond$

Satz 9.5.30. $KOV(X_1, X_2) = E_{P_{(X_1, X_2)}}(X_1 \cdot X_2) - E_P(X_1)E_P(X_2)$

Beweis. $KOV(X_1, X_2) = E_{P_{(X_1, X_2)}}((X_1 - E_P(X_1)) \cdot (X_2 - E_P(X_2))) =$
 $E_{P_{(X_1, X_2)}}(X_1 X_2 - X_1 E_P(X_2) - E_P(X_1) X_2 + E_P(X_1) E_P(X_2)) =$
 $E_{P_{(X_1, X_2)}}(X_1 X_2) - E_{P_{(X_1, X_2)}}(X_1 E_P(X_2)) - E_{P_{(X_1, X_2)}}(E_P(X_1) X_2) + E_{P_{(X_1, X_2)}}(E_P(X_1) E_P(X_2)) =$
 $E_{P_{(X_1, X_2)}}(X_1 X_2) - E_P(X_1) E_P(X_2) - E_P(X_1) E_P(X_2) + E_P(X_1) E_P(X_2) =$
 $E_{P_{(X_1, X_2)}}(X_1 X_2) - E_P(X_1) E_P(X_2).$ $\square$

Satz 9.5.31. $KOV(X_1, X_1) = V(X_1)$

Beweis. Folgt aus der Definition der Kovarianz durch Einsetzen sowie durch die Sätze 9.3.6 und 9.5.26. $\square$

Satz 9.5.32. $KOV(X_1, X_2) = 0$ genau dann, wenn

$$E_{P_{(X_1, X_2)}}(X_1 X_2) = E_P(X_1) E_P(X_2)$$

Beweis. folgt unmittelbar aus Satz 9.5.30. $\square$

Satz 9.5.33. Wenn X_1 und X_2 unabhängig sind und $-\infty < E_P(X_i), E_P(X_i) < \infty$, dann gilt:

$$E_{P_{(X_1, X_2)}}(X_1 X_2) = E_P(X_1) E_P(X_2)$$

Beweis.

$$E_{P_{(X_1, X_2)}}(X_1 X_2) = \sum_{t \in \text{Bild}(X_1 X_2)} t P(X_1 X_2 = t),$$

wobei $X_1 X_2$ eine Zufallsvariable auf $\mathfrak{X}_1 \times \mathfrak{X}_2$ ist. Damit gilt:

$$E_{P_{(X_1, X_2)}}(X_1 X_2) = \sum_{(x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2} t P_{(X_1, X_2)}(\{(x_1, x_2)\})$$

wobei $t = x_1 x_2$ und wegen der Unabhängigkeit gilt dann

$$P_{(X_1, X_2)}(\{(x_1, x_2)\}) = P_{X_1}(x_1) P_{X_2}(x_2) \\ = P(X_1 = x_1) P(X_2 = x_2).$$

Wir erhalten somit:

$$\begin{aligned} E_{P_{(X_1, X_2)}}(X_1 X_2) &= \sum_{(x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2} (x_1 x_2 P(X_1 = x_1) P(X_2 = x_2)) \\ &= \sum_{(x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2} (x_1 P(X_1 = x_1) x_2 P(X_2 = x_2)) \\ &= \sum_{x_1 \in \mathfrak{X}_1} \sum_{x_2 \in \mathfrak{X}_2} (x_1 P(X_1 = x_1) x_2 P(X_2 = x_2)) \\ &= \sum_{x_1 \in \mathfrak{X}_1} x_1 P(X_1 = x_1) \sum_{x_2 \in \mathfrak{X}_2} x_2 P(X_2 = x_2) \\ &= E_P(X_1) E_P(X_2). \end{aligned}$$

$\square$

Satz 9.5.34. *Es gilt:*

$$V_{P_{(X_1, X_2)}}(X_1 + X_2) = V_P(X_1) + V_P(X_2) + 2KOV(X_1, X_2)$$

Beweis. $V_{P_{(X_1, X_2)}}(X_1 + X_2) = E_{P_{(X_1, X_2)}}(X_1 + X_2)^2 - \left[E_{P_{(X_1, X_2)}}(X_1 + X_2) \right]^2 =$
 $E_{P_{(X_1, X_2)}}(X_1^2 + 2X_1X_2 + X_2^2) - [E_P(X_1) + E_P(X_2)]^2 =$
 $E_{P_{(X_1, X_2)}}(X_1^2) + 2E_{P_{(X_1, X_2)}}(X_1X_2) + E_{P_{(X_1, X_2)}}(X_2^2) - [E_P(X_1)]^2 - 2E_P(X_1)E_P(X_2) - [E_P(X_2)]^2 =$
 $E_P(X_1^2) - [E_P(X_1)]^2 + E_P(X_2^2) - [E_P(X_2)]^2 + 2(E_{P_{(X_1, X_2)}}(X_1X_2) - E_P(X_1)E_P(X_2)) =$
 $V(X_1) + V(X_2) + 2KOV(X_1, X_2).$ $\square$

Da wie oben gezeigt bei unabhängigen Zufallsvariablen X_1 und X_2 $KOV(X_1, X_2) = 0$, gilt in diesem Fall: bei unabhängigen Zufallsvariablen ist die Varianz der Statistik Summe gleich der Summe der Varianzen der einzelnen Zufallsvariablen, die in die Berechnung der Statistik einfließen.

Satz 9.5.35. *Für die Summe von n i.i.d. Zufallsvariablen X_i gilt:*

$$V_{P_{(X_1, \dots, X_n)}}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V_P(X_i) = nV_P(X_i)$$

Beweis. Wir beweisen den Satz für zwei unabhängige und identisch verteilte Zufallsvariablen X_1 und X_2 :

$V_{P_{(X_1, X_2)}}(X_1 + X_2) = V_P(X_1) + V_P(X_2) = 2V_P(X_1)$ (da identisch verteilte Zufallsvariablen dieselbe Varianz haben).

Die Fortführung auf n unabhängige und identische Zufallsvariablen ergibt sich zwanglos. $\square$

Der folgende Satz ist von fundamentaler Bedeutung:

Satz 9.5.36. *Seien X_i i.i.d. und $V_P(X_i) < \infty$, dann gilt:*

$$V_{P_{(X_1, \dots, X_n)}}(\bar{X}_n) = \frac{V_P(X_i)}{n}$$

Beweis. $V_{P_{(X_1, \dots, X_n)}}(\bar{X}_n) \stackrel{\text{Definition } \bar{X}_n}{=} V_{P_{(X_1, \dots, X_n)}}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \stackrel{\text{Satz 9.5.21}}{=} \frac{1}{n^2} V_{P_{(X_1, \dots, X_n)}}\left(\sum_{i=1}^n X_i\right)$

$\stackrel{\text{Satz 9.5.35}}{=} \frac{1}{n^2} n V_P(X_i) = \frac{1}{n} V_P(X_i).$ $\square$

Der Satz besagt: Die Varianz des Mittelwertes von n i.i.d.-Zufallsvariablen X_i unter der gemeinsamen Verteilung $P_{(X_1, \dots, X_n)}$ ist n -mal kleiner als die Varianz der einzelnen Zufallsvariable X_i unter P , wenn diese Varianz endlich ist.

Für n gegen Unendlich wird in diesem Fall die Varianz 0, d.h.

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} V(X_i) &= V(X_i) \lim_{n \rightarrow \infty} \frac{1}{n} \\ &= V(X_i) \cdot 0 \\ &= 0. \end{aligned}$$

Die praktischen Auswirkungen dieser Sätze sind:

a) Wegen $E(\bar{X}_n) = E(X_i)$ können wir den Erwartungswert von i.i.d. Zufallsvariablen X_i mit Hilfe des Mittelwertes einer Stichprobe der Grösse n schätzen. „Im Mittel“ treffen wir dadurch $E(X_i)$. Wir betrachten die Stichprobenwerte als n Werte von i.i.d. Zufallsvariablen X_i und können damit den Mittelwert der Stichprobe als Wert der Zufallsvariable $\bar{X}_n$ betrachten. Dies ist nützlich, weil wegen

b) $V(\bar{X}) = \frac{1}{n} V(X_i)$ für $V(X_i) < \infty$ die Varianz der Zufallsvariable $\bar{X}$ um den Faktor $\frac{1}{n}$ kleiner ist als die Varianz von X_i . Lassen wir n genügend gross werden, können wir $E(X_i)$ mit Hilfe

des Mittels einer Stichprobe der Grösse n beliebig genau schätzen. Diese Aussage ist mit dem sogenannte *Schwachen Gesetz der Grossen Zahlen* (s. Seite 224) verwandt, welches unter Anderem auch den Zusammenhang zwischen Wahrscheinlichkeiten und relativen Häufigkeiten näher beleuchtet.

Gilt für eine Statistik $T : E(T) = \delta$, wobei δ ein Kennwert der i.i.d Zufallsvariablen ist, die in die Berechnung von T einfließen, dann nennt man T einen *erwartungstreuen Schätzer von δ* . Wir haben oben gezeigt, dass $\bar{X}$ ein erwartungstreuer Schätzer von $E(X_i)$ ist. Zudem ist

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ein erwartungstreuer Schätzer von $V(X_i)$, wie wir später zeigen werden (d.h. die Statistik Stichprobenvarianz, die konkreten Stichprobenvarianzen einer konkreten Stichprobe als Werte annimmt, ist ein erwartungstreuer Schätzer der Varianz der identisch verteilten Zufallsvariablen X_i , aus deren Werten wir konkrete Stichprobenvarianzen berechnen).

Ein interessantes Licht auf Erwartungswert und Varianz wirft der folgende Satz (vergleichen Sie das Ergebnis mit den Bemerkungen 4.1.9. 5 von Seite 47 - die quadrierte Summe der Abweichungen der Daten von einem Wert wird minimal, wenn der Wert mit dem Mittelwert identisch ist):

Satz 9.5.37. *Sei X eine Zufallsvariable und $a \in \mathbb{R}$ eine Konstante.*

(i) $E((X - a)^2) = V(X) + (E(X) - a)^2$

(ii) Entsprechend wird $E((X - a)^2)$ minimal genau dann, wenn $a = E(X)$.

Beweis. (i)

$$\begin{aligned} E((X - a)^2) &= E(X^2 - 2Xa + a^2) \\ &= E(X^2) - 2aE(X) + a^2 \\ &= E(X^2) - \underbrace{(E(X))^2 + (E(X))^2}_{=0} - 2aE(X) + a^2 \end{aligned}$$

Da mit Satz 9.5.19 $E(X^2) - (E(X))^2 = V(X)$ und $(E(X))^2 - 2aE(X) + a^2 = (E(X) - a)^2$ (Binom), gilt der Satz.

(ii) Da $V(X) \geq 0$ (Satz 9.3.7) und ebenso $(E(X) - a)^2 \geq 0$, wird das Minimum von $V(X) + (E(X) - a)^2$ erreicht, wenn $E(X) - a = 0$, d.h. $E(X) = a$. $\square$

Der Satz besagt, (i) dass die zu erwartende quadrierte Abweichung der Werte der Zufallsvariable vom Wert a sich aus der Varianz von X und der quadrierten Abweichung des Erwartungswertes von a zusammensetzt und (ii) dass die zu erwartende quadrierte Abweichung dann minimal ist, wenn man als Konstante den Erwartungswert von X wählt. In diesem Fall fällt die erwartete quadrierte Abweichung mit der Varianz zusammen.

9.6 Das Schwache Gesetz der Grossen Zahlen

Wir formulieren das Schwache Gesetz der Grossen Zahlen für eine Folge von identisch verteilten Zufallsvariablen.

Definition 9.6.1. Sei $(\Omega, \mathcal{P}(\Omega), P)$ eine diskreter Wahrscheinlichkeitsraum und (X_i) eine Folge von Zufallsvariablen mit $X_i : \Omega \rightarrow \mathbb{R}$ und $\mathfrak{X}_i := \text{Bild}(X_i)$ sowie $E(X_i) = E(X_j)$ für alle $i, j \in \mathbb{N}^*$. Ferner sei

$$\begin{aligned} \bar{X}_n &: \times_{i=1}^n \mathfrak{X}_i \rightarrow \mathbb{R} \\ \bar{X}_n(x) &:= \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{mit } x = (x_1, \dots, x_n)). \end{aligned}$$

Wenn für beliebige $k > 0$ gilt:

$$P(\{x \mid |\bar{X}_n(x) - E(X_i)| > k\}) = 0$$

für $n \rightarrow \infty$ (d.h. $\lim_{n \rightarrow \infty} P(\{x \mid |\bar{X}_n(x) - E(X_i)| > k\}) = 0$), dann gilt für die Folge von Zufallsvariablen X_i das schwache Gesetz der grossen Zahlen

Mit anderen Worten: Wenn das Schwache Gesetz der grossen Zahlen gilt, dann ist die Wahrscheinlichkeit 0, dass der Wert der Zufallsvariable $\bar{X}_n(x)$ mehr als k von $E(X_i)$ entfernt ist, wenn n gegen Unendlich geht. Dabei kann $k > 0$ beliebig klein gewählt werden (Wichtiger als ein Verständnis des folgenden Beweises ist die Durcharbeitung der darauf folgenden zwei Beispiele).

Satz 9.6.2. Für eine Folge von identisch verteilten und unabhängigen Zufallsvariablen X_i mit $V(X_i) < \infty$ gilt das schwache Gesetz der grossen Zahlen.

Beweis. (i) Zuerst sind ein paar Vorarbeiten zu leisten. Wir beweisen zuerst die sogenannte Markoffsche Ungleichung:

$$P(\{\omega \in \Omega \mid |X(\omega)| \geq t\}) \leq \frac{E(|X|^q)}{t^q}, \text{ für } q > 0$$

Beweis: Sei $A := \{\omega \in \Omega \mid |X(\omega)| \geq t\}$. Damit gilt

$$\begin{aligned} E(|X|^q) &= \sum_{x \in \mathfrak{X}} |x|^q P(X = x) \stackrel{\text{Satz 9.2.3}}{=} \sum_{\omega \in \Omega} |X(\omega)|^q P(\{\omega\}) \\ &\geq \sum_{\substack{\text{da } A \subset \Omega \\ \omega \in A}} |X(\omega)|^q P(\{\omega\}) \\ &\stackrel{\text{da } |X(\omega)| \geq t}{\geq} \sum_{\omega \in A} t^q P(\{\omega\}) \quad (\text{man ersetzt alle } |X(\omega)| \geq t \text{ durch } t) \\ &= t^q \sum_{\omega \in A} P(\{\omega\}) \\ &= t^q P(A). \end{aligned}$$

Daraus folgt:

$$P(A) = P(\{\omega \in \Omega \mid |X(\omega)| \geq t\}) \leq \frac{E(|X|^q)}{t^q} \quad \text{q.e.d.}$$

Aus der Markoffschen Ungleichung folgt die Tschebyscheffsche Ungleichung:

$$P(\{\omega \in \Omega \mid |X(\omega) - E(X)| \geq t\}) \leq \frac{V(X)}{t^2} \text{ für } t \in \mathbb{R}, t > 0.$$

Beweis: Wir setzen in der Markoffschen Ungleichung 2 für q und $X - E(X)$ für X . Wir erhalten:

$$P(\{\omega \in \Omega \mid |X(\omega) - E(X)| \geq t\}) \leq \frac{E(|X - E(X)|^2)}{t^2} = \frac{V(X)}{t^2} \text{ für } t \in \mathbb{R}, t > 0 \quad \text{q.e.d.}$$

(ii) Nun kann der eigentliche Satz bewiesen werden:

Setzt man in die Tschebyscheffsche Ungleichung $\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i))$ für X , erhält man:

$$\begin{aligned} (a) &:= P\left(\left\{\omega \in \Omega \mid \left|\frac{1}{n} \sum_{i=1}^n (X_i(\omega) - E(X_i)) - E\left(\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i))\right)\right| > t\right\}\right) \\ &\leq \frac{V\left(\frac{1}{n} \sum_{i=1}^n (X_i - E(X_i))\right)}{t^2} \end{aligned}$$

Nun gilt

$$\begin{aligned} E\left(\frac{1}{n}\sum_{i=1}^n(X_i - E(X_i))\right) \\ = \frac{1}{n}\sum_{i=1}^n(E(X_i) - E(X_i)) = 0 \end{aligned}$$

sowie

$$\begin{aligned} V\left(\frac{1}{n}\sum_{i=1}^n(X_i - E(X_i))\right) & \stackrel{E(X_i) \text{ ist Konstante (Satz 9.5.22)}}{=} V\left(\frac{1}{n}\sum_{i=1}^n(X_i)\right) \\ & \stackrel{\text{Satz 9.5.21, i.i.d. und 9.5.35}}{=} \frac{nV(X_i)}{n^2} = \frac{V(X_i)}{n} \end{aligned}$$

und damit:

$$\begin{aligned} (a) &= P\left(\left\{\omega \in \Omega \mid \left|\frac{1}{n}\sum_{i=1}^n(X_i(\omega) - E(X_i))\right| > t\right\}\right) \\ &\leq \frac{V(X_i)}{nt^2} \text{ für alle } t \in \mathbb{R}, t > 0. \end{aligned}$$

Es gilt

$$\lim_{n \rightarrow \infty} \frac{V(X_i)}{nt^2} = 0 \text{ für } V(X_i) < \infty$$

und damit

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\left\{\omega \in \Omega \mid \left|\frac{1}{n}\sum_{i=1}^n(X_i(\omega) - E(X_i))\right| > t\right\}\right) &= \\ \lim_{n \rightarrow \infty} P(\{x \in \mathfrak{X} \mid |\bar{X}_n(x) - E(X_i)| > t\}) &= 0 \end{aligned}$$

□

Beispiel 9.6.3. Wir nennen einen einzelnen Wurf des Würfels „ein Wurfexperiment“ (mit Augen; Gleichverteilung auf Ω). Wir führen n Würfe durch. Die n Zufallsvariablen X_i ordnen den nach oben liegenden Augenzahlen des Würfels deren Anzahl zu. Wir definieren die Statistik „Mittelwert“ $\bar{X}_n$ auf das kartesische Produkt der Bilder $\mathfrak{X}_i$ (siehe oben). Würden wir z.B. in 6 Würfen $(1, 3, 6, 5, 5, 4)$ werfen, würde $\bar{X}_n(1, 3, 6, 5, 5, 4) = \frac{1+3+6+5+5+4}{6} = 4$ gelten. Es gilt $E(X_i) = 3.5$. Das schwache Gesetz der grossen Zahlen besagt nun für z.B. $k = 1.5$. Die Wahrscheinlichkeit, dass $|\bar{X}_n - 3.5| > 1.5$ beträgt 0, wenn n gegen unendlich geht. Anders formuliert: Die Wahrscheinlichkeit, dass $\bar{X}_n$ mehr als 1.5 von 3.5 abweicht, sinkt mit steigendem n (je mehr Wurfexperimente wir berücksichtigen).

Die entsprechenden Wahrscheinlichkeiten könnten auch berechnet werden (siehe Übung 9.6.1.19). So ergibt sich fürs Beispiel ($k = 1.5$) (s. Abbildung 9.6.3):

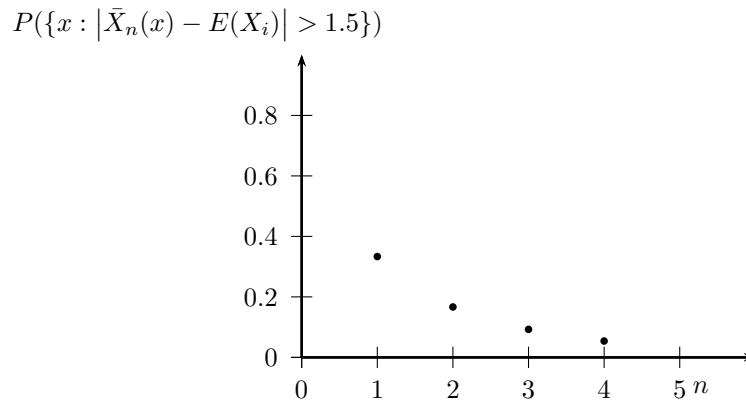


Abbildung 9.6.3: Schwaches Gesetz der Grossen Zahlen; Entwicklung von $P(\{x \mid \bar{X}_n(x) - E(X_i) \mid > 1.5\})$ als Funktion von n (= Anzahl Wurfexperimente; Berechnung Übung 9.6.1.19)

◇

Im besonderen ergibt sich aus dem Schwache Gesetz der Grossen Zahlen eine Aussage über den Zusammenhang von Wahrscheinlichkeiten und relativen Häufigkeiten, wie das folgende Beispiel zeigt.

Beispiel 9.6.4. Wir werfen n mal eine Münze. Bei Kopf wird 0 ausbezahlt, bei Zahl 1. Wir setzen die folgende Wahrscheinlichkeitsfunktion auf Ω voraus: $f = \{(K, 0.7), (Z, 0.3)\}$. Wir haben also $\Omega = \{K, Z\}$,
 $X_i = \{(K, 0), (Z, 1)\}$ ($i \in \mathbb{N}_n^*$).
 $\mathfrak{X}_i = \{0, 1\}$,
 $E(X_i) = 0.7 \cdot 0 + 0.3 \cdot 1 = 0.3 = P(X_i = 1)$ (= Wahrscheinlichkeit, eine 1 zu erhalten)

Wir betrachten nun $\bar{X}_n$. $\bar{X}_n$ ist ein erwartungstreuer Schätzer von $E(X_i) = P(X_i = 1)$. Denn $E(\bar{X}_n) = E(X_i) = P(X_i = 1)$. $\bar{X}_n$ nimmt als Werte das arithmetische Mittel der Komponenten von $(x_1, \dots, x_n)$ an. Dies entspricht der relativen Häufigkeit $h(1)$ der Einsen in einem n -fachen Wurfexperiment der Münze. Das Schwache Gesetz der Grossen Zahlen besagt nun im Beispiel: die relative Häufigkeit der Einsen in einem n -fachen Wurfexperiment weicht mit steigendem n mit einer immer kleineren Wahrscheinlichkeit mehr als k von der Wahrscheinlichkeit $P(X_i = 1) = 0.3$ ab. (z.B. $k = 0.01$). Je mehr Wurfexperimente wir durchführen, desto näher kommt mit hoher Wahrscheinlichkeit die relative Häufigkeit der Einsen der Wahrscheinlichkeit, eine 1 zu erhalten. Formal:

$$\lim_{n \rightarrow \infty} P(\{x \in \mathfrak{X} \mid |h(1) - P(X_i = 1)| > t\}) = 0$$

Wir könnten dieses Verhalten von $\bar{X}_n$ wiederum überprüfen, indem wir entsprechende Wahrscheinlichkeiten für ein spezifisches k berechnen. Wir können es aber auch empirisch überprüfen, indem wir entsprechende Experimente durchführen. Dazu simulieren wir das Werfen der Münzen mit Excel. Mit Excel können wir n Zufallszahlen produzieren, indem wir den Befehl „zufallszahl()“ in eine Zelle einfügen und diese über n Zellen hinunterziehen. Anschliessend fixieren wir die Zufallszahlen. Wir können aus diesen Zufallszahlen P_{X_i} -verteilte Zufallszahlen x_i produzieren - X_i definiert wie oben, indem wir in eine weitere Spalte den Befehl „wenn(A1<0.7;0;1)“ eingeben. Dann berechnen wir die relative Häufigkeit der Einsen (= arithmetisches Mittel) und die Abweichung von 0.3. Dies machen wir für $n = 10$, $n = 100$, $n = 1000$ und $n = 10000$. Im Beispiel der Excel-Tabelle 16Zufallsvariablen.xls erhalten wir:
für $n = 10$ eine Abweichung $|\bar{X}_{10} - 0.3|$ von 0.1

für $n = 100$ eine Abweichung $|\bar{X}_{100} - 0.3|$ von 0.03

für $n = 1000$ eine Abweichung $|\bar{X}_{1000} - 0.3|$ von 0.02

für $n = 10000$ eine Abweichung $|\bar{X}_{10000} - 0.3|$ von 0.01.

Bei der Reproduktion des Beispiels werden Sie vermutlich andere Werte erhalten. Insbesondere kann auch die Tendenz gestört werden. Es handelt sich ja um Zufallsresultate und wir können nur sagen, dass sich mit (angegebener) Wahrscheinlichkeit die Tendenz einstellt, nicht aber, dass sie sich mit Sicherheit ergibt. $\diamond$

Das schwache Gesetz der grossen Zahlen zeigt damit, wie relative Häufigkeiten mit Wahrscheinlichkeiten zusammenhängen. Je mehr Zufallsexperimente wir durchführen, desto näher liegen mit angegebener Wahrscheinlichkeit die relativen Häufigkeiten bei den Wahrscheinlichkeiten. Dies macht es vernünftig, Wahrscheinlichkeiten mit relativen Häufigkeiten zu schätzen und die Schätzung ist mit steigender Wahrscheinlichkeit umso besser, je mehr Zufallsexperimente wir durchführen, um die relativen Häufigkeiten zu berechnen.

9.6.1 Übungen

- Gegeben sei die folgende Wahrscheinlichkeitsfunktion: Graph von $f = \{(rot, 0.7), (grün, 0.2), (gelb, 0.1)\}$ und Graph der Zufallsvariable: $X = \{(rot, 8), (grün, 8), (gelb, 2)\}$. Bestimmen Sie die Graphen der Funktionen f_X und P_X .
- Berechnen Sie den Erwartungswert und die Varianz für die folgende Wahrscheinlichkeitsfunktion auf das Bild einer Zufallsvariable X :
Graph von $f_X = \{(5, 0.2), (6, 0.1), (1, 0.15), (3, 0.55)\}$
- Berechnen Sie den Erwartungswert, die Varianz und die Standardabweichung der Zufallsvariable von Übung 1)
- Gegeben sei ein Glücksspiel mit dem Graphen $f_X = \{(50, 0.01), (1, 0.1), (0, 0.89)\}$. (d.h. 50 wird mit einer Wahrscheinlichkeit von 0.01, 1 mit einer Wahrscheinlichkeit von 0.1 und 0 mit einer Wahrscheinlichkeit von 0.89 ausbezahlt). Berechnen Sie den durchschnittlich zu erwartenden Gewinn. Berechnen Sie die Standardabweichung.
- Wir nehmen an, Sie könnten folgende Entscheidung treffen: Entscheiden Sie sich für A , so erhalten Sie mit einer Wahrscheinlichkeit von $\frac{1}{6}$ 50 SFr. und mit einer Wahrscheinlichkeit von $\frac{5}{6}$ nichts. Entscheiden Sie sich für B , erhalten Sie mit einer Wahrscheinlichkeit von $\frac{2}{3}$ 5 SFr. und mit einer Wahrscheinlichkeit von $\frac{1}{6}$ 1 Sfr. und mit einer Wahrscheinlichkeit von $\frac{1}{6}$ nichts.
 - Für welches Spiel würden Sie sich spontan entscheiden?
 - Welches Risikoverhalten legt Ihre Entscheidung offen?
- Bei Entscheidung für A erhalten sie 40'000 SFr. mit Sicherheit. Bei einer Entscheidung für B erhalten Sie 1 Million SFr. mit einer Wahrscheinlichkeit von $\frac{1}{6}$, und 0 SFr. mit einer Wahrscheinlichkeit von $\frac{5}{6}$. Wofür würden Sie sich entscheiden? Wofür sollte man sich entscheiden?
- Eine Unternehmung entscheidet sich bei der Anlage von Gewinnen zwischen dem Kauf von Obligationen (Zins 3.5%, Wahrscheinlichkeit 1) und dem Aktienkauf, wobei sie mit folgenden Bedingungen rechnet: mit einer Wahrscheinlichkeit von $\frac{1}{5}$ 30% Verlust, mit einer Wahrscheinlichkeit von $\frac{2}{5}$ weder Verlust, noch Gewinn, mit einer Wahrscheinlichkeit von $\frac{1}{5}$ 20% Gewinn und mit einer Wahrscheinlichkeit von $\frac{1}{5}$ 30% Gewinn. Wofür sollte sich die Unternehmung entscheiden?
- Gegeben sei die folgende Verteilungsfunktion:
Graph von $F = \{(0.2, 0.01), (0.3, 0.05), (0.5, 0.1), (2, 0.5), (3, 0.7), (10, 0.9), (11, 0.99), (12, 1)\}$
Berechnen Sie die entsprechenden Graphen der Wahrscheinlichkeitsfunktion. Berechnen Sie

$$P(X = 10),$$

$$P(0.2 < X \leq 3),$$

$$P(X \geq 0.5)$$

9. $\Omega = \{1, 2, 3\}$. P auf $\mathcal{P}(\Omega)$ sei durch den folgenden Graphen der Wahrscheinlichkeitsfunktion f auf Ω gegeben: $\{(1, \frac{1}{4}), (2, \frac{1}{4}), (3, \frac{1}{2})\}$.
 Graph von $X_i = \{(1, 1), (2, 2), (3, 2)\}$ ($i \in \mathbb{N}_2^*$). Überprüfen Sie, ob die Zufallsvariablen X_1 und X_2 unabhängig sind, wenn deren Verteilung durch den folgenden Graphen einer Wahrscheinlichkeitsfunktion g auf $\mathfrak{X}_1 \times \mathfrak{X}_1$ gegeben wird:
 $\{((1, 1), \frac{1}{10}), ((1, 2), \frac{1}{5}), ((2, 1), \frac{3}{10}), ((2, 2), \frac{2}{5})\}$
10. Sei die folgende Tabelle gegeben, welche Paaren von Werten von Zufallsvariablen X_1 und X_2 jeweils eine Wahrscheinlichkeit zuordnet ($\mathfrak{X}_1 = \{1, 2, 5, 8\}$).

	1	2	5	8
3	0.1	0.2	0.1	0.01
4	0.2	0.02	0.1	0.02
6	0.05	0.03	0.022	0.148

- a) Berechnen Sie die Randverteilungen
 b) Wir nehmen an, diese Randverteilungen würden mit den Bild-W-Funktionen von X_1 und X_2 übereinstimmen, wodurch die obige Tabelle eine Gemeinsame W-Funktion von X_1 und X_2 liefert. Geben Sie f_{X_1} und f_{X_2} als Graphen an.
 c) Ist die vorliegende Gemeinsame W-Funktion von X_1 und X_2 die Produkte-W-Funktion bezüglich f_{X_1} und f_{X_2} ? Wenn nein, berechnen Sie diese.
 d) Sei die folgende Tabelle gegeben, welche eine W-Funktion auf $\mathfrak{X}_1 \times \mathfrak{X}_1$ definiert.

	1	2	5	8
3	0.1	0.01	0.02	0.03
4	0.02	0.1	0.55	0.03
6	0.03	0.01	0.03	0.07

Ist dadurch eine Gemeinsame W-Funktion von X_1 und X_2 , wie sie unter b) definiert ist, gegeben? Wenn nicht, geben Sie mit Hilfe von Graphen die W-Funktionen von X'_1 und X'_2 , so dass die obige Tabelle einer Gemeinsamen W-Funktion von X'_1 und X'_2 entspricht.

11. Vervollständigen Sie die Berechnung für $f_T(t)$, $t \in \mathcal{D}_T$ (siehe Aufgabenstellung von Übung 9.5.2 im Text). Berechnen Sie den Erwartungswert und die Varianz von T .
12. Vervollständigen Sie die Berechnung von f_M von Übung 9.5.4. Berechnen Sie den Erwartungswert und die Varianz von M .
13. Vervollständigen Sie die Berechnung von $f_{\bar{X}}$ von Übung 9.5.6. Berechnen Sie den Erwartungswert und die Varianz von $\bar{X}$ sowie von X_i . Vergleichen Sie!
14. Kontrollieren Sie das Ergebnis von Übung 9.5.8. Berechnen Sie den Erwartungswert und die Varianz von $\bar{X}$ und von X_i . Vergleichen Sie!
15. Berechnen Sie das Produktmass für die folgenden drei Zufallsvariablen X_i mit $i \in \{1, 2, 3\}$ und $X_i = \{(1, 2), (3, 4)\}$, wobei $\Omega = \{1, 3\}$ und P (=Wahrscheinlichkeitsmass auf $\mathcal{P}(\Omega)$) durch den Graphen von $f = \{(1, 0.2), (3, 0.8)\}$ (Wahrscheinlichkeitsfunktion auf Ω) gegeben ist.

16. Sei $\Omega = \{Kopf, Zahl\}$ mit Gleichverteilung und $X_i(Kopf) = 0$ und $X_i(Zahl) = 2$ mit $i \in \{1, 2, 3, 4, 5\}$. Die Zufallsvariablen X_i sind unabhängig. Berechnen Sie die Verteilung der Zufallsvariable Summe T auf das Kartesische Produkt $\times_{i=1}^5 \mathfrak{X}_i =: \mathfrak{X}$. Berechnen Sie den Erwartungswert $E(T)$ und die Varianz $V(T)$. Berechnen Sie auch $E(X_i)$ sowie $V(X_i)$ und vergleichen Sie!
17. Sei $\Omega = \{Kopf, Zahl\}$ mit Gleichverteilung und $X_i(Kopf) = 0$ und $X_i(Zahl) = 2$ mit $i \in \{1, 2, 3, 4, 5, 6\}$. Die Zufallsvariablen X_i sind unabhängig. Berechnen Sie die Verteilung der Zufallsvariable Varianz S^2 auf das Kartesische Produkt $\times_{i=1}^6 \mathfrak{X}_i$.
 S^2 ist definiert als:
 $S^2 : \times_{i=1}^n \mathfrak{X}_i \rightarrow \mathbb{R}$
 $S^2((x_1, \dots, x_n)) := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ mit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.
 Berechnen Sie $E(S^2)$ und $V(S^2)$.
18. Würfeln Sie 10 mal, 100 mal und 1000 mal und berechnen Sie die relativen Häufigkeiten des Eintretens der Augenzahlen 1, 2, 3, 4, 5, und 6. Berechnen Sie die Summen des Betrages der Abweichungen der relativen Häufigkeiten von den vorausgesetzten Wahrscheinlichkeiten für $n = 10, 100, 1000$. Beobachten Sie, wie sich die relativen Häufigkeiten den Wahrscheinlichkeiten $\frac{1}{6}$ nähern (am besten simuliert man dabei das Würfelspiel mit Hilfe von in Excel erzeugten Zufallszahlen).
19. Berechnen Sie $P(\{x : |\bar{X}_n(x) - E(X_i)| > 1.5\})$ für $n \in \mathbb{N}_4^*$ (siehe für Definitionen von X_i und $\bar{X}_n$ Beispiel 9.5.8).

9.6.2 Lösungen

- Graph von $f_X = \{(8, 0.7 + 0.2), (2, 0.1)\}$
 $\mathcal{P}(\{8, 2\}) = \{\emptyset, \{8\}, \{2\}, \{8, 2\}\}$ (= Potenzmenge des Bildes der Zufallsvariable)
 $P(\emptyset) = 0$; $P(\{8\}) = 0.9$, $P(\{2\}) = 0.1$, $P(\{8, 2\}) = 1$ oder mit Hilfe des Graphen von
 $P : P_X = \{(\emptyset, 0), (\{8\}, 0.9), (\{2\}, 0.1), (\{8, 2\}, 1)\}$ (= Wahrscheinlichkeitsverteilung)
- $E(X) = (0.2 \cdot 5) + (0.1 \cdot 6) + (0.15 \cdot 1) + (0.55 \cdot 3) = 3.4$
 $V(X) = 0.2(5 - 3.4)^2 + 0.1(6 - 3.4)^2 + 0.15(1 - 3.4)^2 + 0.55(3 - 3.4)^2 = 2.14$
- Graph von $f_X = \{(8, 0.7 + 0.2), (2, 0.1)\}$. Damit gilt dann:
 $E(X) = 0.9 \cdot 8 + 0.1 \cdot 2 = 7.4$
 $V(X) = 0.9(8 - 7.4)^2 + 0.1(2 - 7.4)^2 = 3.24$
 $\sqrt{V(X)} = \sqrt{3.24} = 1.8$
- $E(X) = 0.01 \cdot 50 + 0.1 \cdot 1 + 0.89 \cdot 0 = 0.6$
 $V(X) = 0.01(50 - 0.6)^2 + 0.1(1 - 0.6)^2 + 0.89(0 - 0.6)^2 = 24.74$
 $\sqrt{V(X)} = \sqrt{24.74} = 4.9739$
- Die Entscheidung für A kann dargestellt werden mit: Graph von
 $f_A = \{(50, \frac{1}{6}), (0, \frac{5}{6})\} \implies E(A) = \frac{1}{6} \cdot 50 + 0 \cdot \frac{5}{6} = \frac{25}{3} = 8.3333$
 Die Entscheidung für B kann dargestellt werden mit:
 Graph von $f_B = \{(5, \frac{2}{3}), (1, \frac{1}{6}), (0, \frac{1}{6})\} \implies E(B) = \frac{2}{3} \cdot 5 + 1 \cdot \frac{1}{6} + 0 \cdot \frac{1}{6} = 3.5$
 Der Erwartungswert der ersten Variante ist höher. Allerdings müssten wir eine allfällige Entscheidung unter Berücksichtigung der Standardabweichung beurteilen:
 $\sqrt{(50 - 8.3333)^2 \cdot \frac{1}{6} + (0 - 8.33333)^2 \cdot \frac{5}{6}} = 18.634$
 $\sqrt{(5 - 3.5)^2 \cdot \frac{2}{3} + (1 - 3.5)^2 \cdot \frac{1}{6} + (0 - 3.5)^2 \cdot \frac{1}{6}} = 2.1409$
 Die zweite Alternative ist also weniger risikoreich.

6. Die Entscheidung für A kann man darstellen mit Hilfe von:

Graph von $f_A = \{(40'000, 1)\}$

$$E(A) = 40'000 \cdot 1 = 40'000$$

Die Entscheidung für B kann man darstellen mit Hilfe von:

Graph von $f_B = \{(1'000'000, \frac{1}{6}), (0, \frac{5}{6})\}$

$$E(B) = 1000000 \cdot \frac{1}{6} + 0 \cdot \frac{5}{6} = \frac{500'000}{3} = 1.666\overline{7} \times 10^5$$

Ohne Berücksichtigung des Risikos müsste somit B gewählt werden. Vermutlich gibt es Personen, die sich für A entscheiden. Diese wären risikoavers. Bei so hohen Einsätzen und bei der grossen Standardabweichung der zweiten Alternative von

$\sqrt{(1000000 - 1.666\overline{7} \times 10^5)^2 \cdot \frac{1}{6} + (0 - 1.666\overline{7} \times 10^5)^2 \cdot \frac{5}{6}} = 3.7268 \times 10^5$ ist das jedoch durchaus vernünftig.

7. Das Kapital sei K . Bei einem Zins von 3.5% erhalten wir $1.035K$. Da die Sicherheit maximal ist ($= 1$), ergibt dies einen Erwartungswert von $1.035K$.

Für die Aktien ergibt sich:

$$(0.7 \cdot K \cdot \frac{1}{5}) + (1 \cdot K \cdot \frac{2}{5}) + (1.2 \cdot K \cdot \frac{1}{5}) + (1.3 \cdot K \cdot \frac{1}{5}) = 1.04K$$

Während die erste Variante keine Streuung aufweist, ist die Streuung der zweiten Variante:

$K \sqrt{(0.7 - 1.04)^2 \cdot \frac{1}{5} + (1 - 1.04)^2 \cdot \frac{2}{5} + (1.2 - 1.04)^2 \cdot \frac{1}{5} + (1.3 - 1.04)^2 \cdot \frac{1}{5}} = 0.20591K$. Ob der grössere Gewinn durch das grössere Risiko ausgeglichen wird oder nicht, ist eine Frage der Risikoeinstellung.

8. Graph von $F = \{(0.2, 0.01), (0.3, 0.05), (0.5, 0.1), (2, 0.5), (3, 0.7), (10, 0.9), (11, 0.99), (12, 1)\}$
 Graph von $f_X = \{(0.2, 0.01), (0.3, 0.04), (0.5, 0.05), (2, 0.4), (3, 0.2), (10, 0.2), (11, 0.09), (12, 0.01)\}$
 Kontrolle: $0.01 + 0.04 + 0.05 + 0.4 + 0.2 + 0.2 + 0.09 + 0.01 = 1$

$$P(X = 10) = F(10) - F(3) = 0.9 - 0.7 = 0.2$$

$$P(0.2 < X \leq 3) = F(3) - F(0.2) = 0.7 - 0.01 = 0.69$$

$$P(X \geq 0.5) = 1 - P(X \leq 0.3) = 1 - F(0.3) = 1 - 0.05 = 0.95$$

9. Die Graphen von $f_{X_i} = \{(1, \frac{1}{4}), (2, \frac{3}{4})\}$. Die Randverteilungen von g stimmen nicht mit f_{X_i} überein. Deshalb kann g keine Gemeinsame Verteilung der Zufallsvariablen X_1 und X_2 sein:

	1	2	Summen
1	0.1	0.2	0.3
2	0.3	0.4	0.7
Summen	0.4	0.6	1

Tabelle 9.6.5:

Da Produktmasse von Zufallsvariablen Gemeinsame Verteilungen dieser Zufallsvariablen, kann g nicht das Produktmass sein, was man noch rechnerisch bestätigen kann: Die Wahrscheinlichkeitsfunktion $f_{(X_1, X_2)}$ des Produktmasses ist:

$$\{((1, 1), \frac{1}{16}), ((1, 2), \frac{3}{16}), ((2, 1), \frac{3}{16}), ((2, 2), \frac{9}{16})\}$$

$$\text{Probe: } \frac{1}{16} + \frac{3}{16} + \frac{3}{16} + \frac{9}{16} = 1$$

Der Graph der Wahrscheinlichkeitsfunktion des Produktmasses stimmt nicht mit dem gelieferten Graphen g überein. Es hat wenig Sinn bezüglich g von Unabhängigkeit oder Abhängigkeit zu reden, da keine Gemeinsame Verteilung der Zufallsvariablen vorliegt.

10. a) Randverteilungen

	1	2	5	8	Summen
3	0.1	0.2	0.1	0.01	0.41
4	0.2	0.02	0.1	0.02	0.34
6	0.05	0.03	0.022	0.148	0.25
Summen	0.35	0.25	0.222	0.178	1

- b) Graph von $f_{X_1} = (1, 0.35), (2, 0.25), (5, 0.222), (8, 0.178)$ und $f_{X_2} = (3, 0.41), (4, 0.34), (6, 0.25)$
c) Die vorliegende Gemeinsame W-Funktion von X_1 und X_2 ist nicht das Produkte-W-Mass, dieses ist gegeben durch:

	1	2	5	8	Summen
3	0.1435	0.1025	0.09102	0.07298	0.41
4	0.119	0.085	0.07548	0.06052	0.34
6	0.0875	0.0625	0.0555	0.0445	0.25
Summen	0.35	0.25	0.222	0.178	1

Tabelle 9.6.6:

Man sieht, dass die Produkte-W-Funktion nicht mit der obigen Gemeinsamen W-Funktion übereinstimmt.

d)

	1	2	5	8	Summen
3	0.1	0.01	0.02	0.03	0.16
4	0.02	0.1	0.55	0.03	0.7
6	0.03	0.01	0.03	0.07	0.14
Summen	0.15	0.12	0.6	0.13	1

Tabelle 9.6.7:

Die Randverteilungen stimmen nicht mit f_{X_1} und f_{X_2} überein. Es liegt keine Gemeinsame Verteilung von X_1 und X_2 vor. Demgegenüber wird durch die Tabelle eine Gemeinsame Verteilung der Zufallsvariablen X'_1 und X'_2 geliefert, deren Verteilung durch die Graphen von und $f_{X'_1} : \{(1, 0.15), (2, 0.12), (5, 0.6), (8, 0.13)\}$ und $f_{X'_2} : \{(3, 0.16), (4, 0.7), (6, 0.14)\}$ gegeben sind. Diese Gemeinsame Verteilung von X'_1 und X'_2 ist nicht das Produktmass, da z.B. $(1, 3)$ 0.1 statt $0.15 \cdot 0.16 = 0.024$ zugeordnet ist.

11. Es gilt: $T((1, 1)) = 2 \implies f_T(2) = \frac{1}{36}$.
 $T((1, 2)) = T((2, 1)) = 3 \implies f_T(3) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$
 $T((1, 3)) = T((2, 2)) = T((3, 1)) = 4 \implies f_T(4) = \frac{3}{36}$
 $T((1, 4)) = T((2, 3)) = T((3, 2)) = T((4, 1)) = 5 \implies f_T(5) = \frac{4}{36}$
 $T((1, 5)) = T((2, 4)) = T((3, 3)) = T((4, 2)) = T((5, 1)) = 6 \implies f_T(6) = \frac{5}{36}$
 $T((1, 6)) = T((2, 5)) = T((3, 4)) = T((4, 3)) = T((5, 2)) = T((6, 1)) = 7 \implies f_T(7) = \frac{6}{36}$
 $T((2, 6)) = T((3, 5)) = T((4, 4)) = T((5, 3)) = T((6, 2)) = 8 \implies f_T(8) = \frac{5}{36}$
 $T((3, 6)) = T((4, 5)) = T((5, 4)) = T((6, 3)) = 9 \implies f_T(9) = \frac{4}{36}$
 $T((4, 6)) = T((5, 5)) = T((6, 4)) = 10 \implies f_T(10) = \frac{3}{36}$
 $T((5, 6)) = T((6, 5)) = 11 \implies f_T(11) = \frac{2}{36}$
 $T((6, 6)) = 12 \implies f_T(12) = \frac{1}{36}$

$$E(T) = \sum_{t \in \mathfrak{D}_T} t \cdot P(T = t) =$$

$$2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + \\ 8 \cdot \frac{7}{36} + 9 \cdot \frac{8}{36} + 10 \cdot \frac{9}{36} + 11 \cdot \frac{10}{36} + 12 \cdot \frac{11}{36} = 7$$

$$V(T) = \frac{1}{36}(2-7)^2 + \frac{2}{36}(3-7)^2 + \frac{3}{36}(4-7)^2 + \frac{4}{36}(5-7)^2 + \\ \frac{5}{36}(6-7)^2 + \frac{6}{36}(7-7)^2 + \frac{7}{36}(8-7)^2 + \frac{8}{36}(9-7)^2 + \frac{9}{36}(10-7)^2 + \\ \frac{10}{36}(11-7)^2 + \frac{11}{36}(12-7)^2 = \frac{35}{6} = 5.833\bar{3}$$

12. Graph von $f_M = \{(1, \frac{1}{36}), (2, \frac{2}{36}), (3, \frac{3}{36}), (4, \frac{4}{36}), (5, \frac{5}{36}), (6, \frac{6}{36}), (8, \frac{2}{36}), (9, \frac{1}{36}), (10, \frac{2}{36}), (12, \frac{4}{36}), (15, \frac{2}{36}), (16, \frac{1}{36}), (18, \frac{2}{36}), (20, \frac{2}{36}), (24, \frac{2}{36}), (25, \frac{1}{36}), (30, \frac{2}{36}), (36, \frac{1}{36})\}$ (Kontrolle: Die Summe der Elemente des Bildes von $f_X = 1$).

$$E(M) = \sum_{m \in \mathfrak{M}} m \cdot P(M = m) = 1 \cdot \frac{1}{36} + 2 \cdot \frac{2}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{2}{36} + \\ 6 \cdot \frac{4}{36} + 8 \cdot \frac{2}{36} + 9 \cdot \frac{1}{36} + 10 \cdot \frac{2}{36} + 12 \cdot \frac{4}{36} + 15 \cdot \frac{2}{36} + 16 \cdot \frac{1}{36} + 18 \cdot \frac{2}{36} + \\ 20 \cdot \frac{2}{36} + 24 \cdot \frac{2}{36} + 25 \cdot \frac{1}{36} + 30 \cdot \frac{2}{36} + 36 \cdot \frac{1}{36} = \frac{49}{4} = 12.25.$$

$$V(M) = \frac{1}{36}(1-12.25)^2 + \frac{2}{36}(2-12.25)^2 + \frac{2}{36}(3-12.25)^2 + \frac{3}{36}(4-12.25)^2 + \\ \frac{2}{36}(5-12.25)^2 + \frac{4}{36}(6-12.25)^2 + \frac{2}{36}(8-12.25)^2 + \frac{1}{36}(9-12.25)^2 + \\ \frac{2}{36}(10-12.25)^2 + \frac{4}{36}(12-12.25)^2 + \frac{2}{36}(15-12.25)^2 + \frac{1}{36}(16-12.25)^2 + \\ \frac{2}{36}(18-12.25)^2 + \frac{2}{36}(20-12.25)^2 + \frac{2}{36}(24-12.25)^2 + \frac{1}{36}(25-12.25)^2 + \\ \frac{2}{36}(30-12.25)^2 + \frac{1}{36}(36-12.25)^2 = 79.965$$

13. Graph von $f_{\bar{X}} = \{(1, \frac{1}{36}), (1.5, \frac{2}{36}), (2, \frac{3}{36}), (2.5, \frac{4}{36}), (3, \frac{5}{36}), (3.5, \frac{6}{36}), (4, \frac{5}{36}), (4.5, \frac{4}{36}), (5, \frac{3}{36}), (5.5, \frac{2}{36}), (6, \frac{1}{36})\}$ (Kontrolle: Die Summe der Elemente des Bildes von $f_{\bar{X}} = 1 = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36}$).

$$E(\bar{X}) = E(X) = \sum_{\bar{x} \in \Omega_{\bar{X}}} \bar{x} \cdot P(\bar{X} = \bar{x}) = 1 \cdot \frac{1}{36} + 1.5 \cdot \frac{2}{36} + 2 \cdot \frac{3}{36} + 2.5 \cdot \frac{4}{36} + 3 \cdot \frac{5}{36} + 3.5 \cdot \frac{6}{36} + 4 \cdot \frac{5}{36} +$$

$$4.5 \cdot \frac{4}{36} + 5 \cdot \frac{3}{36} + 5.5 \cdot \frac{2}{36} + 6 \cdot \frac{1}{36} = 3.5$$

$$V(\bar{X}) = V(X) = \frac{1}{36}(1-3.5)^2 + \frac{2}{36}(1.5-3.5)^2 + \frac{3}{36}(2-3.5)^2 + \frac{4}{36}(2.5-3.5)^2 + \\ \frac{5}{36}(3-3.5)^2 + \frac{6}{36}(3.5-3.5)^2 + \frac{5}{36}(4-3.5)^2 + \frac{4}{36}(4.5-3.5)^2 + \frac{3}{36}(5-3.5)^2 + \\ \frac{2}{36}(5.5-3.5)^2 + \frac{1}{36}(6-3.5)^2 = 1.458\bar{3}$$

$$E(X_i) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$$

$$V(X_i) = \frac{1}{6}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \frac{1}{6}(3-3.5)^2 + \frac{1}{6}(4-3.5)^2 +$$

$$+ \frac{1}{6}(5-3.5)^2 + \frac{1}{6}(6-3.5)^2 = 2.916\bar{7}$$

Wir stellen fest: $E(X_i) = E(\bar{X})$

$$V(\bar{X}) = \frac{1}{2}V(X_i).$$

Es handelt sich um wesentliche Ergebnisse, die wir später allgemein beweisen werden und die für die Statistik eine wichtige Rolle spielen.

14. $E(\bar{X}) = E(X_i) = 1 \cdot \frac{1}{16} + 1.25 \cdot \frac{4}{16} + 1.5 \cdot \frac{6}{16} + 1.75 \cdot \frac{4}{16} + 2 \cdot \frac{1}{16} = 1.5$

$$V(\bar{X}) = V(X_i) = (1-1.5)^2 \frac{1}{16} + (1.25-1.5)^2 \frac{4}{16} + (1.5-1.5)^2 \frac{6}{16} +$$

$$(1.75-1.5)^2 \frac{4}{16} + (2-1.5)^2 \frac{1}{16} = 0.0625$$

$$E(X_i) = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2} = 1.5$$

$$V(X_i) = (1-1.5)^2 \frac{1}{2} + (2-1.5)^2 \frac{1}{2} = 0.25$$

Es gilt wiederum: $E(X_i) = E(\bar{X})$ sowie

$$V(\bar{X}) = \frac{1}{4}V(X_i)$$

15. Das Produktemass liefern wir mit Hilfe des Graphen der Wahrscheinlichkeitsfunktion auf $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 : \{((2, 2, 2), 0.2^3), ((2, 2, 4), 0.2^2 \cdot 0.8), ((2, 4, 2), 0.2^2 \cdot 0.8), ((4, 2, 2), 0.2^2 \cdot 0.8), ((2, 4, 4), 0.2 \cdot 0.8^2), ((4, 2, 4), 0.2 \cdot 0.8^2), ((4, 4, 2), 0.2 \cdot 0.8^2), ((4, 4, 4), 0.8^3)\}$

$$\text{Kontrolle: } 0.2^3 + (3 \cdot 0.2^2 \cdot 0.8) + (3 \cdot 0.2 \cdot 0.8^2) + 0.8^3 = 1$$

Weist eine Wahrscheinlichkeitsverteilung auf $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3$ nicht diese Verteilung auf, so sind die betroffenen Zufallsvariablen nicht unabhängig bezüglich P .

16. Es ist leichter, die Dinge mit Excel zu berechnen (siehe 16Zufallsvariablen.xls).

17. Auch hier rechnen wir mit Excel (siehe 16Zufallsvariablen.xls).

18. Siehe 16Zufallsvariablen.xls: Im Beispiel erhalten wir als *Summe* des Betrages der Abweichungen der relativen Häufigkeiten von den vorausgesetzten Wahrscheinlichkeiten für $n = 10 : 0.466666667$, $n = 100 : 0.326666667$; $n = 1000 : 0.109333333$). Beachten Sie: die Tendenz muss sich im konkreten Fall nicht einstellen, da die konkrete Abweichung vom Zufall abhängt. Die entsprechenden Wahrscheinlichkeiten, dass sich die Abweichungen ausserhalb eines bestimmten Intervalls befinden, könnten berechnet werden.
19. (I) Für $n = 1$ wird $\bar{x}$ als die Augenzahl der Ergebnisses betrachtet. Somit ergeben sich vier Ergebnisse auf 6, die innerhalb des Intervalls $[1.5 - E(X), E(X) + 1.5] = [2, 5]$ ($E(X) = 3.5$) liegen (nämlich 2, 3, 4, 5) was eine Wahrscheinlichkeit von $\frac{2}{6} = \frac{1}{3}$ für $P(\{x : |\bar{X}_n(x) - E(X)| > 1.5\}) = P(\bar{x} : |\bar{x} - E(X)| > 1.5)$ ergibt).

(II) $n = 2$: Wie gross ist die Wahrscheinlichkeit, dass bei *zwei* Wurfexperimenten des obigen Typs $|\bar{x} - E(X)| > 1.5$. $\mathfrak{X}_1 \times \mathfrak{X}_2 = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\}$. Es gibt 6^2 Paare (x, y) . Jedem ist ein Mittelwert (Mittelwert der Komponenten = Mittelwert der Augenzahlen) zugeordnet. Diese Zuordnung sind die Werte der Zufallsvariable $\bar{X}$. Graph von $\bar{X} = \{((x_1, x_2), \bar{x}) : (x_1, x_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2 \text{ und } \bar{x} = \frac{x_1 + x_2}{2}\}$.

Wir suchen nun die (x_1, x_2) für die gilt: $\bar{x} < 2$ oder $\bar{x} > 5$:

$\bar{x}$ von $(1, 1) = \frac{1+1}{2} = 1 = \bar{X}((1, 1))$ (1 Fall) $\implies P(\bar{X} = 1) = \frac{1}{6^2}$

$\bar{x}$ von $(2, 1), (1, 2) = 1.5$ (2 Fälle) $\implies P(\bar{X} = 1.5) = 2 \cdot \frac{1}{6^2}$

$\bar{x}$ von $(2, 2) = 2$ (Ausserhalb des untersuchten Intervalls).

$\bar{x}$ von $(6, 6) = 6$ (1 Fall) $\implies P(\bar{X} = 6) = \frac{1}{6^2}$

$\bar{x}$ von $(6, 5), (5, 6) = 5.5$ (2 Fälle) $\implies P(\bar{X} = 5.5) = 2 \cdot \frac{1}{6^2}$

$\bar{x}$ von $(5, 5) = 5$ (Ausserhalb des untersuchten Intervalls).

$P(\{\bar{x} : |\bar{x} - E(X)| > 1.5\}) = 2 \cdot (3 \cdot \frac{1}{6^2}) = \frac{1}{6} = 0.16667$

Die Wahrscheinlichkeit, dass $\bar{x} \in [2, 5]$ beträgt somit $\frac{5}{6}$.

Einfacher kann man wie folgt rechnen: es gibt drei Paare, so dass das Mittel unter $3.5 - 1.5 = 2$ liegt. $(1, 1), (1, 2), (2, 1)$. Wegen der Symmetrie der Verteilung gibt es dann ebenso viele auf der andren Seite. Also ist die Wahrscheinlichkeit $\frac{2 \cdot 3}{36} = 0.16667$

(III) $n = 3$: Wie gross ist die Wahrscheinlichkeit, dass bei 3 Wurfexperimenten des obigen Typs $|\bar{x} - E(X)| > 1.5$.

$\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 = \{(x, y, z) : x, y, z \in \{1, 2, 3, 4, 5, 6\}\}$

Es gibt 6^3 Triple (x, y, z) . Jedem dieser ist ein Mittelwert (= Mittelwert der Augenzahlen) zugeordnet. Graph von $\bar{X} = \{((x_1, x_2, x_3), \bar{x}) : (x_1, x_2, x_3) \in \mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 \text{ und } \bar{x} = \frac{x_1 + x_2 + x_3}{3}\}$.

$\bar{x}$ von $(1, 1, 1) = 1. \implies P(\bar{X} = 1) = \frac{1}{6^3}$

$\bar{x}$ von $(1, 1, 2), (1, 2, 1)$ und $(2, 1, 1) = 1.\bar{3} \implies P(\bar{X} = 1.\bar{3}) = 3 \cdot \frac{1}{6^3}$

$\bar{x}$ von $(1, 2, 2), (2, 1, 2)$ und $(2, 2, 1) = 1.\bar{6} \implies P(\bar{X} = 1.\bar{6}) = 3 \cdot \frac{1}{6^3}$

$\bar{x}$ von $(2, 2, 2) = 2$

$\bar{x}$ von $(1, 1, 3), (1, 3, 1), (3, 1, 1) = 1.\bar{6} \implies P(\bar{X} = 1.\bar{6}) = 3 \cdot \frac{1}{6^3}$

$\bar{x}$ von $(1, 1, 4) = 2$

$\bar{x}$ von $(1, 2, 3) = 2$ (es gibt keine weiteren zu untersuchenden Tripel für $\bar{x} < 2$).

$\bar{x}$ von $(6, 6, 6) = 6 \implies P(\bar{X} = 6) = \frac{1}{6^3}$

$\bar{x}$ von $(6, 6, 5), (6, 5, 6)$ und $(5, 6, 6) = 5.\bar{6} \implies P(\bar{X} = 5.\bar{6}) = 3 \cdot \frac{1}{6^3}$.

$\bar{x}$ von $(6, 5, 5), (5, 6, 5)$ und $(5, 5, 6) = 5.\bar{3} \implies P(\bar{X} = 5.\bar{3}) = 3 \cdot \frac{1}{6^3} = 0.013889$.

$\bar{x}$ von $(5, 5, 5) = 5$.

$\bar{x}$ von $(6, 6, 4), (6, 4, 6), (4, 6, 6) = 5.\bar{3} \implies P(\bar{X} = 5.\bar{3}) = 3 \cdot \frac{1}{6^3} = 0.013889$.

(es gibt keine weiteren zu untersuchenden Tripel für $\bar{x} > 5$).

Somit gilt: Die Wahrscheinlichkeit, dass $|\bar{x} - E(X)| > 1.5 = 2 \cdot (10 \cdot \frac{1}{6^3}) = 0.092593$

Die Wahrscheinlichkeit dass $\bar{x} \in [2, 5]$ beträgt somit $1 - 0.092593 = 0.90741$.

Auch hier geht es einfacher: man zähle die Anzahl der Tripel, die ein Mittel unter 2 haben:

$(1, 1, 1), (1, 1, 2), (1, 2, 1), (2, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1), (3, 1, 1), (1, 3, 1), (1, 1, 3)$: Man erhält wegen der Symmetrie $\frac{2 \cdot 10}{6^3} = 9.2593 \times 10^{-2}$

(IV) $n = 4$: Wie hoch ist die Wahrscheinlichkeit, dass bei 4 Wurfexperimenten des obigen Typs $|\bar{x} - E(X)| > 1.5$. $\Omega := \mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 \times \Omega_{X_4} = \{(x_1, x_2, x_3, x_4) : x_1, x_2, x_3, x_4 \in \{1, 2, 3, 4, 5, 6\}\}$. Es gibt 6^4 4-Tupel (x_1, x_2, x_3, x_4) .

Graph von $\bar{X} = \{((x_1, x_2, x_3, x_4), \bar{x}) : (x_1, x_2, x_3, x_4) \in \Omega \text{ und } \bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{4}\}$.

$\bar{x}$ von $(1, 1, 1, 1) : 1$ (1 Fall)

$\bar{x}$ von $(1, 1, 1, 2) : 1.25$ (4 Fälle = $\binom{4}{1}$)

$\bar{x}$ von $(1, 1, 2, 2) : 1.5$ (6 Fälle = $\binom{4}{2}$)

$\bar{x}$ von $(1, 2, 2, 2) : 1.75$ ($\binom{4}{1} = 4$ Fälle)

$\bar{x}$ von $(2, 2, 2, 2) : 2 \in [2, 5]$.

$\bar{x}$ von $(1, 1, 1, 3) : 1.25$ (4 Fälle)

$\bar{x}$ von $(1, 1, 2, 3) : 1.75$ ($\frac{4!}{2} = 12$ Fälle).

$\bar{x}$ von $(1, 1, 3, 3) : 2 \in [2, 5]$.

$\bar{x}$ von $(1, 2, 2, 3) : 2 \in [2, 5]$.

$\bar{x}$ von $(1, 1, 1, 4) : 1.75$ (4 Fälle).

$\bar{x}$ von $(1, 1, 2, 4) : 2 \in [2, 5]$.

Damit wurden alle zu betrachtenden Fälle bezüglich $\bar{x} < 2$ berücksichtigt. Für $\bar{x} > 5$ gibt es gleich viele Fälle. Somit gilt: $P(\{\bar{x} : |\bar{x} - E(X)| > 1.5\}) = 2 \cdot (35 \cdot \frac{1}{6^4}) = 0.054012$

Wir sehen, dass - in Bestätigung des Gesetzes der grossen Zahlen - die Wahrscheinlichkeit dafür, dass $\bar{x}$ ausserhalb von $[2, 5]$ liegt, von $\frac{1}{3}$ auf, 0.16667, 0.092593 und 0.054012 gesunken ist.

Bemerkung: effizient kann man die obige Aufgabe mit R lösen: für

$n = 2$: `n2=expand.grid(1:6,1:6); n2s=apply(n2,1,mean); n2t=table(abs(n2s-3.5)>1.5); n2t[2]/sum(n2t)`

$n = 3$: `n3=expand.grid(1:6,1:6,1:6); n3s=apply(n3,1,mean); n3t=table(abs(n3s-3.5)>1.5); n3t[2]/sum(n3t)`

$n = 4$: `n4=expand.grid(1:6,1:6,1:6,1:6); n4s=apply(n4,1,mean); n4t=table(abs(n4s-3.5)>1.5); n4t[2]/sum(n4t)`

etc. (Bei mehr als so grossen Zahlen dauert die Berechnung dann lange oder kann ganz aussteigen. bei $n = 10$ produziert `expand.grid` eine Tabelle mit 6^{10} Zeilen.

9.7 Lernziele

- „Zufallsvariable“ definieren können und zwei Gründe für deren Einführung angeben können. Die Graphen von Zufallsvariablen hinschreiben können.
- Auf Grund eines auf einen Ergebnismenge gegebenen Wahrscheinlichkeitsmasses die Wahrscheinlichkeitsfunktion des Bildes der Zufallsvariable und Wahrscheinlichkeiten beliebiger Ereignisse des Bildes der Zufallsvariable berechnen können.
- Erwartungswert einer Zufallsvariable definieren und berechnen können. Die Sätze des Unterkapitels erläutern können (wenn diese gegeben sind).
- Varianz und Standardabweichung einer Zufallsvariable definieren und berechnen können. Die Sätze des Unterkapitels erläutern können (wenn diese gegeben sind).
- Die Idee der Varianz als Risikomass erläutern können.
- Die Idee der Unabhängigkeit von zwei Zufallsvariablen erläutern können.
- An Beispielen überprüfen können, ob Zufallsvariablen unabhängig sind oder nicht.

- Die Idee der „Zufallsvariablen auf das kartesische Produkt der Bilder von Zufallsvariablen“ verstehen und die Verteilung solcher Zufallsvariablen (= Statistiken) an einfachen Beispielen berechnen können.
- Für i.i.d. Zufallsvariablen X_i die Sätze

$$E_{P_{(X_1, \dots, X_n)}} \left(\sum_{i=1}^n X_i \right) = nE_P(X_i)$$

$$E_{P_{(X_1, \dots, X_n)}} (\bar{X}_n) = E_P(X_i)$$

$$V_{P_{(X_1, \dots, X_n)}} (\bar{X}_n) = \frac{1}{n} V_P(X_i)$$

erläutern können (wenn diese gegeben sind).

Den Satz $V(bX) = b^2 V(X)$ verstehen und erläutern können (wenn er gegeben ist).

- Die Idee des Schwachen Gesetzes der Grossen Zahlen am Würfelbeispiel erläutern können. Die Bedeutung des Schwachen Gesetzes der Grossen Zahlen für den Zusammenhang von Wahrscheinlichkeiten und relativen Häufigkeiten am Münzbeispiel erläutern können.
- Übungen der obigen Art lösen können.
- Die folgenden Symbole oder Ausdrücke in die folgende Tabelle einordnen können (die sich entsprechenden auf einer Zeile): $\bar{x}$, $V(X)$, $P(X = x_i)$, s , $E(X)$, s^2 , $h(x_i)$ (für relative Häufigkeit von x_i), Kenngrösse, Statistik :

empirische Begriffe oder Symbole	wahrscheinlichkeitstheoretische Begriffe oder Symbole
$\bar{x}$	$E(X)$
s^2	$V(X)$
$h(x_i)$	$P(X = x_i)$
s	$\sqrt{V(X)}$
Kenngrösse	Statistik

Tabelle 9.7.8: Übersicht über empirische und wahrscheinlichkeitstheoretische Symbole

Kapitel 10

Diskrete Verteilungsfamilien

Viele Wahrscheinlichkeitsverteilungen sind zu anderen ähnlich. Sie bilden sogenannte Verteilungsfamilien. Wir betrachten einige dieser Verteilungsfamilien, wobei wir künftig auch Bildmasse „Wahrscheinlichkeitsverteilungen“ nennen - faktisch werden wir es fast nur mehr mit Bildmassen zu tun haben:

10.1 Familie der Bernoulli-Verteilungen

Definition 10.1.1. Die Zufallsvariable X ist bernoulliverteilt (Jakob I. Bernoulli, 1655 - 1705, Schweizer Mathematiker und Physiker) genau dann, wenn

$$\begin{aligned}\mathfrak{X} &= \{0, 1\} \\ P_X(\{1\}) &= p \\ P_X(\{0\}) &= 1 - p\end{aligned}$$

$$(0 \leq p \leq 1)$$

$\mathfrak{X}$ = Bild von X ; p ist der Parameter der Verteilungsfamilie)

$X \sim B(1, p)$ ist eine Abkürzung für „die Zufallsvariable X ist bernoulliverteilt mit dem Parameter p “.

Damit gilt für eine Bernoulli-Verteilung P_X :

Graph von $f_X = \{(1, p), (0, 1 - p)\}$

Graph von $P_X = \{(\emptyset, 0), (\{0\}, 1 - p), (\{1\}, p), (\{0, 1\}, 1)\}$

P_X ist die Bernoulli-Verteilung der Zufallsvariable X . Je nach dem p erhalten wir eine andere Bernoulli-Verteilung. Es gibt überabzählbar unendlich viele Bernoulli-Verteilungen, da p im Intervall $[0, 1]$ beliebig wählbar ist. Durch die obige Definition wird somit eine ganze *Familie* von Bernoulli-Verteilungen festgelegt. p nennen wir den Parameter der Verteilung oder der Verteilungsfamilie. Durch ihn ist eine spezifische Bernoulli-Verteilung festgelegt und damit charakterisiert.

Beachten Sie, dass der Definitionsbereich der Zufallsvariable X in der obigen Definition nicht mehr thematisiert wird. Wir interessieren uns nur mehr für das Bildmass von X . Der Definitionsbereich von X kann zwei Elemente oder auch mehr Elemente enthalten. Das Bild von X enthält jedoch genau zwei Elemente.

Beispiel 10.1.2. a) Wurf einer Münze: $\Omega = \{K, Z\}$;

Graph von $f = \{(K, 0.3), (Z, 0.7)\}$

Graph von $X = \{(K, 1), (Z, 0)\}$ (bijektive Zufallsvariable X);

Graph von $f_X = \{(1, 0.3), (0, 0.7)\}$

Die Bernoulli-Verteilung ist damit:

Graph von $P_X = \{(\emptyset, 0), (\{0\}, 1 - 0.3), (\{1\}, 0.3), (\{0, 1\}, 1)\}$

b) Probe auf Wurmstichigkeit von Äpfeln: $\Omega = \{W, NW\}$;

Graph von $f = \{(W, 0.02), (NW, 0.98)\}$

Graph von $X = \{(W, 1), (NW, 0)\}$ (bijektive Zufallsvariable X);

Graph von $f_X = \{(1, 0.02), (0, 0.98)\}$ (mit W für „wurmstichig“, NW für „nicht wurmstichig“)

Die Bernoulli-Verteilung ist damit:

Graph von $P_X = \{(\emptyset, 0), (\{0\}, 0.98), (\{1\}, 0.02), (\{0, 1\}, 1)\}$

c) Probe einer Warenlieferung: $\Omega = \{na, ka, a\}$; (mit na für „nicht akzeptierbar“, ka für „knapp akzeptierbar“, a für „gute Qualität“).

Graph von $f = \{(na, 0.05), (ka, 0.45), (a, 0.5)\}$

Graph von $X = \{(na, 0), (ka, 1), (a, 1)\}$ (nicht bijektive Zufallsvariable X);

Graph von $f_X = \{(1, 0.95), (0, 0.05)\}$

Die Bernoulli-Verteilung ist damit:

Graph von $P_X = \{(\emptyset, 0), (\{0\}, 0.05), (\{1\}, 0.95), (\{0, 1\}, 1)\}$ ◇

Damit die betrachtete Verteilung tatsächlich eine Wahrscheinlichkeitsverteilung ist, muss gelten:

Satz 10.1.3. $\sum_{k=0}^1 P(X = k) = 1$ (für $X \sim B(1, p)$)

Beweis. $\sum_{k=0}^1 P(X = k) = (1 - p) + p = 1$ □

Satz 10.1.4. Wenn $X \sim B(1, p)$, dann $E(X) = p$

Beweis. $E(X) = \sum_{k=0}^1 k \cdot P(X = k) = 0(1 - p) + 1p = p$ □

Satz 10.1.5. Wenn $X \sim B(1, p)$, dann $V(X) = p(1 - p)$

Beweis. $V(X) = \sum_{k=0}^1 P(X = k)(k - E(X))^2 =$
 $(1 - p)(0 - p)^2 + p(1 - p)^2 \underset{\text{durch Ausklammern}}{=} (1 - p)p(p + 1 - p) = p(1 - p)$ □

Damit wäre z.B. fürs Beispiel 10.1.2c) die Varianz: $0.05 \cdot 0.95 = 0.0475$

10.2 Familie der Binomialverteilungen

Wir führen die Problemlage mit einem Beispiel ein: Wir nehmen an, wir würden aus einer Urne mit 20 roten und 30 schwarzen Kugeln 10 Kugeln ziehen - mit Zurücklegen und ohne Reihenfolge. Es stellt sich die Frage, wie hoch die Wahrscheinlichkeit ist, 6 rote Kugeln zu erhalten. Die Wahrscheinlichkeit, in einem Zug eine rote Kugel zu erhalten, beträgt $\frac{2}{5} = 0.4$, in einem Zug eine schwarze Kugel zu erwischen $\frac{3}{5} = 0.6$. Wir betrachten zuerst eine konkrete Reihenfolge, die bei 6 roten und 4 schwarzen Kugeln möglich ist: $(r, s, r, r, s, s, s, r, r, r)$. Die Wahrscheinlichkeit, genau diese Reihenfolge zu erhalten, ist bei vorausgesetzter Unabhängigkeit

$$\begin{array}{cccccccccc} r & s & r & r & s & s & s & r & r & r \\ 0.4 \cdot & 0.6 \cdot & 0.4 \cdot & 0.4 \cdot & 0.6 \cdot & 0.6 \cdot & 0.6 \cdot & 0.4 \cdot & 0.4 \cdot & 0.4 \cdot \\ = & 0.4^6 \cdot 0.6^4 = 0.00053084. \end{array}$$

Jede solche Reihenfolge mit gleichvielen roten und schwarzen Kugeln hat dieselbe Wahrscheinlichkeit (Kommutativität der Multiplikation). Um die Wahrscheinlichkeit zu berechnen, 6 rote und

4 schwarze Kugeln zu erhalten, muss man wissen, wie viele solcher Reihenfolgen es gibt. Diese Aufgabe wurde in der Kombinatorik behandelt (s. Übung 6.2.5.18, S. 155). Es gibt

$$\binom{10}{6} = \binom{10}{4} = 210$$

solcher Reihenfolgen. Entsprechend beträgt die Wahrscheinlichkeit, 6 rote Kugeln zu erhalten,

$$\binom{10}{6} 0.4^6 \cdot 0.6^4 = 210 \cdot 0.00053084 = 0.1148.$$

Nach diesem Beispiel behandeln wir den Fall allgemeiner. Sei p die Wahrscheinlichkeit, das ausgezeichnete Ergebnis zu erhalten - im Beispiel waren das die roten Kugeln. Die Wahrscheinlichkeit, das ausgezeichnete Ergebnis zu erhalten, verändere sich zudem bei wiederholten identischen Zufallsexperimenten nicht. Wir möchten wissen, wie viel die Wahrscheinlichkeit beträgt, bei n Zügen x ausgezeichnete Ergebnisse zu erhalten. Ordnen wir dem einzelnen ausgezeichneten Ergebnis durch eine Zufallsvariable 1 zu und dem Eintreffen der anderen Ergebnisse 0, so ist x die Summe der Zahlen eines Zahlenvektors $y \in \{0, 1\}^n$. Die Wahrscheinlichkeit, in einem konkreten Zahlenvektor y genau x ausgezeichnete Ergebnisse zu erhalten, ist dann bei vorausgesetzter Unabhängigkeit der Zufallsvariablen

$$p^x (1 - p)^{n-x}.$$

Es gibt $\binom{n}{x}$ Zahlenvektoren (= n -Tupel) aus $\{0, 1\}^n$, welche eine Summe von x aufweisen. Entsprechend ist die Wahrscheinlichkeit, x ausgezeichnete Ergebnisse zu erhalten

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}.$$

Man beachte, dass X die Summe von n bernoulli-verteilten i.i.d. Zufallsvariablen ist, wie wir später noch eigens beweisen und mit Hilfe eines Beispiels durchrechnen werden. Die obigen Überlegungen führen uns zur folgenden Definition:

Definition 10.2.1. X ist eine binomialverteilte Zufallsvariable genau dann, wenn

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$$(x \in \mathbb{N}_n).$$

(Abkürzung: $X \sim B(n, p)$ für „die Zufallsvariable X ist binomialverteilt“)

n und p werden die Parameter der Binomialverteilung genannt. Durch sie ist eine Binomialverteilung bestimmt.

Wir können Wahrscheinlichkeitsfunktionen von Wahrscheinlichkeitsverteilungen auch graphisch darstellen: auf die x-Achse tragen wir die möglichen Werte der Zufallsvariable X ab (= Argumente der Wahrscheinlichkeitsfunktion). Auf der y-Achse tragen wir die Wahrscheinlichkeiten $P(X = x)$ ab. Achtung: die Punkte sollte nicht verbunden werden, da die Funktion auf $\mathbb{N}_n$ definiert ist und den Zwischenwerten (z.B. 0.8) durch die Funktion keine Werte zugeordnet werden (s. Abbildung 10.2.1).

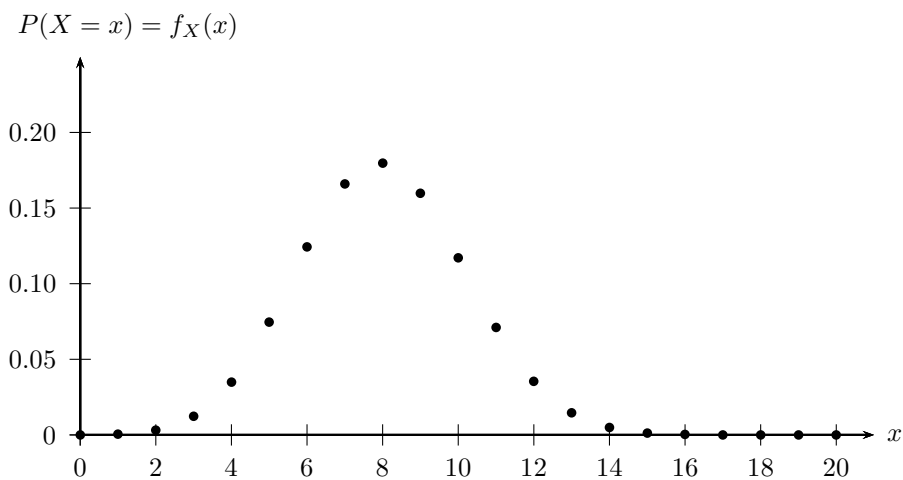


Abbildung 10.2.1: Graphische Darstellung der Wahrscheinlichkeitsfunktion einer binomial verteilten Zufallsvariable

Die Verteilungsfunktion einer binomialverteilten Zufallsvariable ist damit (für die graphische Darstellung s. Abbildung 10.2.2):

$$F(x) = P(X \leq x) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}$$

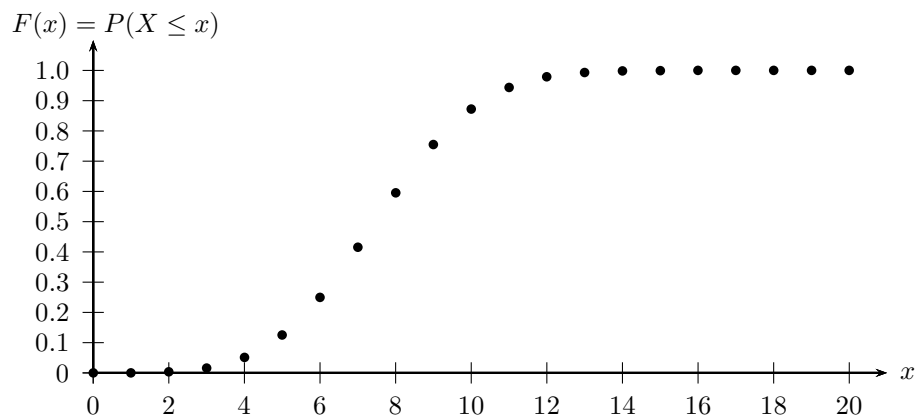


Abbildung 10.2.2: Graphische Darstellung der Verteilungsfunktion einer binomial verteilten Zufallsvariable $X \sim B(20, 0.4)$

Damit die definierte Verteilung eine Wahrscheinlichkeitsverteilung ist, muss gelten:

Satz 10.2.2. $\sum_{x=0}^n f_X(x) = \sum_{x=0}^n P(X=x) = \sum_{x=0}^n \binom{n}{x} p^x \cdot (1-p)^{n-x} = 1$

Beweis. $\sum_{x=0}^n \binom{n}{x} p^x \cdot (1-p)^{n-x} = (p + (1-p))^n = 1^n = 1$ (s. Kombinatorik, Satz 6.2.17, Erläuterungen zu den Binomialkoeffizienten). $\square$

Eine binomialverteilte Zufallsvariable X mit $X \sim B(n, p)$ ist identisch mit der Zufallsvariable „Summe“, definiert auf das kartesische Produkt der Bilder von n bernoulli-verteilten Zufallsvariablen. Dies zeigt sich einerseits am Eingangsbeispiel. Andererseits können wir das auch beweisen:

Satz 10.2.3. Für die Statistik „Summe“ $\sum_{i=1}^n X_i$ definiert als

$$\begin{aligned} \sum_{i=1}^n X_i &:= \times_{i=1}^n \mathfrak{X}_i \rightarrow \mathbb{N}_n \\ \sum_{i=1}^n X_i(x) &= \sum_{i=1}^n x_i \quad (x = (x_1, \dots, x_n)) \end{aligned}$$

mit

$$\text{unabhängigen } X_i \sim B(1, p) \quad (\text{für } i \in \mathbb{N}_n^*)$$

gilt:

$$\sum_{i=1}^n X_i \sim B(n, p)$$

Beweis. Wir beweisen durch Induktion:

(i) Der Satz gilt für $n = 2$. Sei $X_i \sim B(1, p)$. Das kartesische Produkt der Bilder der Zufallsvariablen X_i ist: $\mathfrak{X}_1 \times \mathfrak{X}_1 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Wegen der Unabhängigkeit der Zufallsvariablen gilt: Graph von

$$f_{(X_1, X_2)} = \{((0, 0), (1-p)^2), ((0, 1), p(1-p)), ((1, 0), p(1-p)), ((1, 1), p^2)\}.$$

Das Bild der Summe $X_1 + X_2$ ist: $\{0, 1, 2\}$ und der Graph von

$$f_{X_1+X_2} = \{(0, (1-p)^2), (1, 2 \cdot p(1-p)), (2, p^2)\}.$$
 Für die Binomialverteilung erhalten wir:

$$P(X = 0) = f_X(0) = \binom{2}{0} p^0 (1-p)^{2-0} = (1-p)^2.$$

$$P(X = 1) = f_X(1) = \binom{2}{1} p^1 (1-p)^{2-1} = 2 \cdot p(1-p)$$

$$P(X = 2) = f_X(2) = \binom{2}{2} p^2 (1-p)^{2-2} = p^2. \text{ Damit weist } X_1 + X_2 \text{ und } X \text{ dieselbe Verteilung auf, da sie in den Wahrscheinlichkeitsfunktionen übereinstimmen.}$$

(ii) Wir nehmen an, der Satz gelte für n , d.h. für die Summe $X_n^s := \sum_{i=1}^n X_i$ auf das kartesische Produkt der Bilder der $X_i \sim B(1, p)$, $i \in \mathbb{N}_n^*$, gilt $X_n^s \sim B(n, p)$. Wir versuchen zu zeigen, dass

für die Summe $X_{n+1}^s := \sum_{i=1}^{n+1} X_i$ auf das kartesische Produkt der Bilder der $X_i \sim B(1, p)$, $i \in \mathbb{N}_{n+1}^*$, gilt: $X_{n+1}^s \sim B(n+1, p)$. Sei $x \in \mathfrak{D}_{X_{n+1}^s} = \mathfrak{D}_{X_n^s} \cup \{n+1\}$ mit $\mathfrak{D}_{X_n^s} := \text{Bild}(X_n^s)$. Zu berechnen ist:

$$P(X_{n+1}^s = x).$$

Eine spezifische Summe $x \in \mathbb{N}_{n+1}^*$ kann dabei durch zwei Arten zustande kommen:

(i) $X_n^s = x'$ und der Wert der Zufallsvariable X_{n+1} ist 1. Damit gilt $x = x' + 1$ oder

(ii) $X_n^s = x$ und der Wert der Zufallsvariable X_{n+1} ist 0, womit gilt $x = x + 0$.

Wir erhalten wegen der Unabhängigkeit der Zufallsvariablen:

$$P(X_{n+1}^s = x) = f_{X_n^s}(x') \cdot f_{X_{n+1}}(1) + f_{X_n^s}(x) \cdot f_{X_{n+1}}(0)$$

Zudem gilt laut Induktionsvoraussetzung $P(X_n^s = x') = f_{X_n^s}(x') = \binom{n}{x'} p^{x'} (1-p)^{n-x'}$ sowie $P(X_n^s = x) = \binom{n}{x} p^x (1-p)^{n-x}$.

Damit gilt dann durch Einsetzen: $P(X_{n+1}^s = x) =$

$$\binom{n}{x'} p^{x'} (1-p)^{n-x'} \cdot p + \binom{n}{x} p^x (1-p)^{n-x} \cdot (1-p) \stackrel{x=x'+1}{=}$$

$$\binom{n}{x} p^x (1-p)^{n+1-x} + \binom{n}{x} p^x (1-p)^{n+1-x} =$$

$\left(\binom{n}{x'} + \binom{n}{x'+1}\right) p^x (1-p)^{n+1-x} = \binom{n+1}{x'+1} p^x (1-p)^{n+1-x} =$
 $\binom{n+1}{x} p^x (1-p)^{n+1-x}$ (vorletzter Schritt mit Satz 6.2.14, Kombinatorik)
 Damit gilt: $X_{n+1}^s \sim B(n+1, p)$.

(iii) Aus (ii) und (i) folgt durch Induktion der Satz. $\square$

Beispiel 10.2.4. Mit einem weiteren Beispiel wollen wir diese Betrachtungsweise nochmals verdeutlichen.

Für $X_i \sim B(1, 0.4)$ und $i \in \mathbb{N}_3^*$ erhalten wir als kartesisches Produkt der Bilder: $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}$. X ist nun auf dieses Bild definiert, indem den Elementen von $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3$ jeweils die Summe der Komponenten zugeordnet wird. Der Graph von X ist damit:

$\{((0, 0, 0), 0), ((1, 0, 0), 1), ((0, 1, 0), 1), ((0, 0, 1), 1), ((1, 1, 0), 2),$
 $((1, 0, 1), 2), ((0, 1, 1), 2), ((1, 1, 1), 3)\}$. Das Bild von X ist damit $\mathfrak{D}_X = \{0, 1, 2, 3\}$.

Als nächstes wollen wir nun die Wahrscheinlichkeitsverteilung von X mit Hilfe des Graphen der Wahrscheinlichkeitsfunktion angeben (Unabhängigkeit der X_i vorausgesetzt):

Graph von $f_{(X_1, X_2, X_3)}$:
 $\{((0, 0, 0), 0.6^3), ((1, 0, 0), 0.4 \cdot 0.6^2), ((0, 1, 0), 0.4 \cdot 0.6^2), ((0, 0, 1), 0.4 \cdot 0.6^2),$
 $((1, 1, 0), 0.4^2 \cdot 0.6), ((1, 0, 1), 0.4^2 \cdot 0.6), ((0, 1, 1), 0.4^2 \cdot 0.6), ((1, 1, 1), 0.4^3)\}$. Für die Verteilung von X erhalten wir damit: Graph von $f_X = \{(0, 0.6^3), (1, 3 \cdot 0.4 \cdot 0.6^2), (2, 3 \cdot 0.4^2 \cdot 0.6), (3, 0.4^3)\} =$
 $\{(0, 0.216), (1, 0.432), (2, 0.288), (3, 0.064)\}$. Dies ist der Graph der Wahrscheinlichkeitsfunktion von $X \sim B(3, 0.5)$, wie Sie nachrechnen können. $\diamond$

Mit Hilfe der Tatsache, dass eine binomialverteilte Zufallsvariable die Summe von i.i.d. bernoulli-verteilten Zufallsvariablen ist, können wir den Beweis für die Formeln des Erwartungswertes und der Varianz binomialverteilter Zufallsvariablen sehr einfach führen.

Satz 10.2.5. Für $X \sim B(n, p)$ gilt: $E(X) = n \cdot p$.
 (d.h. der Erwartungswert einer binomialverteilten Zufallsvariable beträgt „ np “)

Beweis. Da $X \sim B(n, p)$ die Summe von n unabhängigen Zufallsvariablen

$X_i \sim B(1, p)$ ist, gilt $E\left(\sum_{i=1}^n X_i\right) \stackrel{\text{Satz 9.5.17}}{=} nE(X_i) \stackrel{\text{Satz 10.1.4}}{=} np$. $\square$

Satz 10.2.6. Für $X \sim B(n, p)$ gilt: $V(X) = n \cdot p \cdot (1 - p)$.

Beweis. Da $X \sim B(n, p)$ die Summe von n unabhängigen Zufallsvariablen $X_i \sim B(1, p)$ ist, gilt

$V\left(\sum_{i=1}^n X_i\right) \stackrel{\text{Satz 9.5.35}}{=} nV(X_i) \stackrel{\text{Satz 10.1.5}}{=} np(1 - p)$. $\square$

Man könnte diese zwei Sätze auch durch Einsetzen in die Definition des Erwartungswertes und der Varianz herleiten, wobei dann allerdings viel Summenzeichenjonglieren nötig ist.

Beispiel 10.2.7. Eine statistische Anwendung: Aus einer Urne mit roten und schwarzen Kugeln kann auf einmal nur eine Kugel herausgenommen werden. Diese wird nach der Entnahme wieder zurückgelegt. Wir möchten schätzen, wieviel der Anteil der roten Kugeln in der Urne ist. Dazu ziehen wir n Kugeln (mit Zurücklegen) und zählen die Anzahl x der roten Kugeln. Die Anzahl x der roten Kugeln ist Wert der Zufallsvariable Summe X und $\frac{x}{n}$ ist der Wert der Statistik Mittelwert $\bar{X}_n$. Der Erwartungswert der Statistik Summe fällt mit np zusammen (s. oben, Satz 10.2.5). Der Wert der Statistik Mittelwert ist $E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$. Wir können damit den Anteil p durch $\frac{x}{n}$ schätzen. Je grösser n , desto präziser ist die Schätzung, da $V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{1}{n^2}np(1 - p) = \frac{p(1-p)}{n}$ und $\lim_{n \rightarrow \infty} \frac{p(1-p)}{n} = 0$. Konkret: Wir nehmen an, wir hätten $n = 200$ Züge gemacht. Der Anteil der roten Kugeln sei 60 gewesen. p wird dann geschätzt

durch $\frac{60}{200} = \frac{3}{10}$ und wir schreiben $\hat{p} = \frac{3}{10}$ - geschätzte Werte werden oft mit einem Hütchen versehen. Die geschätzte Varianz der Statistik Mittelwert ist $\frac{\frac{3}{10}(1-\frac{3}{10})}{200} = \frac{21}{20\,000} = 0.001\,05$. Damit liegt eine recht präzise Schätzung vor. $\diamond$

10.3 Familie der Poisson-Verteilungen

Während die bisherigen Verteilungsfamilien jeweils bezüglich der Bildbereiche von Zufallsvariablen mit endlich vielen Werten definiert sind, weisen poissonverteilte Zufallsvariablen einen unendlichen Bildbereich auf (Siméon Denis Poisson, 1781-1840, französischer Physiker und Mathematiker). Die Verteilung ist trotzdem diskret, da der Definitionsbereich der Wahrscheinlichkeitsfunktion abzählbar unendlich ist. Im vorliegenden Fall ist er mit der Menge der natürlichen Zahlen identisch.

Bei der Poisson-Verteilungen geht es um *pro Einheit* seltene Ereignisse (Ereignisse, die pro Einheit selten sind, können pro andere, grössere Einheit häufig vorkommen). Im Allgemeinen geht es um *pro Zeitperiode* seltene Ereignisse. Die Poisson-Verteilung kann aber auch verwendet werden, um das seltene Auftreten von Ereignissen auf Flächeneinheiten oder in Raumeinheiten zu beschreiben.

Beispiel 10.3.1. *Wahrscheinlichkeit einer spezifischen Anzahl von Autounfällen in einem Quartier pro Jahr; Wahrscheinlichkeit einer spezifischen Anzahl von Maschinenstillständen pro Tag. Wahrscheinlichkeit einer Anzahl von Betriebsunfällen pro Jahr. Wahrscheinlichkeit einer Anzahl von defekten Geräten pro Jahr; Wahrscheinlichkeit einer Anzahl von Klienten an einem Schalter pro Zeiteinheit; Wahrscheinlichkeit einer Anzahl von Keimen pro Quadratmeter, einer Anzahl von Asbestfasern pro dm^3 .* $\diamond$

Definition 10.3.2. Eine Zufallsvariable X ist poisson-verteilt genau dann, wenn

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

($x \in \mathbb{N}$, $\lambda > 0$).

(Abkürzung: $X \sim P(\lambda)$)

(„ λ “ wird als „Lamda“ gelesen. λ ist der Parameter der Verteilung).

Wir stellen für $X \sim P(5)$ die Wahrscheinlichkeitsfunktion graphisch dar, wobei wir uns auf den Teil $[0, 24] \cap \mathbb{N}$ des Definitionsbereichs beschränken (s. Abbildung 10.3.3):

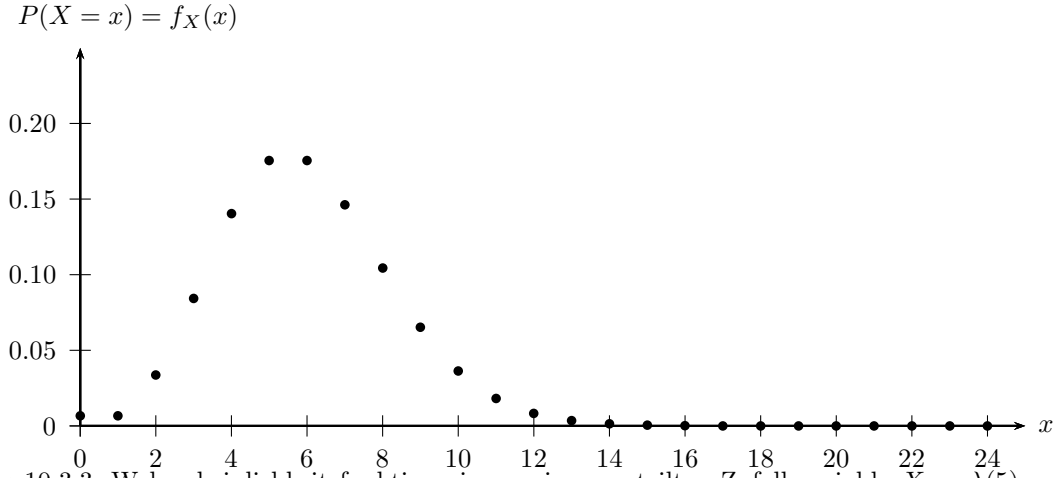


Abbildung 10.3.3: Wahrscheinlichkeitsfunktion einer poissonverteilten Zufallsvariable $X \sim \lambda(5)$ im Bereich $[0, 24]$

Die Poisson-Verteilungsfunktion wird beschrieben durch:

$$F(x) = \sum_{k=0}^x \frac{\lambda^k}{k!} e^{-\lambda}$$

Bei der folgenden graphischen Darstellung (s. Abbildung 10.3.4) beschränken wir uns wiederum auf den interessanten Bereich (von 25 weg ist $F(x)$ näherungsweise 1).

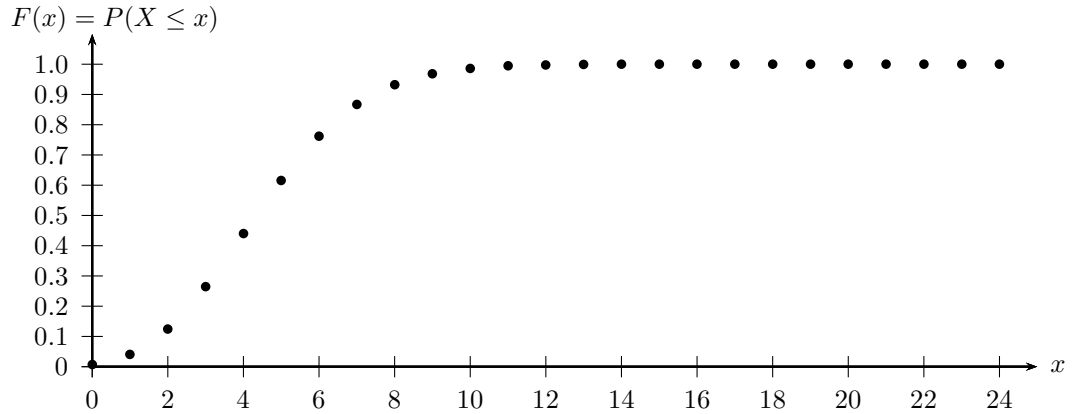


Abbildung 10.3.4: Graphische Darstellung der Verteilungsfunktion einer poissonverteilten Zufallsvariable $X \sim \lambda(5)$ im Bereich $[0, 24]$

Um nachzuweisen, dass durch die obige Definition eine Wahrscheinlichkeitsverteilung festgelegt ist, muss man zeigen, dass

Satz 10.3.3. $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1.$

Beweis. Auf Grund der Definition von e^{λ} gilt: $e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$

Damit wird: $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = e^{\lambda-\lambda} = e^0 = 1$

Somit ist die durch $P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$ definierte Verteilung tatsächlich eine Wahrscheinlichkeitsverteilung. $\square$

Der Erwartungswert einer poissonverteilten Zufallsvariable ist ihr Parameter:

Satz 10.3.4. Wenn $X \sim P(\lambda)$, dann $E(X) = \lambda$

Beweis. $E(X) = \sum_{k=0}^{\infty} \left(k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \right) = \sum_{k=1}^{\infty} \left(k \frac{\lambda^k}{k(k-1)!} e^{-\lambda} \right) = \sum_{k=1}^{\infty} \left(\frac{\lambda^k}{(k-1)!} e^{-\lambda} \right) =$
 $\sum_{k=1}^{\infty} \left(\frac{\lambda \lambda^{k-1}}{(k-1)!} e^{-\lambda} \right) = \lambda \sum_{k=1}^{\infty} \left(\frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \right) = \lambda \sum_{k=0}^{\infty} \left(\frac{\lambda^k}{k!} e^{-\lambda} \right) = \lambda,$
 da wie oben gezeigt $\sum_{k=0}^{\infty} \left(\frac{\lambda^k}{k!} e^{-\lambda} \right) = 1$. $\square$

Der Erhöhung der unteren Summationsgrenze im zweiten Schritt ist gerechtfertigt, da für $k = 0$ der Ausdruck 0 wird. Die Tiefersetzung der Summationsgrenze im vorletzten Schritt ist gerechtfertigt, da es auf dasselbe herauskommt, ob wir mit $k - 1$ bei 1 anfangen oder mit k bei 0.

Satz 10.3.5. Wenn $X \sim P(\lambda)$, dann $V(X) = \lambda$

Beweis. $V(X) = \sum_{k=0}^{\infty} \left(\frac{\lambda^k}{k!} e^{-\lambda} (k - \lambda)^2 \right) = \sum_{k=0}^{\infty} \left(\frac{\lambda^k}{k!} e^{-\lambda} (k^2 - 2\lambda k + \lambda^2) \right) =$
 $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} k^2 - \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} 2\lambda k + \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \lambda^2 =$
 $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} k^2 - \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} 2\lambda k + \lambda^2 = (\lambda^2 + \lambda) - 2\lambda^2 + \lambda^2 = \lambda$ denn:
 $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} k^2 = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} k k = \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} e^{-\lambda} k = \sum_{k=1}^{\infty} \frac{\lambda \lambda^{k-1}}{(k-1)!} e^{-\lambda} k =$
 $\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} k = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} ((k-1) + 1) =$
 $\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} (k-1) + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} 1 = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda =$
 $\lambda^2 + \lambda$
 $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} 2\lambda k = 2\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} k = 2\lambda^2$ (da $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} k = \lambda$, s. Beweis von 10.3.4)
 $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \lambda^2 = \lambda^2$ da $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1$ (s. Beweis von 10.3.4) $\square$

λ ist, wie gezeigt, der Erwartungswert einer Poisson-Verteilung. Haben wir k Häufigkeiten pro Einheit, so können wir diese Häufigkeiten bei fehlender gegenseitiger Beeinflussung der Wahrscheinlichkeiten als Werte von *i.i.d.* poissonverteilten Zufallsvariable X_i betrachten. Bilden wir den Durchschnitt der Werte dieser Zufallsvariablen X_i , so erhalten wir einen konkreten Wert der Statistik Mittelwert $\bar{X}$. Da mit Satz 9.5.18 und 10.3.4

$$E(\bar{X}) = E(X_i) = \lambda,$$

können wir mit Hilfe dieses Durchschnitts λ schätzen: wir setzen das geschätzte λ in die Formel $\frac{\lambda^k}{k!} e^{-\lambda}$ und erhalten: $\frac{\hat{\lambda}^k}{k!} e^{-\hat{\lambda}}$ (geschätzte Größen werden jeweils mit einem Hütchen versehen). Wegen $V(\bar{X}) = \frac{1}{n} V(X_i) = \frac{\lambda}{n}$ nimmt die Präzision dieses Schätzers mit steigendem n zu. Man erhält also eine aus den Daten geschätzte Wahrscheinlichkeitsfunktion. Diese Formel können wir nun verwenden, um z.B. die Wahrscheinlichkeit zu berechnen, dass $x = 0, 1, 2, 3, 4, 5, 6, 7$, etc. Unfälle pro Jahr, wenn die Werte von X_i die Anzahl der Unfälle des i -ten Jahres sind. Dies ist erlaubt, sofern die geschätzte Poissonverteilung in etwa den empirischen Befunden entspricht. Die Angemessenheit eines Modells können wir etwas handgestrickt graphisch und von Auge überprüfen, indem wir die relativen Häufigkeiten und die Werte der geschätzten Wahrscheinlichkeitsfunktion in dasselbe Koordinatennetz zeichnen (später werden wir rechnerische Methoden entwickeln).

Übung 10.3.6. In einem Betrieb wurden folgende Unfallhäufigkeiten pro Jahr beobachtet: In einem Jahr gibt es keinen Unfall, in 2 Jahren gibt es einen Unfall, in 5 Jahren gibt es zwei Unfälle, in 4 Jahren gibt es drei Unfälle, in 3 Jahren gibt es vier Unfälle, in 2 Jahren gibt es 5 Unfälle, in einem Jahr gibt es 6 Unfälle. In keinem Jahr gibt es mehr Unfälle. Insgesamt wurden 18 Jahre untersucht. Die Wahrscheinlichkeit, dass sich in einem Jahr x Unfälle ergeben, kann sinnvoller Weise mit der relativen Häufigkeit der Jahre mit x Unfällen geschätzt werden. Dies ergibt die folgende relative Häufigkeitsverteilung (s. Tabelle 10.3.1 und Abbildung 10.3.5):

Anzahl Jahre mit x Unfällen	Anzahl Unfälle pro Jahr ($= x$)	relative Häufigkeiten (der Jahre mit x Unfällen)	Anzahl Jahre · Anzahl Unfälle	Werte der Wahrscheinlichkeitsfunktion $f_X; X \sim P(\hat{\lambda})$ $\hat{\lambda} = 2.88888$
1	0	0.055555556	0	0.055637998
2	1	0.111111111	2	0.160731995
5	2	0.277777778	10	0.232168437
4	3	0.222222222	12	0.223569606
3	4	0.166666667	12	0.161466938
2	5	0.111111111	10	0.093292009
1	6	0.055555556	6	0.044918374
Σ 18		1	52	0.971785358

Tabelle 10.3.1: Tabelle für die Berechnung einer geschätzten Poisson-Wahrscheinlichkeitsfunktion mit Hilfe einer absoluten Häufigkeitsverteilung ($\hat{\lambda} = 52/18 = 2.\bar{8} =$ durchschnittliche Anzahl Unfälle pro Jahr)

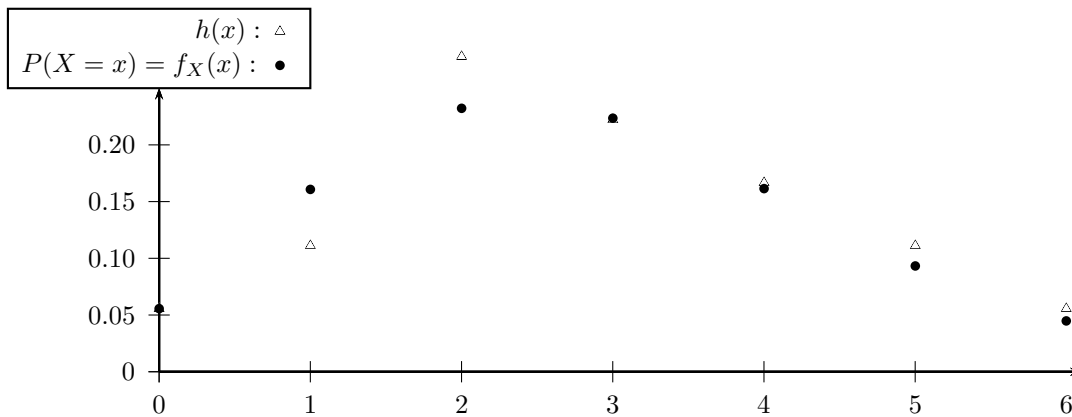


Abbildung 10.3.5: Graphische Darstellung der relativen Häufigkeiten und der geschätzten Wahrscheinlichkeitsfunktion (Poisson-Verteilung; geschätztes $\lambda = \hat{\lambda} = 52/18 = 2.88888$)

Vermutlich sind die Daten poisson-verteilt. Wir können damit die entsprechende Poisson-verteilung als Modell verwenden und z.B. die Wahrscheinlichkeit dafür berechnen, dass es sieben oder mehr Unfälle pro Jahr gibt: $P(X \geq 7) = P(X > 6) = 1 - P(X \leq 6) = 1 - \sum_{k=0}^6 \frac{2.888888^k}{k!} e^{-2.88888888} = 0.028215$.

Beispiel 10.3.7. In Dorf A hat es über die letzten Jahre durchschnittlich 3 Autounfälle pro Jahr gegeben. Wir setzen voraus, dass die Anzahl Unfälle pro Jahr poissonverteilt sind. Wir berechnen

die folgenden Werte der Poisson-Verteilung:

Die Wahrscheinlichkeit, dass pro Jahr Jahr x Unfälle passieren, beträgt:

$$P(X = 0) = \frac{3^0}{0!}e^{-3} = 0.049787 \quad (\text{Wahrscheinlichkeit, dass kein Unfall passiert})$$

$$P(X = 1) = \frac{3^1}{1!}e^{-3} = 0.14936 \quad (\text{Wahrscheinlichkeit, dass ein Unfall passiert})$$

$$P(X = 2) = \frac{3^2}{2!}e^{-3} = 0.22404 \quad (\text{Wahrscheinlichkeit, dass zwei Unfälle passieren})$$

$$P(X = 3) = \frac{3^3}{3!}e^{-3} = 0.22404 \quad (\text{Wahrscheinlichkeit, dass drei Unfälle passieren})$$

$$P(X = 4) = \frac{3^4}{4!}e^{-3} = 0.16803 \quad (\text{Wahrscheinlichkeit, dass vier Unfälle passieren})$$

$$P(X = 5) = \frac{3^5}{5!}e^{-3} = 0.10082 \quad (\text{Wahrscheinlichkeit, dass fünf Unfälle passieren}).$$

Die Wahrscheinlichkeit, dass im nächsten Jahr mehr als 5 Unfälle passieren, ist

$$1 - P(X \leq 5) = 1 - \left(\frac{3^0}{0!}e^{-3} + \frac{3^1}{1!}e^{-3} + \frac{3^2}{2!}e^{-3} + \frac{3^3}{3!}e^{-3} + \frac{3^4}{4!}e^{-3} + \frac{3^5}{5!}e^{-3} \right) = 0.083918. \quad \diamond$$

Beispiel 10.3.8. Ein Unternehmen fertigt täglich 2'000'000 Transistoren, darunter täglich durchschnittlich 5 defekte. Wie gross ist die Wahrscheinlichkeit, dass an einem bestimmten Tag mehr als 3 defekte Transistoren produziert werden? dass genau ein defekter Transistor produziert wird? (Poissonverteilung vorausgesetzt).

$\lambda = 5$.

$$P(X \leq 3) = \sum_{k=0}^3 \frac{5^k}{k!}e^{-5} = \frac{5^0}{0!}e^{-5} + \frac{5^1}{1!}e^{-5} + \frac{5^2}{2!}e^{-5} + \frac{5^3}{3!}e^{-5} = 0.26503$$

$$P(X > 3) = 1 - \left(\sum_{k=0}^3 \frac{5^k}{k!}e^{-5} \right) = 0.73497$$

$$P(X = 1) = \frac{5^1}{1!}e^{-5} = 0.033690.$$

$$P(X < 2) = \sum_{k=0}^1 \frac{5^k}{k!}e^{-5} = \frac{5^0}{0!}e^{-5} + \frac{5^1}{1!}e^{-5} = 4.0428 \times 10^{-2}$$

$$P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3) = 1 - 0.26503 = 0.73497 \quad \diamond$$

Satz 10.3.9. Sei $X_1 \sim P(\lambda_1)$ und $X_2 \sim P(\lambda_2)$ sowie unabhängig und die Statistik „Summe“ X der Zufallsvariablen X_1 und X_2 :

$$X : \mathfrak{X}_1 \times \mathfrak{X}_2 \rightarrow \mathbb{N}$$

$$X((n, m)) := n + m.$$

Dann gilt: $X \sim P(\lambda_1 + \lambda_2)$.

$$\text{Beweis. } P(X = x) = \sum_{k=0}^x P(X_1 = k \text{ und } X_2 = x - k) = \sum_{k=0}^x \left(\frac{\lambda_1^k}{k!} e^{-\lambda_1} \cdot \frac{\lambda_2^{(x-k)}}{(x-k)!} e^{-\lambda_2} \right) =$$

$$\sum_{k=0}^x \left(\frac{\lambda_1^k}{k!} \frac{\lambda_2^{(x-k)}}{(x-k)!} e^{-(\lambda_1 + \lambda_2)} \right)$$

Wir erweitern mit $x!$ und $(\lambda_1 + \lambda_2)^x$:

$$\sum_{k=0}^x \left(\frac{x! \lambda_1^k}{x! k!} \frac{\lambda_2^{(x-k)}}{(x-k)! (\lambda_1 + \lambda_2)^x} (\lambda_1 + \lambda_2)^x e^{-(\lambda_1 + \lambda_2)} \right) \stackrel{a^x = a^{x-k} a^k}{=} =$$

$$\sum_{k=0}^x \left(\frac{x!}{k! (x-k)!} \frac{\lambda_1^k \lambda_2^{(x-k)}}{(\lambda_1 + \lambda_2)^k (\lambda_1 + \lambda_2)^{(x-k)}} \frac{(\lambda_1 + \lambda_2)^x}{x!} e^{-(\lambda_1 + \lambda_2)} \right) =$$

$$\left(\sum_{k=0}^x \binom{x}{k} \left(\frac{\lambda_1}{(\lambda_1 + \lambda_2)^k} \right) \left(\frac{\lambda_2^{(x-k)}}{(\lambda_1 + \lambda_2)^{(x-k)}} \right) \right) \frac{(\lambda_1 + \lambda_2)^x}{x!} e^{-(\lambda_1 + \lambda_2)} =$$

$$\left(\sum_{k=0}^x \binom{x}{k} \left(\frac{\lambda_1}{(\lambda_1 + \lambda_2)} \right)^k \left(\frac{\lambda_2}{(\lambda_1 + \lambda_2)} \right)^{x-k} \right) \frac{(\lambda_1 + \lambda_2)^x}{x!} e^{-(\lambda_1 + \lambda_2)}$$

$$\text{Da } \frac{\lambda_1}{(\lambda_1 + \lambda_2)} + \frac{\lambda_2}{(\lambda_1 + \lambda_2)} = 1, \text{ gilt } \frac{\lambda_2}{(\lambda_1 + \lambda_2)} = 1 - \frac{\lambda_1}{(\lambda_1 + \lambda_2)}$$

Somit gilt:

$$\sum_{k=0}^x \binom{x}{k} \left(\frac{\lambda_1}{(\lambda_1 + \lambda_2)} \right)^k \left(\frac{\lambda_2}{(\lambda_1 + \lambda_2)} \right)^{x-k} =$$

$$\sum_{k=0}^x \binom{x}{k} \left(\frac{\lambda_1}{(\lambda_1 + \lambda_2)} \right)^k \left(1 - \frac{\lambda_1}{(\lambda_1 + \lambda_2)} \right)^{x-k} = 1 \quad (\text{Summe der Wahrscheinlichkeiten einer Binomialverteilung})$$

lung)

Wir erhalten: $P(X = x) = \frac{(\lambda_1 + \lambda_2)^x}{x!} e^{-(\lambda_1 + \lambda_2)}$. □

Aus dem obigen Satz folgt speziell:

Satz 10.3.10. Für die Summe X_n von n unabhängigen, poissonverteilten Zufallsvariablen $X_i \sim P(\lambda)$ gilt $X_n \sim P(n\lambda)$.

Beweis. Wir beweisen durch Induktion.

(i) Der Satz gilt für $n = 2$. Durch die Anwendung des obigen Satzes gilt nämlich: $X_2 \sim P(\lambda + \lambda) = P(2\lambda)$.

(ii) Der Satz gelte für n . Damit gilt $X_n \sim P(n\lambda)$. Laut obigem Satz gilt nun aber $X_{n+1} \sim P(n\lambda + \lambda) = P((n+1)\lambda)$.

(iii) Damit gilt der Satz durch Induktion. □

Gilt $\lambda = 1$, folgt aus dem Satz 10.3.10: Für die Summe X_n von n unabhängigen, poissonverteilten Zufallsvariablen $X_i \sim P(1)$ gilt $X_n \sim P(n)$.

Man kann zeigen, dass für kleines p die Binomialverteilung durch die Poissonverteilung approximiert werden kann, indem man $\lambda = np$ setzt. Die unten stehende Herleitung der Poissonverteilung aus der Binomialverteilung stellt eine Begründung dafür dar. In den Übungen werden zwei bestätigende Beispiele betrachtet.

Herleitung der Formel für die Poisson-Verteilung (fakultativ)

Die Poisson-Verteilung ist die Formel, die wir aus der Binomial-Verteilung erhalten, wenn wir $p = \frac{\lambda}{n}$ setzen ($p = \frac{\lambda}{n}$ folgt aus der Gleichsetzung der Erwartungswerte $\lambda = np$) und n in

$$\binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

gegen Unendlich streben lassen.

Wir setzen in die Formel für die Binomialverteilung $\frac{\lambda}{n}$ für p ein und wandeln um:

$$\begin{aligned} \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} &= \frac{n!}{(n-k)!k!} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \\ &= \frac{n!}{(n-k)!n^k} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

Jetzt wird der Grenzwert dieser Formel gebildet. Wir nehmen zwei Teilanalysen vor:

A)

$$\lim_{n \rightarrow \infty} \left(\frac{n!}{(n-k)!n^k} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \right) = 1,$$

da

$$\begin{aligned}
\frac{n!}{(n-k)!n^k} \cdot \frac{1}{(1-\frac{\lambda}{n})^k} &= \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-(k+1))}{n^k} \cdot \frac{1}{(1-\frac{\lambda}{n})^k} \\
&= \frac{n}{n(1-\frac{\lambda}{n})} \cdot \frac{n-1}{n(1-\frac{\lambda}{n})} \cdot \frac{n-2}{n(1-\frac{\lambda}{n})} \cdot \dots \cdot \frac{n-(k+1)}{n(1-\frac{\lambda}{n})} \\
&= \frac{n}{(n-\lambda)} \cdot \frac{n-1}{(n-\lambda)} \cdot \frac{n-2}{(n-\lambda)} \cdot \dots \cdot \frac{(n-(k+1))}{(n-\lambda)} \\
&= \frac{\frac{n}{n}}{(\frac{n}{n}-\frac{\lambda}{n})} \cdot \frac{\frac{n-1}{n}}{(\frac{n}{n}-\frac{\lambda}{n})} \cdot \frac{\frac{n-2}{n}}{(\frac{n}{n}-\frac{\lambda}{n})} \cdot \dots \cdot \frac{\frac{n-(k+1)}{n}}{(\frac{n}{n}-\frac{\lambda}{n})} \\
&= \frac{1}{(1-\frac{\lambda}{n})} \cdot \frac{1-\frac{1}{n}}{(1-\frac{\lambda}{n})} \cdot \frac{1-\frac{2}{n}}{(1-\frac{\lambda}{n})} \cdot \dots \cdot \frac{1-\frac{(k+1)}{n}}{(1-\frac{\lambda}{n})}.
\end{aligned}$$

Nun ist

$$\lim_{n \rightarrow \infty} \frac{1 - \frac{m}{n}}{(1 - \frac{\lambda}{n})} = 1.$$

(Geht n gegen unendlich, geht $\frac{m}{n}$ und $\frac{\lambda}{n}$ gegen 0).

Somit gilt:

$$\begin{aligned}
\lim_{n \rightarrow \infty} \left(\frac{1}{(1-\frac{\lambda}{n})} \cdot \frac{1-\frac{1}{n}}{(1-\frac{\lambda}{n})} \cdot \frac{1-\frac{2}{n}}{(1-\frac{\lambda}{n})} \cdot \dots \cdot \frac{1-\frac{(k+1)}{n}}{(1-\frac{\lambda}{n})} \right) &= 1 \cdot 1 \cdot 1 \cdot \dots \cdot 1 \\
&= 1
\end{aligned}$$

B)

$$\lim_{n \rightarrow \infty} \left(\frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \right) = \frac{\lambda^k}{k!} \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n$$

Es gilt:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Somit gilt:

$$\lim_{n \rightarrow \infty} \left(\frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \right) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Aus A) und B) folgt:

$$\lim_{n \rightarrow \infty} \left(\binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \right) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Umgekehrt erhält man die Binomialverteilung aus der Poissonverteilung, indem man zwei unabhängige poisson-verteilte Zufallsvariablen mit fixer Summe betrachtet. Analog zu

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

gilt

$$P_X(A|B) = \frac{P_X(A \cap B)}{P_X(B)}$$

und mit der Schreibweise $P(X \in A) := P_X(A)$

$$P(X \in A | X \in B) := P_X(A|B) = \frac{P(X \in A \text{ und } X \in B)}{P(X \in B)} := \frac{P_X(A \cap B)}{P_X(B)}$$

Seien nun zwei unabhängige poisson-verteilte ZV N_i gegeben, mit den Erwartungswerten λ_i . Dann ist die Wahrscheinlichkeit, dass die Werte n_i angenommen werden gegeben durch

$$P(N_1 = n_1 \text{ und } N_2 = n_2) = e^{-\lambda_1} \frac{\lambda_1^{n_1}}{n_1!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n_2}}{n_2!}$$

Die Wahrscheinlichkeit, dass $N_1 + N_2$ den Wert n annimmt, ist mit $\lambda = \lambda_1 + \lambda_2$

$$P(N_1 + N_2 = n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

Sei nun n fix mit fixem Vektor (n_1, n_2) , so dass $n_1 + n_2 = n$. Damit ist

$$\{(n_1, n_2)\} \cap \{(n_{1j}, n_{2j}) \mid n_{1j} + n_{2j} = n\} = \{(n_1, n_2)\}$$

und $P(N_1 = n_1 \text{ und } N_2 = n_2 \text{ und } N_1 + N_2 = n) = P(N_1 = n_1 \text{ und } N_2 = n_2)$. Man erhält also

$$\begin{aligned} P(N_1 = n_1 \text{ und } N_2 = n_2 \mid N_1 + N_2 = n) &= \frac{P(N_1 = n_1 \text{ und } N_2 = n_2 \text{ und } N_1 + N_2 = n)}{P(N_1 + N_2 = n)} \\ &= \frac{P(N_1 = n_1 \text{ und } N_2 = n_2)}{P(N_1 + N_2 = n)} \\ &= \frac{e^{-\lambda_1} \frac{\lambda_1^{n_1}}{n_1!} \cdot e^{-\lambda_2} \frac{\lambda_2^{n_2}}{n_2!}}{e^{-\lambda} \frac{\lambda^n}{n!}} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1^{n_1}}{n_1!} \cdot \frac{\lambda_2^{n_2}}{n_2!}}{e^{-\lambda} \frac{\lambda^n}{n!}} = \frac{\frac{\lambda_1^{n_1}}{n_1!} \cdot \frac{\lambda_2^{n_2}}{n_2!}}{\frac{\lambda^n}{n!}} = \frac{n!}{n_1! n_2!} \frac{\lambda_1^{n_1} \lambda_2^{n_2}}{\lambda^n} \\ &= \frac{n!}{n_1! n_2!} \frac{\lambda_1^{n_1} \lambda_2^{n_2}}{\lambda^{n_1 + n_2}} = \frac{n!}{n_1! n_2!} \frac{\lambda_1^{n_1} \lambda_2^{n_2}}{\lambda^{n_1} \lambda^{n_2}} = \frac{n!}{n_1! n_2!} \left(\frac{\lambda_1}{\lambda} \right)^{n_1} \left(\frac{\lambda_2}{\lambda} \right)^{n_2} \end{aligned}$$

Mit $p_i := \frac{\lambda_i}{\lambda}$ ergibt sich

$$P(N_1 = n_1 \text{ und } N_2 = n_2 \mid N_1 + N_2 = n) = \frac{n!}{n_1! n_2!} p_1^{n_1} p_2^{n_2}$$

und mit $p_1 + p_2 = 1$, $n_1 + n_2 = n$ und $x = n_1$

$$\frac{n!}{x! (n-x)!} p_1^x (1-p_1)^{n-x} = \binom{n}{x} p_1^x (1-p_1)^{n-x}.$$

10.4 Familie der hypergeometrischen Verteilungen

Betrachtet man eine Urnenzufallsexperiment mit zwei möglichen Ausgängen (z.B. rote oder schwarze Kugel; defektes oder nicht-defektes Objekt; 0 oder 1), legt aber im Gegensatz zur Situation, die auf eine Binomialverteilung führt, die gezogenen Objekte nicht zurück, so kann man die sogenannte hypergeometrische Verteilung entwickeln:

Beispiel 10.4.1. Aus einem Sortiment von $N = 12$ Transistoren werden $n = 4$ zufällig (ohne Zurücklegen, ohne Reihenfolge) gezogen. Es ist bekannt, dass $M = 5$ Transistoren des Sortiments defekt sind. Wie gross ist die Wahrscheinlichkeit, genau $x = 2$ defekte unter den 4 gezogenen Transistoren zu finden. In einer Stichprobe von $n = 4$ und 2 Defekten tauchen auch 2 nicht Defekte auf. Es gibt $\binom{7}{2}$ Möglichkeiten, zwei nicht Defekte aus den 7 nicht Defekten zu ziehen. Für jede dieser Möglichkeiten gibt es $\binom{5}{2}$ Möglichkeiten, 2 Defekte aus den 5 Defekten zu ziehen. Wir erhalten damit $\binom{5}{2} \binom{7}{2}$ „günstige Möglichkeiten“. Insgesamt erhalten wir $\binom{12}{4}$ gleichwahrscheinliche Möglichkeiten, 4er Stichproben zu ziehen. Damit erhalten wir $P(X = 2) = \frac{\binom{5}{2} \binom{7}{2}}{\binom{12}{4}} = \frac{14}{33} = 0.42424$

◇

Allgemein: Ziehen wir ohne Zurücklegen (ohne Reihenfolge) aus N Objekten n Gegenstände, so gibt es $\binom{N}{n}$ mögliche gezogene Mengen. Es gibt $\binom{M}{x}$ Möglichkeiten, aus M ausgezeichneten Gegenständen x ausgezeichnete zu ziehen (ohne Zurücklegen und ohne Reihenfolge). Eine Stichprobe, die x ausgezeichnete Gegenstände enthält, enthält $n-x$ nicht-ausgezeichnete. Es gibt $\binom{N-M}{n-x}$ Möglichkeiten, diese zu ziehen. Somit gibt es für jede der $\binom{M}{x}$ Möglichkeiten, x ausgezeichnete Gegenstände zu ziehen, $\binom{N-M}{n-x}$ Möglichkeiten, nicht-ausgezeichnete zu ziehen. Dies ergibt $\binom{M}{x} \binom{N-M}{n-x}$ „günstige“ Fälle. Dies führt uns auf folgende Definition:

Definition 10.4.2. X ist eine hypergeometrisch verteilte Zufallsvariable genau dann, wenn

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$0 \leq x \leq n; x \leq M; n-x \leq N-M; x \in \mathbb{N}$$

(N = Anzahl der Objekte der Grundgesamtheit; M = Anzahl der ausgezeichneten Objekte in der Grundgesamtheit; n = Stichprobengröße, x = Anzahl der ausgezeichneten Objekte in der Stichprobe)

Abkürzung: $X \sim H(N, n, M)$

N, M und n sind die Parameter der Verteilung)

Die Wahrscheinlichkeitsfunktion einer hypergeometrisch verteilten Zufallsvariable mit den Parametern 200, 30 und 20 sieht wie folgt aus (s. Abbildung 10.4.6)

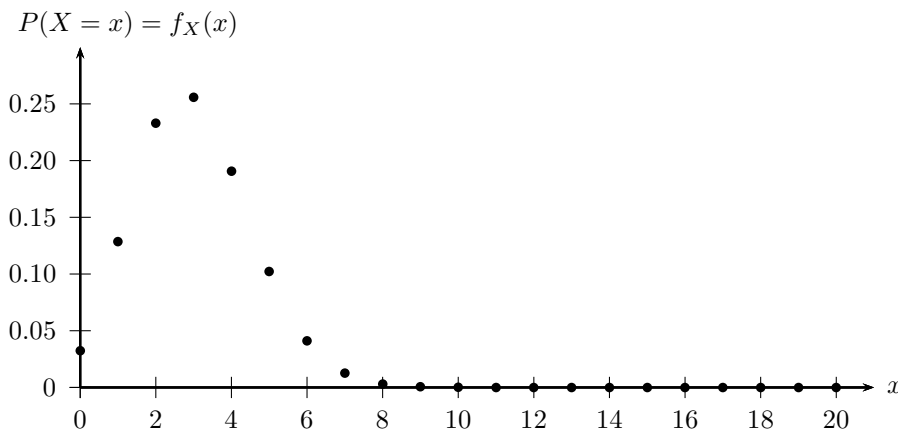


Abbildung 10.4.6: Graphische Darstellung der Wahrscheinlichkeitsfunktion einer hypergeometrisch verteilten Zufallsvariable $X \sim H(200, 20, 30)$

Zu beachten ist, dass die Formel für $x > M$ keinen Sinn ergibt. Man kann in der Stichprobe nicht mehr defekte Objekte haben, als es in der Grundgesamtheit gibt. Zudem kann es nicht mehr als $N-M$ nicht ausgezeichnete Objekte in der Stichprobe haben.

Die hypergeometrische Verteilung kommt bei kleinen Grundmengen zum Einsatz. Bei grossen Grundmengen macht es einen vernachlässigbaren Unterschied, ob man die gezogenen Objekte zurücklegt oder nicht. Entsprechend kann man bezüglich grosser Grundmengen die Binomial-Verteilung verwenden. Betrachten wir zwei Beispiele, um ein Gefühl für die Unterschiede zu entwickeln:

Im obigen Beispiel erhalten wir bei der Verwendung der Binomialverteilung (d.h. mit Zurücklegen) und $p = \frac{5}{12}$:

$$P(X = 2) = \binom{4}{2} \left(\frac{5}{12}\right)^2 \left(1 - \frac{5}{12}\right)^{4-2} = \frac{1225}{3456} = 0.354456019 \quad (10.4.1)$$

Die Wahrscheinlichkeiten unterscheiden sich bei so kleinen Grundmengen ($N = 12$) im Prozentbereich ($0.354\,46 - 0.329\,59 = 0.024\,87$). Betrachten wir ein Beispiel mit einer Grundmenge von 6000. Wir ziehen 4. Der Anteil der defekten an der Grundgesamtheit betrage ebenfalls $\frac{5}{12}$. Somit ergibt sich mit der hypergeometrischen Verteilung ($\frac{5}{12} \cdot 6000 = 2500$)

$$\begin{aligned} \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} &= \frac{\binom{2500}{2} \binom{6000-2500}{4-2}}{\binom{6000}{4}} \\ &= 0.354567463 \end{aligned}$$

Bei der Binomialverteilung spielt die Grösse der Grundgesamtheit keine Rolle - da zurückgelegt wird. Entsprechend erhält man den Wert 0.354456019 von Gleichung (10.4.1). Die Unterschiede sind zu vernachlässigen: $0.354567463 - 0.354456019 = 0.000111444$ - weniger als Promillebereich. Die hypergeometrische Verteilung wird etwa in der Qualitätskontrolle eingesetzt, wenn die Grundmengen klein sind (z.B. $N = 200$) und die gezogenen Gegenstände nicht zurückgelegt werden. Bei Zurücklegen wird jedoch auch bei kleinen Mengen die Binomialverteilung verwendet. Bei grossen Mengen kann auch ohne Zurücklegen die Binomialverteilung verwendet werden.

Die hypergeometrische Verteilung als Verteilung einer Summe von Zufallsvariablen (fakultativer Stoff)

Ähnlich wie die Binomialverteilung kann man die hypergeometrische Verteilung mit den Parametern (N, n, M) als Verteilung der Statistik „Summe von n bernoulliverteilten Zufallsvariablen“ betrachten. Im Unterschied zur Binomialverteilung sind diese Zufallsvariablen weder identisch verteilt noch unabhängig, da die gezogenen Objekte nicht zurückgelegt werden. Wir betrachten ein Beispiel: wie hoch ist die Wahrscheinlichkeit, den Vektor

$$y = (0, 1, 1, 0, 1, 0, 1, 1, 0, 0)$$

zu erhalten, wenn M ausgezeichnete Objekte in N Objekten vorkommen und nicht zurückgelegt werden. Es gilt $n = 10$ (Stichprobengrösse) und die Anzahl der erhaltenen ausgezeichneten Objekte $x = 5 = \sum_{i=1}^{10} y_i$. Wir betrachten y als Wert des Zufallsvektors $Y = (Y_1, \dots, Y_{10})$. Die einzelne Zufallsvariable Y_i ist dabei bernoulliverteilt, da sie die Werte 0 oder 1 annimmt, wobei sich wegen des fehlenden Zurücklegens für jede Zufallsvariable eine andere Wahrscheinlichkeitsverteilung ergibt, in Abhängigkeit von den vorausgegangenen Ergebnissen. Die Unabhängigkeit der Zufallsvariablen gilt also nicht. Fürs Beispiel ergibt sich als Wahrscheinlichkeit, dass die Zufallsvariable Y den Vektor y als Wert annimmt, das hinter $P(Y = y)$ stehende Produkt:

$$P(Y = y) = \frac{0}{N-M} \cdot \frac{1}{N-1} \cdot \frac{1}{N-2} \cdot \frac{0}{N-3} \cdot \frac{1}{N-4} \cdot \frac{0}{N-5} \cdot \frac{1}{N-6} \cdot \frac{1}{N-7} \cdot \frac{0}{N-8} \cdot \frac{0}{N-9}$$

Auf Grund des Kommutativgesetzes für die Multiplikation kann man das Produkt wie folgt umschreiben:

$$\underbrace{\frac{M}{N} \frac{M-1}{N-1} \frac{M-2}{N-2} \frac{M-3}{N-3} \frac{M-4}{N-4}}_{\text{Zähler}} \underbrace{\frac{N-M}{N-5} \frac{N-M-1}{N-6} \frac{N-M-2}{N-7} \frac{N-M-3}{N-8} \frac{N-M-4}{N-9}}_{\text{Nenner}}$$

Für beliebiges n, y und der Summe x der Komponenten von y gilt damit offensichtlich:

$$P(Y = y) = \underbrace{\frac{M}{N} \cdot \frac{M-1}{N-1} \cdot \dots \cdot \frac{M-x+1}{N-x+1}}_{\frac{M!}{(M-x)!}} \cdot \frac{N-M}{N-x} \cdot \underbrace{\frac{N-M-1}{N-x-1} \cdot \dots \cdot \frac{N-M-n+x+1}{N-n+1}}_{\frac{(N-M)!}{(N-M-n+x)!}}$$

Dabei gilt (s. Satz 6.2.7, Kombinatorik) :

$$N \cdot \dots \cdot N-x+1 \cdot N-x \cdot \dots \cdot N-n+1 = \frac{N!}{(N-n)!}$$

$$M \cdot \dots \cdot M-x+1 = \frac{M!}{(M-x)!}$$

$$N-M \cdot \dots \cdot N-M-n+x+1 = \frac{(N-M)!}{(N-M-n+x)!}$$

$$\text{und damit } \frac{M}{N} \cdot \dots \cdot \frac{M-x+1}{N-x+1} \cdot \frac{N-M}{N-x} \cdot \dots \cdot \frac{N-M-n+x+1}{N-n+1} = \frac{\frac{M!}{(M-x)!} \frac{(N-M)!}{(N-M-n+x)!}}{\frac{N!}{(N-n)!}}$$

Als nächstes berechnen wir die Verteilung der Statistik Summe $X := \sum_{i=1}^n Y_i$, die auf das kartesische Produkt der Bilder dieser n Zufallsvariablen Y_i definiert ist. Offensichtlich ergibt sich eine identische Wahrscheinlichkeit $P(X = x)$ für jeden Vektor $(y_1, \dots, y_n)$ mit fixem $x := \sum_{i=1}^n y_i$ unabhängig davon, wie die Summe x zustande kommt. Es gibt $\binom{n}{x}$ solcher Vektoren (s. Binomialverteilung). Entsprechend gilt dann für die Wahrscheinlichkeit, eine Summe von x zu erhalten

$$\begin{aligned} \binom{n}{x} \cdot \frac{\frac{M!}{(M-x)!} \frac{(N-M)!}{(N-M-n+x)!}}{\frac{N!}{(N-n)!}} \\ = \frac{n!}{(n-x)!x!} \cdot \frac{\frac{M!}{(M-x)!} \frac{(N-M)!}{(N-M-n+x)!}}{\frac{N!}{(N-n)!}} \\ = \frac{\frac{M!}{(M-x)!x!} \frac{(N-M)!}{(N-M-n+x)!}}{\frac{N!}{(N-n)!n!}} = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \end{aligned}$$

Weitere Eigenschaften der hypergeometrischen Verteilung

Um zu überprüfen, dass die Funktion $f_X(x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$ eine Wahrscheinlichkeitsfunktion ist, müsste man zeigen, dass

$$\textbf{Satz 10.4.3.} \quad \sum_{x=0}^n \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = 1$$

Beweis. Statt eines algebraischen Beweises der Aussagen wird ein Beweis mittels Wahrscheinlichkeitstheorie geliefert. Die hypergeometrische Verteilung kann als Verteilung der Statistik „Summe“ $X := \sum X_i$ (s. vorangegangener Unterabschnitt) betrachtet werden, die auf den $\mathfrak{X} = \{y : y \in \{0, 1\}^n \text{ und } \sum_{i=1}^n y_i \leq M, \left(n - \sum_{i=1}^n y_i\right) \leq N - M\}$ definiert ist. Es gilt $P(\mathfrak{X}) = 1$. Als nächstes

betrachten wir die folgenden Teilmengen von $\mathfrak{X}$: $M_x := \{y : y \in \{0, 1\}^n, \sum_{i=1}^n y_i = x\} \subset \mathfrak{X}$.

Sei S die Menge der derartigen x . Die Menge der M_x ist offensichtlich eine Zerlegung von $\mathfrak{X}$ und es gilt deshalb $\sum_{x \in S} P(M_x) = P(\mathfrak{X}) = 1$. Zudem gilt auf Grund der Definition der Wahr-

scheinlichkeitsfunktion der hypergeometrischen Funktion $f_X(x) : f_X(x) = P(M_x)$, woraus folgt

$$\sum_{x \in S} f_X(x) = \sum_{x=0}^n \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = 1. \quad \square$$

Dieser Beweis nimmt auf die konkrete Gestalt der Wahrscheinlichkeitsfunktion der hypergeometrischen Funktion $f_X(x)$ keinen Bezug. Er könnte auch auf die Berechnung der Summe der Werte der Wahrscheinlichkeitsfunktion einer binomialverteilten Zufallsvariable angewendet werden.

Satz 10.4.4. Wenn $X \sim H(N, n, M)$, dann $E(X) = n \frac{M}{N}$

Beweis.

$$E(X) = \sum_{x=0}^n x \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} = \sum_{x=1}^n x \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

Mit Satz $\binom{n}{m} = \frac{n}{m} \binom{n-1}{m-1}$ (s. in Kapitel „Kombinatorik“, Satz 6.2.16) erhält man daraus:

$$\begin{aligned} \sum_{x=1}^n x \frac{\frac{M}{x} \binom{M-1}{x-1} \binom{N-1-(M-1)}{n-1-(x-1)}}{\frac{N}{n} \binom{N-1}{n-1}} &= n \frac{M}{N} \sum_{x=1}^n \frac{\binom{M-1}{x-1} \binom{N-1-(M-1)}{n-1-(x-1)}}{\binom{N-1}{n-1}} = \\ n \frac{M}{N} \sum_{x=0}^{n-1} \frac{\binom{M-1}{x} \binom{N-1-(M-1)}{n-1-x}}{\binom{N-1}{n-1}} &= n \frac{M}{N} \end{aligned}$$

(Die Summe in der letzten Gleichung ist 1, da die Summe der Wahrscheinlichkeiten der Wahrscheinlichkeitsfunktion einer hypergeometrischen Funktion mit den Parametern $(N-1, n-1, M-1)$ 1 ergibt). $\square$

Mit $p = \frac{M}{N}$ entspricht dies dem Erwartungswert einer binomialverteilten Zufallsvariable.

Satz 10.4.5. Wenn $X \sim (N, n, M)$, dann $V(X) = \frac{N-n}{N-1} n \frac{M}{N} (1 - \frac{M}{N})$

Beweis.

$$\begin{aligned} V(X) &= \sum_{x=0}^n (x - n \frac{M}{N})^2 \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \\ &= \sum_{x=0}^n \left(x^2 - 2nx \frac{M}{N} + n^2 \frac{M^2}{N^2} \right) \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \\ &= \sum_{x=0}^n x^2 \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} - 2n \frac{M}{N} \sum_{x=0}^n x \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} + n^2 \frac{M^2}{N^2} \sum_{x=0}^n \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \end{aligned}$$

Die Summe im mittleren Ausdruck entspricht dem Erwartungswert einer hypergeometrisch verteilten Zufallsvariablen, ergibt also $n \frac{M}{N}$. Die Summe im letzten Ausdruck ist die Summe der Werte einer hypergeometrisch verteilten Zufallsvariable und ergibt somit 1. Man erhält also:

$$\begin{aligned} V(X) &= \sum_{x=0}^n x^2 \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} - 2n \frac{M}{N} n \frac{M}{N} + n^2 \frac{M^2}{N^2} \\ &= \sum_{x=0}^n x^2 \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} - n^2 \frac{M^2}{N^2} \end{aligned}$$

Für den linken Ausdruck im letzten Term gilt:

$$\begin{aligned} \sum_{x=0}^n x^2 \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} &= \sum_{x=1}^n x^2 \frac{\frac{M}{x} \binom{M-1}{x-1} \binom{N-M}{n-x}}{\frac{N}{n} \binom{N-1}{n-1}} \\ &= \frac{Mn}{N} \sum_{x=1}^n x \frac{\binom{M-1}{x-1} \binom{N-M}{n-x}}{\binom{N-1}{n-1}} \end{aligned}$$

Mit $x = (x-1) + 1$ erhält man:

$$= \frac{Mn}{N} \sum_{x=1}^n (x-1) \frac{\binom{M-1}{x-1} \binom{N-M}{n-x}}{\binom{N-1}{n-1}} + \frac{Mn}{N} \sum_{x=1}^n \frac{\binom{M-1}{x-1} \binom{N-M}{n-x}}{\binom{N-1}{n-1}}$$

und mit $l := x - 1$ und einigen bereits bekannten Umformungen:

$$= \frac{Mn}{N} \sum_{l=0}^{n-1} l \frac{\binom{M-1}{l} \binom{N-1-(M-1)}{n-1-l}}{\binom{N-1}{n-1}} + \frac{Mn}{N} \sum_{x=0}^{n-1} \frac{\binom{M-1}{x} \binom{N-M}{n-1-x}}{\binom{N-1}{n-1}}$$

Die Summe im ersten Ausdruck ist der Erwartungswert einer $(N-1, n-1, M-1)$ verteilten Zufallsvariable und damit $(n-1) \frac{M-1}{N-1}$, die Summe im zweiten Ausdruck ist die Summe der Werte einer hypergeometrischen Wahrscheinlichkeitsfunktion mit den Parametern $(N-1, n-1, M-1)$ und damit mit 1 identisch. Entsprechend erhält man:

$$= \frac{Mn}{N} \frac{(n-1)(M-1)}{N-1} + \frac{Mn}{N}$$

Damit ist dann

$$\begin{aligned} V(X) &= -n^2 \frac{M^2}{N^2} + \frac{Mn}{N} \frac{(n-1)(M-1)}{N-1} + \frac{Mn}{N} \\ &= \frac{-n^2 M^2 (N-1) + NMn(n-1)(M-1) + MnN(N-1)}{N^2(N-1)} \\ &= \frac{nM[-nM(N-1) + N(n-1)(M-1) + N(N-1)]}{N^2(N-1)} \\ &= \frac{nM[-nMN + nM + N - MN - Nn + MNn + N^2 - N]}{N^2(N-1)} \\ &= \frac{nM[nM - MN - Nn + N^2]}{N^2(N-1)} \\ &= \frac{nM(N-n)(n-M)}{N^2(N-1)} = \frac{nM}{N} \frac{N-n}{N-1} \frac{N-M}{N} \\ &= \frac{nM}{N} \frac{N-n}{N-1} \left(1 - \frac{M}{N}\right) \end{aligned}$$

□

$\frac{N-n}{N-1}$ stellt eine Endlichkeitskorrektur dar. Bei grösser werdendem N und konstantem n nähert sich der Ausdruck 1, d.h. $\lim_{N \rightarrow \infty} \frac{N-n}{N-1} = 1$, womit wir mit $p := \frac{M}{N}$ die Formel für die binomialverteilte Zufallsvariable erhalten. Geht n gegen N , so ergibt sich

$$\lim_{n \rightarrow N} V(X) = \lim_{n \rightarrow N} \frac{nM}{N} \frac{N-n}{N-1} \left(1 - \frac{M}{N}\right) = \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N - \lim_{n \rightarrow N} n}{N-1} \lim_{n \rightarrow N} n = \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-N}{N-1} N = 0$$

(wenn man alle zieht, gibt es keine Varianz mehr!).

Die graphische Darstellung der Verteilungsfunktion ist (s. Abbildung 10.4.7):

$$F(x) = P(X \leq x) = \sum_{k=0}^x \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

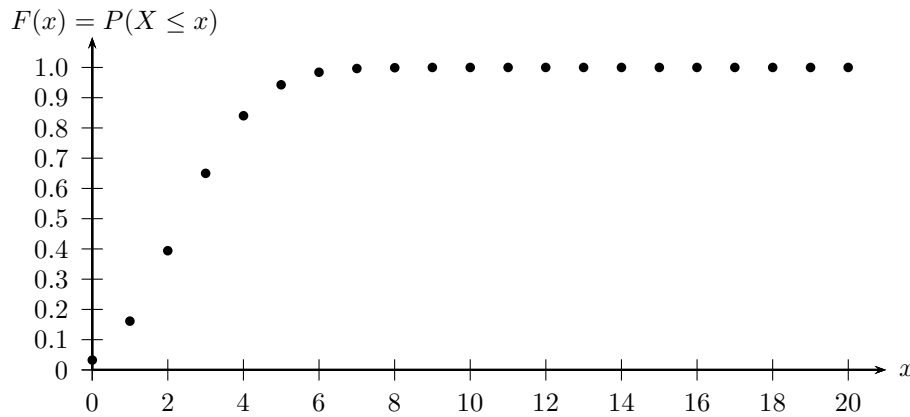


Abbildung 10.4.7: Graphische Darstellung der Verteilungsfunktion einer hypergeometrisch verteilten Zufallsvariablen $X \sim H(200, 20, 30)$

Die bisherigen Verteilungsfamilien sind alle nahe verwandt. Es gibt noch weitere diskrete Verteilungsfamilien, die mit Namen versehen werden und die mit den obigen verwandt sind, die wir jedoch nicht behandeln (z.B. Multinomialverteilung, Geometrische Verteilung, Negative Binomialverteilung, etc.). Zum Schluss behandeln wir noch eine Verteilungsfamilie, die mit den obigen weniger verwandt ist, die uns jedoch schon mehrmals begegnet ist.

10.5 Familie der diskreten uniformen Verteilungen (fakultativer Stoff)

Definition 10.5.1. Die Zufallsvariable X ist diskret uniform (= diskret gleichverteilt) verteilt genau dann, wenn

$$P(X = x) = \frac{1}{k}$$

für alle x ($k = |\mathfrak{X}| = \text{Anzahl Elemente des Bildes von } X$)

Abkürzung: $X \sim U(k)$

Beispiele für uniform verteilte Zufallsvariablen haben wir bereits einige gesehen. Sei etwa

$S = \{\text{ein Auge, zwei Augen, drei Augen, vier Augen, fünf Augen, sechs Augen}\}$

Graph von $f = \{(\text{ein Auge}, \frac{1}{6}), (\text{zwei Augen}, \frac{1}{6}), (\text{drei Augen}, \frac{1}{6}), (\text{vier Augen}, \frac{1}{6}), (\text{fünf Augen}, \frac{1}{6}), (\text{sechs Augen}, \frac{1}{6})\}$

Graph von $X = \{(\text{ein Auge}, 1), (\text{zwei Augen}, 2), (\text{drei Augen}, 3), (\text{vier Augen}, 4), (\text{fünf Augen}, 5), (\text{sechs Augen}, 6)\}$

Damit gilt dann Graph von $f_X = \{(1, \frac{1}{6}), (2, \frac{1}{6}), (3, \frac{1}{6}), (4, \frac{1}{6}), (5, \frac{1}{6}), (6, \frac{1}{6})\}$

Eine uniform verteilte Zufallsvariable weist für jeden Wert dieselbe Wahrscheinlichkeit auf. Im folgenden Beispiel gilt $\mathfrak{X} = \mathbb{N}_{10}^*$ (s. Abbildung 10.5.8).

Satz 10.5.2.

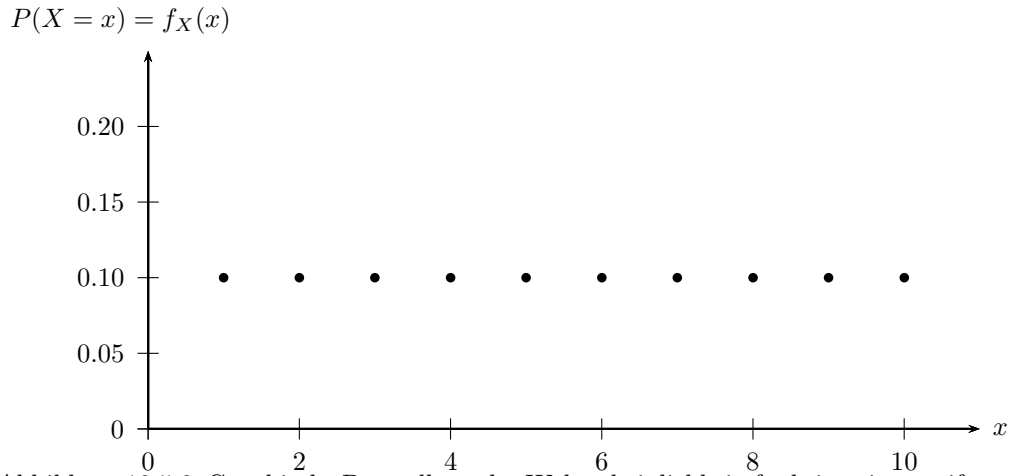


Abbildung 10.5.8: Graphische Darstellung der Wahrscheinlichkeitsfunktion einer uniform verteilten diskreten Zufallsvariable mit $\mathfrak{X} = \mathbb{N}_{10}^*$

$$\text{Wenn } X \sim U(k), \text{ dann } E(X) = \frac{1}{k} \sum_{x=1}^k x_i$$

$$\text{Beweis. } \sum_{i=1}^k x_i P(X = x_i) = \sum_{i=1}^k x_i \frac{1}{k} = \frac{1}{k} \sum_{i=1}^k x_i. \quad \square$$

Der Erwartungswert einer uniformverteilten Zufallsvariablen ist demnach das arithmetische Mittel der Werte der Zufallsvariablen.

So wäre der Erwartungswert des obigen Würfelbeispiels: $\frac{6+1}{2} = 3.5 = \frac{1+2+3+4+5+6}{2}$.

Satz 10.5.3. Wenn $X \sim U(k)$, dann $V(X) = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2$

$$\text{Beweis. } V(X) = \sum_{i=1}^k P(X = x_i) \cdot (x_i - \bar{x})^2 = \frac{1}{k} \sum_{i=1}^k (x_i - \bar{x})^2. \quad \square$$

Für die Verteilungsfunktion gilt: $F(x_i) = P(X \leq x_i) = |\{x_j : x_j \leq x_i \text{ und } x_j \in \mathfrak{X}\}| \cdot \frac{1}{k}$.

Die Punkte der Verteilungsfunktion einer uniform verteilten Zufallsvariable, deren Bildbereich mit $\mathbb{N}_k^*$ zusammenfällt, liegen auf einer Gerade, die durch den 0-Punkt verläuft und deren Steigung $\frac{1}{k}$ ist (Bei anderen diskreten Gleichverteilungen brauchen die Punkte der Verteilungsfunktion nicht auf einer Geraden zu liegen!). Im folgenden Beispiel gilt $\mathfrak{X} = \mathbb{N}_{10}^*$ (s. Abbildung 10.5.9).

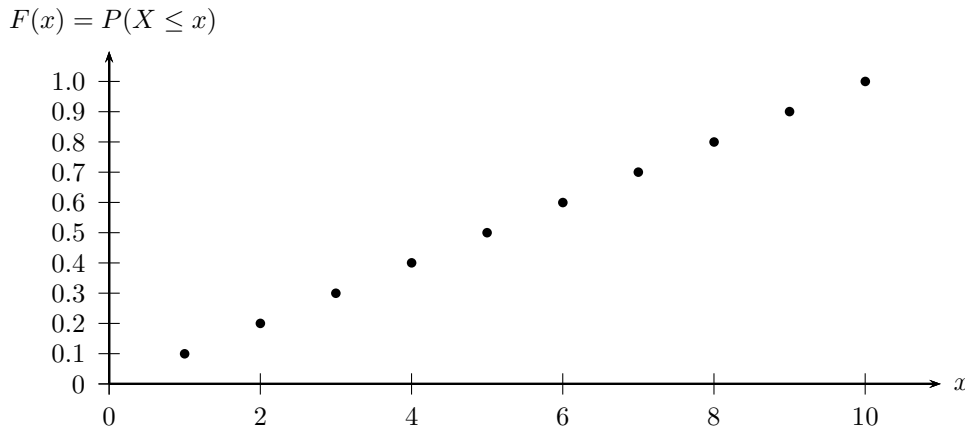


Abbildung 10.5.9: Graphische Darstellung der Verteilungsfunktion einer diskret uniform verteilten Zufallsvariable mit $\mathfrak{X} = \mathbb{N}_{10}^*$.

Wir überprüfen noch die für Wahrscheinlichkeitsverteilungen zu erfüllende Bedingung: $\sum_{i=1}^x P(X = i) = \sum_{i=1}^x \frac{1}{k} = \frac{x}{k} = 1$. Die uniforme Verteilung ist tatsächlich eine Wahrscheinlichkeitsverteilung.

10.5.1 Übungen

- Berechnen sie für
 $X \sim B(2, 0.1); \quad X \sim B(2, 0.6)$
 $X \sim B(3, 0.5); \quad X \sim B(4, 0.2);$
 die Werte $P(X = x) = f_X(x)$
 (Für die ersten drei Fälle mit Excel ohne Verwendung der Excelformel „binom.vert“, für die vierte mit der Excelformel). Berechnen Sie jeweils den Erwartungswert und die Varianz. Erstellen Sie für die Binomialverteilung mit $n = 8$ und $p = 0.8$ das Diagramm der Wahrscheinlichkeitsfunktion. Erstellen Sie für die Binomialverteilung mit $n = 15$ und $p = 0.3$ das Diagramm der Verteilungsfunktion. Berechnen Sie die entsprechenden Erwartungswerte und Varianzen.
- Die relative Häufigkeit von unbrauchbaren Zahnbürsten betrage 0.2.
 - Wie hoch ist die Wahrscheinlichkeit bei $n = 10$ Zügen (mit Zurücklegen und jeweiligem Mischen) 2 unbrauchbare Bürsten zu erhalten.
 - höchstens 2 unbrauchbare Bürsten zu erhalten.
 - Wie hoch ist die Wahrscheinlichkeit 8 unbrauchbare Bürsten zu erhalten.
 - mindestens 8 unbrauchbare Bürsten zu erhalten.
 - Wie hoch ist die Wahrscheinlichkeit bei $n = 20$ Zügen 18 unbrauchbare Bürsten zu erhalten.
- Fluggesellschaften wissen, wie viele der buchenden Passagiere durchschnittlich erscheinen. Um die Flugzeuge möglichst gut auszulasten, werden diese überbucht. Nach Angaben der KLM erscheinen die einzelnen Passagiere mit einer Wahrscheinlichkeit von 0.86 (SAS Communications, 4/1992). Wir nehmen an, dass die Passagiere unabhängig voneinander ihre Reise antreten. Wie gross ist dann die Wahrscheinlichkeit, dass ein Flugzeug mit 82 Plätzen und 85 verkauften Flugtickets alle sich präsentierenden Passagiere mitnehmen kann?
- Die Lieferfirma A und die Bezugsfirma B legen vertraglich folgendes fest: 99% der Güter

müssen verwendbar sein, damit eine Lieferung angenommen wird. Sonst darf sie zurückgewiesen werden. Da es sich jeweils um Millionen von Einzelteilen (z.B. Schrauben, CDs, Borsten) handelt, ist eine vollumfängliche Überprüfung der Grundgesamtheit finanziell nicht vertretbar. Die Firmen einigen sich darauf, dass 30 Objekte zufällig gezogen werden. Sind von dieser Stichprobe mindestens 25 in Ordnung, so muss die Lieferung akzeptiert werden. Berechnen Sie die Wahrscheinlichkeit dafür, dass mindestens 25 in Ordnung sind, wenn 99% in der Grundgesamtheit in Ordnung sind.

5. Bei einem Herstellungsprozess hat man mit 15 % Ausschussware zu rechnen. Berechnen Sie die Wahrscheinlichkeit,
 - (a) unter 10 Exemplaren keinen Ausschuss,
 - (b) unter 20 Exemplaren mindestens ein Ausschusstück zu finden (es wird jeweils zurückgelegt).
6. Durchschnittlich fällt eine Maschine einer Fabrik 0.5 mal pro Tag aus. Wir nehmen an, die Ausfälle pro Tag seien poissonverteilt.
 - a) Berechnen Sie für die entsprechende Poisson-Verteilung $P(X = x)$ für $x \leq 4$.
 - b) Berechnen für das Resultat von a) die Wahrscheinlichkeit, dass mehr als 3 Maschinenausfälle erfolgen.
 - c) Erstellen Sie eine Graphik der Wahrscheinlichkeitsfunktion.
7. In einer Grossstadt erfolge im Durchschnitt pro Tag ein Todesfall infolge eines Verkehrsunfalles. Man berechne die Wahrscheinlichkeit dafür, dass an einem Tag 0, 1, 2, 3, mehr als vier Todesfälle passieren. Die Unfallzahlen pro Tag seien poissonverteilt.
8. Bei einer Serienproduktion habe sich gezeigt, dass mit 2 Promille Ausschuss zu rechnen ist. Die Produkte werden in Packungen von je 500 Stück abgefüllt, ohne dass vorher eventuelle Ausschusstücke ausgeschieden werden können. a) Wie gross ist die Wahrscheinlichkeit, dass alle 500 Stück einer bestimmten Packung in Ordnung sind? b) Bei 5000er Packungen. Wie gross ist die Wahrscheinlichkeit, dass alle 5000 Stück in Ordnung sind?
9. Ungefähr 1 % der Bevölkerung ist farbenblind. Wie gross muss man eine Stichprobe wählen, wenn man mit einer Wahrscheinlichkeit von 99% mindestens eine farbenblinde Person in der Stichprobe haben möchte? Man rechne mit der Binomial- und der Poisson-Verteilung.
10. Die Hersteller einer elektronischen Waage offerieren eine einjährige Garantie. Falls die Waage während dieser Zeit einen Defekt erleidet, wird sie durch eine neue Waage ersetzt. Durchschnittlich fallen im ersten Jahr 0.1 Defekte pro Waage an.
 - a) Wie gross ist der Prozentsatz der Geräte, die während der Garantiezeit ersetzt werden müssen?
 - b) Die Produktionskosten pro Waage betragen 80 Franken, der Verkaufspreis beträgt 120 Franken. Wie gross ist der zu erwartende Gewinn pro Gerät unter Berücksichtigung auftretender Garantieleistungen.
11. Berechnen Sie die Werte der Wahrscheinlichkeitsfunktion für: 10 Taschenlampen werden gezogen (ohne Zurücklegen) aus 15. Der Anteil der defekten betrage 2.
12. In einer Stichprobe von $n = 25$ Glühbirnen aus einem Produktionslos mit $N = 700$ Glühbirnen finden sich 2 defekte Glühbirnen. Man vermutet, dass der Anteil der defekten Glühbirnen in der Grundgesamtheit 0.1 ist.
 - a) Wie hoch ist die Wahrscheinlichkeit, genau 2 defekte Glühbirnen zu ziehen (ohne Zurücklegen).
 - b) Mindestens 2 defekte Glühbirnen zu ziehen (ohne Zurücklegen).
 - c) Mindestens 4 defekte Glühbirnen zu ziehen. (ohne Zurücklegen)
 - d) Berechnen Sie a), b), c) mit der Binomialverteilung.
13. Zeichnen Sie ein Diagramm der Wahrscheinlichkeits- und Verteilungsfunktion der hypergeometrisch verteilten Zufallsvariable X mit $X \sim (500, 20, 200)$

14. Gegeben sei die folgende Tabelle 10.5.2 (Maschinendefekte pro Tag; s. Diskrete Verteilungsfamilien.xls). Überprüfen Sie, ob eine Poissonverteilung angemessen ist und berechnen Sie $P(X \geq 4)$.

Anzahl Tage	Anzahl Defekte
2	0
10	1
20	2
30	3
40	4
25	5
20	6
15	7
5	8
2	9
0	mehr als 9

Tabelle 10.5.2: Eventuell poissonverteilte Daten

15. Berechnen sie Verteilung der Summe von 4 unabhängigen, bernoulliverteilten Zufallsvariablen mit $p = 0.6$, indem Sie die Wahrscheinlichkeitsfunktion auf das Bild der Summe angeben.
16. Vertraglich haben ein Lieferant und der Belieferte abgemacht, dass bei Lieferungen maximal 2% der gelieferten Objekte fehlerhaft sein dürfen. Es wird zudem vereinbart, dass die Lieferung zurückgewiesen werden kann, wenn bei einer Zufallsstichprobe gilt: $P(X \geq x) < 0.01$ ist (x ist die Anzahl der defekten Objekte in der Stichprobe). Wie viele defekte Objekte maximal darf eine Zufallsstichprobe von 40 Objekten haben, damit die Lieferung nicht zurückgewiesen wird (Lieferumfang 450).
17. Schätzen Sie für die Daten der Exceltabelle „Diskrete_Verteilungsfamilien“ (Uebung21) eine geeignete Verteilung und überprüfen Sie die Angemessenheit des geschätzten Modells graphisch.
18. (fakultativ) Sei $|\mathfrak{X}| = 22$ mit $X \sim U(22)$ und $\mathfrak{X} = \mathbb{N}_{22}^*$. Zeichnen Sie die Verteilungsfunktion und die Wahrscheinlichkeitsfunktion. Berechnen Sie die folgenden Wahrscheinlichkeiten: $P(X = 3), P(X = 4), P(2 \leq X \leq 20), P(X \geq 5), P(X < 4)$
19. (fakultativ) Wie hoch ist die Wahrscheinlichkeit dafür, bei einem Kartenspiel nur Karten von einer Farbe zu erhalten? (4 Spieler, 36 Karten, vier Farben, pro Farbe 9 Karten, pro Farbe je ein As)
20. (fakultativ) Wie hoch ist die Wahrscheinlichkeit dafür, bei einem Kartenspiel vier As zu erhalten (4 Spieler, 36 Karten, vier Farben, pro Farbe 9 Karten, pro Farbe je ein As)

10.5.2 Lösungen

1. 1) **n = 2 und p = 0.1:** $\binom{2}{0} 0.1^0 \cdot 0.9^2 = 0.81$
 $\binom{2}{1} 0.1^1 \cdot 0.9^{2-1} = 0.18$
 $\binom{2}{2} 0.1^2 \cdot 0.9^{2-2} = 0.01$
 $E(X) = 2 \cdot 0.1 = 0.2; \quad V(X) = 2 \cdot 0.1 \cdot 0.9 = 0.18$

n = 2 und p = 0.6:

$$\binom{2}{0} 0.6^0 \cdot 0.4^{2-0} = 0.16$$

$$\binom{2}{1} 0.6^1 \cdot 0.4^{2-1} = 0.48$$

$$\binom{2}{2} 0.6^2 \cdot 0.4^{2-2} = 0.36$$

$$E(X) = 2 \cdot 0.6 = 1.2; \quad V(X) = 2 \cdot 0.6 \cdot 0.4 = 0.48$$

$$\mathbf{n = 3; p = 0.5}$$

$$\binom{3}{0} 0.5^0 \cdot 0.5^{3-0} = 0.125$$

$$\binom{3}{1} 0.5^1 \cdot 0.5^{3-1} = 0.375$$

$$\binom{3}{2} 0.5^2 \cdot 0.5^{3-2} = 0.375$$

$$\binom{3}{3} 0.5^3 \cdot 0.5^{3-3} = 0.125$$

$$E(X) = 3 \cdot 0.5 = 1.5; V(X) = 3 \cdot 0.5 \cdot 0.5 = 0.75$$

$$\mathbf{n = 4; p = 0.2}$$

$$\binom{4}{0} 0.2^0 \cdot 0.8^{4-0} = 0.4096$$

$$\binom{4}{1} 0.2^1 \cdot 0.8^{4-1} = 0.4096$$

$$\binom{4}{2} 0.2^2 \cdot 0.8^{4-2} = 0.1536$$

$$\binom{4}{3} 0.2^3 \cdot 0.8^{4-3} = 0.0256$$

$$\binom{4}{4} 0.2^4 \cdot 0.8^{4-4} = 0.0016$$

$$E(X) = 4 \cdot 0.2 = 0.8; V(X) = 4 \cdot 0.2 \cdot 0.8 = 0.64$$

$$\mathbf{n = 8; p = 0.8}$$

$$E(X) = 8 \cdot 0.8 = 6.4; V(X) = 8 \cdot 0.8 \cdot 0.2 = 1.28$$

(für die graphische Darstellung der Wahrscheinlichkeitsfunktion s. Abbildungen [10.5.10](#))

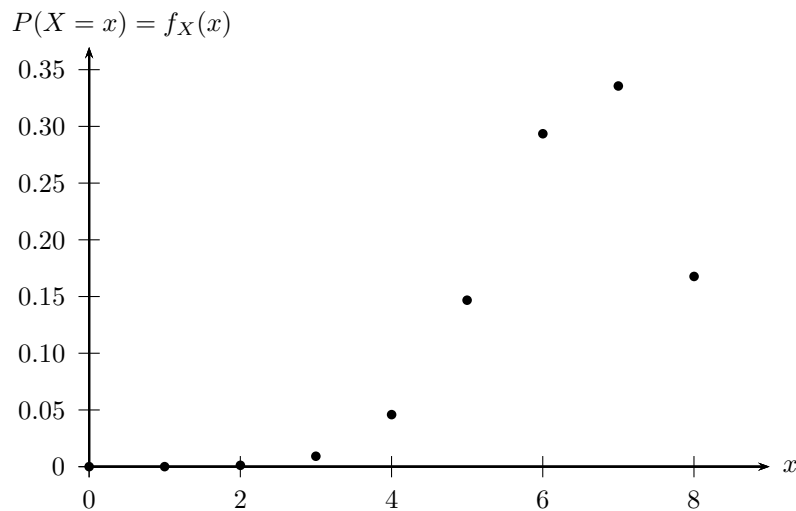


Abbildung 10.5.10: Wahrscheinlichkeitsfunktion einer binomialverteilten Zufallsvariable ($n = 8; p = 0.8$)

$$\mathbf{n = 15, p = 0.3}$$

$$E(X) = 15 \cdot 0.3 = 4.5; V(X) = 15 \cdot 0.3 \cdot 0.7 = 3.15 \text{ (für die graphische Darstellung der Verteilungsfunktion s. Abbildungen [10.5.11](#))}$$

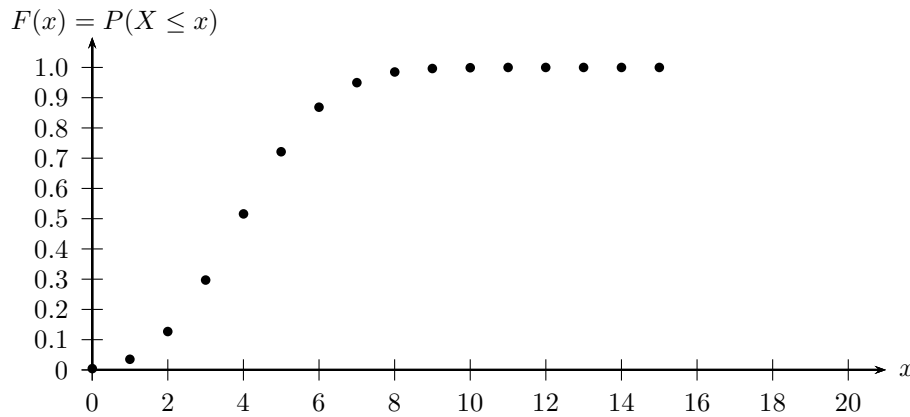


Abbildung 10.5.11: Verteilungsfunktion einer binomialverteilten Zufallsvariable ($n = 15; p = 0.3$)

Eine Bürste ist unbrauchbar oder brauchbar. Entsprechend liegt eine Binomialverteilung vor. $n = 10$ und $p = 0.2$ für unbrauchbare Bürsten.

(i) Die Wahrscheinlichkeit genau zwei unbrauchbare Bürsten zu erhalten beträgt, d.h. $P(X = 2)$ für $X \sim B(10, 0.2)$: $P(X = 2) = \binom{10}{2} 0.2^2 \cdot 0.8^{10-2} = 0.30199$

(ii) Man erhält höchstens 2 unbrauchbare Bürsten, wenn man keine, eine oder zwei unbrauchbare Bürsten erhält, d.h. $P(X \leq 2)$ für $X \sim B(10, 0.2)$

Somit beträgt die entsprechende Wahrscheinlichkeit $= P(X \leq 2) = \sum_{k=0}^2 \binom{10}{k} 0.2^k (1-0.2)^{10-k} = \binom{10}{0} 0.2^0 \cdot 0.8^{10-0} + \binom{10}{1} 0.2^1 \cdot 0.8^{10-1} + \binom{10}{2} 0.2^2 \cdot 0.8^{10-2} = 0.30199 + 0.10737 + 0.26844 = 0.6778$

(iii) $P(X = 8)$ für $X \sim B(10, 0.2)$: $P(X = 8) = \binom{10}{8} 0.2^8 \cdot 0.8^{10-8} = 7.3728 \times 10^{-5}$

(iv) Mindestens 8 werden erhalten, wenn man 8, 9 oder 10 erhält. Somit gilt:

$$P(X \geq 8) = 1 - P(X < 8) = 1 - P(X \leq 7) = 1 - \sum_{k=0}^7 \binom{10}{k} 0.2^k (1-0.2)^{10-k} = 7.7926 \times 10^{-5}$$

Oder durch direkte Berechnung:

$$\binom{10}{8} 0.2^8 \cdot 0.8^{10-8} + \binom{10}{9} 0.2^9 \cdot 0.8^{10-9} + \binom{10}{10} 0.2^{10} \cdot 0.8^{10-10} = 7.7926 \times 10^{-5}$$

(Die erste Berechnungsart ist schneller, wenn man den HP oder Excel benutzt).

(v) $\binom{20}{18} 0.2^{18} (1-0.2)^2 = 3.1877 \times 10^{-11}$

3. Passagiere präsentieren sich oder nicht. Es geht somit um eine Binomialverteilung ($n = 85$, $p = 0.86$).

Wir berechnen die Wahrscheinlichkeit $P(X \geq 83)$.

$$\binom{85}{83} 0.86^{83} \cdot 0.14^{85-83} = 2.5603 \times 10^{-4}$$

$$\binom{85}{84} 0.86^{84} \cdot 0.14^{85-84} = 3.7447 \times 10^{-5}$$

$$\binom{85}{85} 0.86^{85} \cdot 0.14^{85-85} = 2.7063 \times 10^{-6}$$

$$2.5603 \times 10^{-4} + 3.7447 \times 10^{-5} + 2.7063 \times 10^{-6} = 2.9618 \times 10^{-4} = 0.00029618$$

Die Wahrscheinlichkeit, dass man alle mitnehmen kann, beträgt somit:

$$1 - 2.9618 \times 10^{-4} = 0.9997$$

Auf lange Sicht können somit in 99.97% der Fälle alle sich präsentierenden Passagiere mitfliegen.

4. Wir können eine Binomialverteilung voraussetzen. Wir berechnen die Wahrscheinlichkeit dafür, dass 25, 26, 27, 28, 29 oder 30 in Ordnung sind: $P(X \geq 25)$ für $X \sim (30, 0.99)$:
- $$\binom{30}{25} 0.99^{25} \cdot 0.01^{30-25} + \binom{30}{26} 0.99^{26} \cdot 0.01^{30-26} + \binom{30}{27} 0.99^{27} \cdot 0.01^{30-27} + \binom{30}{28} 0.99^{28} \cdot 0.01^{30-28} + \binom{30}{29} 0.99^{29} \cdot 0.01^{30-29} + \binom{30}{30} 0.99^{30} \cdot 0.01^{30-30} \approx 1.00000$$

(mit Excel berechnen wir $1 - \sum_{k=0}^{24} \binom{30}{k} 0.99^k (0.01)^{30-k}$)

Die Wahrscheinlichkeit, mindestens 25 brauchbare zu ziehen, wenn die Lieferung nicht mehr als 1% schlechte Ware oder 99% gute Ware enthält, beläuft sich dann auf beinahe 1.

Zu wessen Gunsten wurde der Test vereinbart?

5. $p = 0.15$

(a) Wir berechnen $P(X = 0)$ für $X \sim B(10, 0.15)$: $\binom{10}{0} 0.15^0 \cdot (1 - 0.15)^{10-0} = 0.19687$

(b) Wir berechnen $P(X \geq 1) = 1 - P(X = 0)$ für $X \sim B(20, 0.15)$: $1 - \binom{20}{0} 0.15^0 (1 - 0.15)^{20-0} = 0.96124$

6. a) $P(X = 0) = \frac{0.5^0}{0!} e^{-0.5} = 0.60653$

$$P(X = 1) = \frac{0.5^1}{1!} e^{-0.5} = 0.30327$$

$$P(X = 2) = \frac{0.5^2}{2!} e^{-0.5} = 0.075816$$

$$P(X = 3) = \frac{0.5^3}{3!} e^{-0.5} = 0.012636$$

$$P(X = 4) = \frac{0.5^4}{4!} e^{-0.5} = 0.0015795$$

Dies ergibt: $P(X \leq 4) = 0.99983$ (was allerdings nicht gefragt ist)

b) Die Wahrscheinlichkeit, dass mehr als 3 Maschinenausfälle erfolgen:

$$1 - \left(\frac{0.5^0}{0!} e^{-0.5} + \frac{0.5^1}{1!} e^{-0.5} + \frac{0.5^2}{2!} e^{-0.5} + \frac{0.5^3}{3!} e^{-0.5} \right) = 0.0017516$$

Mehr als drei Maschinen würden damit durchschnittlich nur alle Tausend Tage ausfallen. Würden an mehreren Tagen mehr Maschinen als 3 ausfallen, so wäre es an der Zeit, die Hypothese $\lambda = 0.5$ in Frage zu stellen.

(c) Graphik s. Abbildung [10.5.12](#)

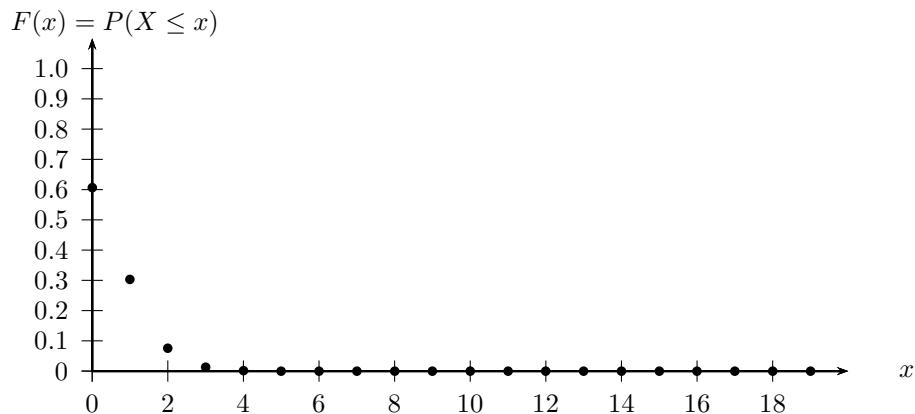


Abbildung 10.5.12: Graphische Darstellung der Wahrscheinlichkeitsfunktion einer poissonverteilten Zufallsvariable ($\lambda = 0.5$)

7. $P(X = 0) = \frac{1^0}{0!} e^{-1} = 0.36788$

$$P(X = 1) = \frac{1^1}{1!} e^{-1} = 0.36788$$

$$P(X = 2) = \frac{1^2}{2!} e^{-1} = 0.18394$$

$$P(X = 3) = \frac{1^3}{3!}e^{-1} = 0.061313$$

$$P(X = 4) = \frac{1^4}{4!}e^{-1} = 0.015328$$

$$P(X > 4) = 1 - \left(\frac{1^0}{0!}e^{-1} + \frac{1^1}{1!}e^{-1} + \frac{1^2}{2!}e^{-1} + \frac{1^3}{3!}e^{-1} + \frac{1^4}{4!}e^{-1} \right) = 0.0036598$$

8. Die Situation kann durch das Modell einer Binomialverteilung dargestellt werden. Da p klein ist, könnten wir auch mit der Poissonverteilung rechnen.

a) Der näherungsweise Erwartungswert beträgt $np = 500 \cdot 0.002 = 1$

$$P(X = 0) = \frac{1^0}{0!}e^{-1} = 0.36788$$

Mit der Binomialverteilung: 0.367511

Wir sehen, dass die Resultate sehr nahe liegen.

b) Der näherungsweise Erwartungswert beträgt: $5000 \cdot 0.002 = 10$

$$P(X = 0) = \frac{10^0}{0!}e^{-10} = 0.0000454$$

Mit der Binomialverteilung: 0.000049475

9. Da p klein ist, können wir mit der Binomial- oder der Poissonverteilung rechnen:

a) Mit Binomialverteilung: $P(X > 0) = 1 - P(X = 0) = 0.99$

$$P(X = 0) = \binom{n}{0} 0.01^0 (1 - 0.01)^{n-0} = 1 \cdot 1 \cdot 0.99^n = 0.99^n$$

Somit ist: $1 - P(X = 0) = 1 - 0.99^n = 0.99$

Wir haben eine Gleichung mit einer Unbekannten. Es gilt, diese nach n aufzulösen.

$$-0.99^n = 0.99 - 1 \implies 0.99^n = 1 - 0.99 = 0.01 \implies$$

$$n \log 0.99 = \log 0.01 \implies n = \frac{\log 0.01}{\log 0.99} = 458.21$$

Somit muss $n > 458$ gelten.

b) Mit Poisson-Verteilung: $P(X > 0) = 1 - P(X = 0) = 0.99$

Der näherungsweise Erwartungswert λ ergibt sich durch $n \cdot 0.01$.

$$P(X = 0) = \frac{(n \cdot 0.01)^0}{0!} e^{-(n \cdot 0.01)} = e^{-(n \cdot 0.01)} \quad (\text{Da } (n \cdot 0.01)^0 = 1; 0! = 1)$$

$$\text{Es gilt somit: } 1 - e^{-(n \cdot 0.01)} = 0.99 \implies -e^{-(n \cdot 0.01)} = 0.99 - 1 \implies e^{-(n \cdot 0.01)} = 1 - 0.99 = 0.01$$

Wir lösen diese Gleichung nach n auf:

$$\ln e^{-(n \cdot 0.01)} = \ln 0.01 \implies -(n \cdot 0.01) \ln e = \ln 0.01 \implies (\text{da } \ln e = 1)$$

$$-(n \cdot 0.01) = \ln 0.01$$

$$n = -\frac{\ln 0.01}{0.01} = 460.52$$

$$n > 460.$$

10. a) Zurückgenommen und ersetzt werden Geräte, bei denen mindestens ein Defekt auftritt. Die Wahrscheinlichkeit dafür, dass mindestens ein Defekt auftritt ist:

$$1 - P(X = 0) = 1 - \frac{0.1^0}{0!}e^{-0.1} = 0.095163$$

Somit fallen durchschnittlich 9.5163 % der Geräte als defekt an.

b) Ein ersetztes Gerät verursacht Kosten von 80 Franken. Auf 100 produzierte Geräte ergibt sich ein Gewinn von $100 \cdot 40$ und ein Verlust von $9.5163 \cdot 80$.

$$\text{Dies ergibt: } 100 \cdot 40 - (9.5163 \cdot 80) = 3238.7$$

Pro Gerät ergibt sich ein durchschnittlicher Gewinn von 32.387 Franken.

$$11. P(X = 0) = \frac{\binom{2}{0} \binom{15-2}{10-0}}{\binom{15}{10}} = 0.095238095; P(X = 1) = \frac{\binom{2}{1} \binom{15-2}{10-1}}{\binom{15}{10}} = 0.476190476$$

$$P(X = 2) = \frac{\binom{2}{2} \binom{15-2}{10-2}}{\binom{15}{10}} = 0.428571429; (\text{da es in der Grundgesamtheit nur 2 hat, kann es nicht mehr in der Stichprobe haben: } P(X = i) = 0 \text{ für } i \geq 3)$$

$$12. 700 \cdot 0.1 = 70 = M$$

$$a) P(X = 2) = \frac{\binom{70}{2} \binom{700-70}{25-2}}{\binom{700}{25}} = 0.26932$$

b) gesucht ist: $1 - [P(X = 1) + P(X = 0)] =$

$$1 - \left[\frac{\binom{70}{1} \binom{700-70}{25-1}}{\binom{700}{25}} + \frac{\binom{70}{0} \binom{700-70}{25-0}}{\binom{700}{25}} \right] = 1 - 0.2658 = 0.7342$$

c) gesucht ist: $1 - [P(X = 3) + P(X = 2) + P(X = 1) + P(X = 0)] =$

$$1 - \left[\frac{\binom{70}{3} \binom{700-70}{25-3}}{\binom{700}{25}} + \frac{\binom{70}{2} \binom{700-70}{25-2}}{\binom{700}{25}} + 0.2658 \right] = 1.0 - 0.76605 = 0.23395$$

d) für die Binomialverteilung: $P(X = 2) = \binom{25}{2} 0.1^2 0.9^{25-2} = .26589$

$P(X \geq 2) =$

$$1 - \left[\binom{25}{1} 0.1^1 0.9^{25-1} + \binom{25}{0} 0.1^0 0.9^{25-0} \right] = 1.0 - 0.27121 = 0.72879$$

$$P(X \geq 4) = 1 - \left[\binom{25}{3} 0.1^3 0.9^{25-3} + \binom{25}{2} 0.1^2 0.9^{25-2} + 0.27121 \right] =$$

$$1 - 0.76360 = 0.2364$$

$$13. P(X = 0) = 2.8121 \cdot 10^{-5}$$

$$P(X = 1) = 0.000400299;$$

$$P(X = 2) = 0.002683566$$

$$P(X = 3) = 0.011265288$$

$$P(X = 4) = 0.033210782$$

$$P(X = 5) = 0.073087027 \text{ (s. Abbildungen 10.5.13 und 10.5.14)}$$

$$P(X = x) = f_X(x)$$

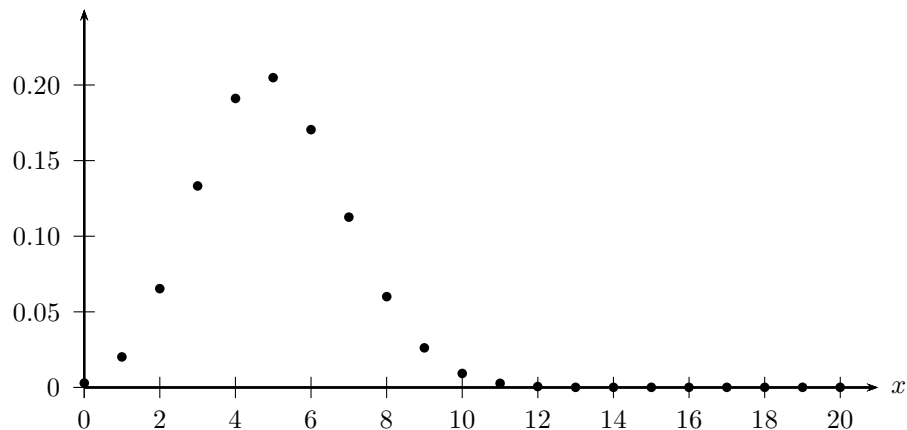


Abbildung 10.5.13: Wahrscheinlichkeitsfunktion einer hypergeometrisch verteilten Zufallsvariable ($N = 500; n = 20; M = 200$)

$$F(x) = P(X \leq x)$$

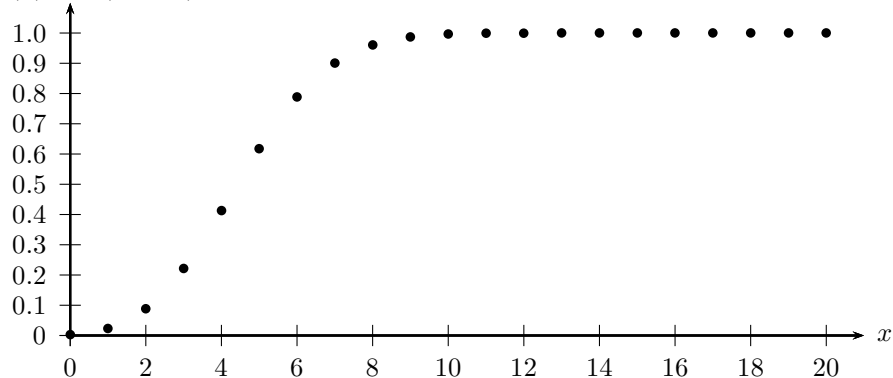


Abbildung 10.5.14: Verteilungsfunktion einer hypergeometrisch verteilten Zufallsvariable ($N = 500; n = 20; M = 200$)

14. Wir erhalten die folgende Tabelle 10.5.3

Anzahl Tage	Anzahl Defekte	Tage* Defekte	relative Häufigkeiten	Werte der geschätzten Poissonverteilung
2	0	0	0.01183432	0.015156147
10	1	10	0.059171598	0.063494389
20	2	40	0.118343195	0.133000082
30	3	90	0.177514793	0.185727925
40	4	160	0.236686391	0.194519779
25	5	125	0.147928994	0.162982253
20	6	120	0.118343195	0.113798259
15	7	105	0.088757396	0.068105805
5	8	40	0.029585799	0.035664874
2	9	18	0.01183432	0.016601401
0	mehr als 9	0	0	0.010949085
Σ	169	708	1	1

Tabelle 10.5.3: Berechnung der relativen Häufigkeiten sowie der geschätzten Wahrscheinlichkeitsfunktion

Graphik der empirischen und der theoretischen Wahrscheinlichkeitsfunktion (wir können die relativen Häufigkeiten, der Tage mit x Ereignissen als eine Art empirische Wahrscheinlichkeitsfunktion betrachten, s. Abbildung 10.5.15).

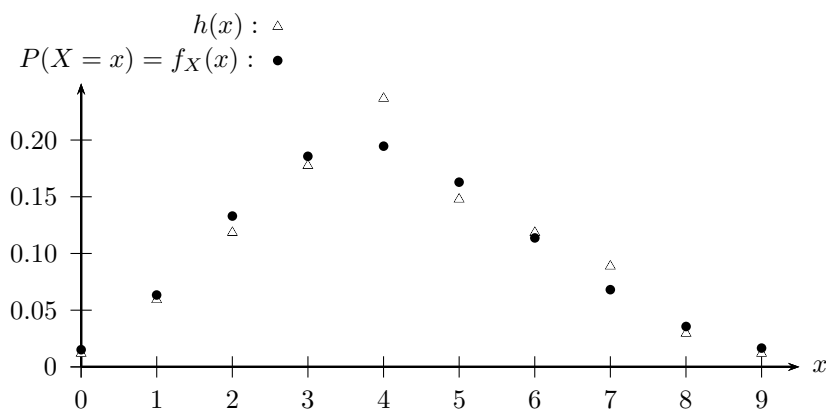


Abbildung 10.5.15: Graphischer Vergleich der relativen Häufigkeiten $h(x)$ und der Werte der geschätzten Wahrscheinlichkeitsfunktion $\hat{f}_X$ einer poissonverteilten Zufallsvariable; $\hat{\lambda} = 4.189349112$

Wir gehen davon aus, dass das Modell der Poissonverteilung sinnvoll ist und berechnen:

$$P(X \geq 4) = 1 - P(X \leq 3) =$$

$$1 - \left(\sum_{k=0}^3 \frac{4.189349112^k}{k!} e^{-4.189349112} \right) = 0.60262$$

15. $2^4 = 16$ Zeilen für das kartesische Produkt der Bilder $\{0,1\}$ der Zufallsvariablen (s. Tabelle 10.5.4).

	Var1	Var2	Var3	Var4	w1	w2	w3	w4	gem	sum	bild	Wfun
1	0	0	0	0	0.40	0.40	0.40	0.40	0.03	0	0	0.0256
2	1	0	0	0	0.60	0.40	0.40	0.40	0.04	1	1	0.1536
3	0	1	0	0	0.40	0.60	0.40	0.40	0.06	1	2	0.3456
4	1	1	0	0	0.60	0.60	0.40	0.40	0.09	2	3	0.3456
5	0	0	1	0	0.40	0.40	0.60	0.40	0.04	1	4	0.1296
6	1	0	1	0	0.60	0.40	0.60	0.40	0.06	2		
7	0	1	1	0	0.40	0.60	0.60	0.40	0.09	2		
8	1	1	1	0	0.60	0.60	0.60	0.40	0.13	3		
9	0	0	0	1	0.40	0.40	0.40	0.60	0.03	1		
10	1	0	0	1	0.60	0.40	0.40	0.60	0.04	2		
11	0	1	0	1	0.40	0.60	0.40	0.60	0.06	2		
12	1	1	0	1	0.60	0.60	0.40	0.60	0.09	3		
13	0	0	1	1	0.40	0.40	0.60	0.60	0.04	2		
14	1	0	1	1	0.60	0.40	0.60	0.60	0.06	3		
15	0	1	1	1	0.40	0.60	0.60	0.60	0.09	3		
16	1	1	1	1	0.60	0.60	0.60	0.60	0.13	4		

Tabelle 10.5.4: Berechnung der Statistik Summe von 4 i.i.d. bernoulliverteilten Zufallsvariablen

Oder in alternativer Darstellung: $\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3 \times \mathfrak{X}_4 =$
 $\{(0, 0, 0, 0), (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 1),$
 $(1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 1, 1),$
 $(1, 1, 1, 0), (1, 1, 0, 1), (1, 0, 1, 1), (0, 1, 1, 1), (1, 1, 1, 1)\}$

Bild der Statistik Summe $X = \{0, 1, 2, 3, 4\}$

Wir geben die Verteilung mit Hilfe des Graph der Wahrscheinlichkeitsfunktion von X an:
 $f_X = \{(0, 0.4^4), (1, 4 \cdot 0.4^3 \cdot 0.6), (2, 6 \cdot 0.4^2 \cdot 0.6^2),$
 $(3, 4 \cdot 0.4 \cdot 0.6^3), (4, 0.6^4)\}$

16. Es ist die hypergeometrische Verteilung zu verwenden (s. Tabelle 10.5.5).

Es gilt: $M = 0.02 \cdot 450 = 9$

Werte x der Zufallsvariable	Werte der Wahrscheinlich- keitsfunktion	$P(X \geq x)$	wenn $P(X \geq x)$ < 0.01 dann 1 sonst 0
0	0.429246765	1	0
1	0.384400088	0.570753235	0
2	0.148800034	0.186353147	0
3	0.032657433	0.037553113	0
4	0.004475278	0.004895679	1
5	0.000396823	0.000420402	1
6	2.27499E-05	2.35789E-05	1
7	8.12495E-07	8.29026E-07	1
8	1.6389E-08	1.65311E-08	1
9	1.42127E-10	1.42127E-10	1

Tabelle 10.5.5: Werte der umgekehrt kumulativen Werte der Wahrscheinlichkeitsfunktion einer hypergeometrischverteilten Zufallsvariable

Damit darf die Stichprobe bei einem Lieferumfang von 450 Stücken höchstens drei defekte Stücke enthalten.

17. Wahrscheinlichkeitsfunktion $f_X(x) = P(X = x)$ (Es gilt: $\mathfrak{X} = \mathbb{N}_{22}^*$, s. Abbildung 10.5.16):

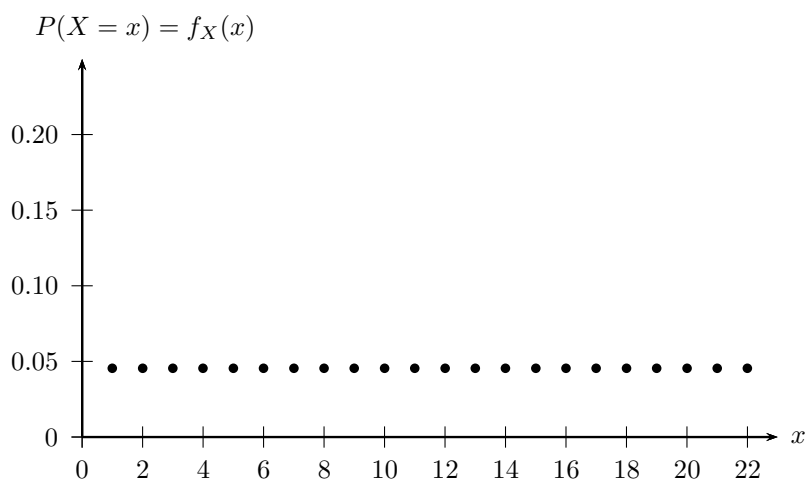


Abbildung 10.5.16: Graphische Darstellung der Wahrscheinlichkeitsfunktion einer diskreten, uniformverteilten Zufallsvariable X ($\mathfrak{X} = \mathbb{N}_{22}^*$)

- Verteilungsfunktion (Es gilt: $\mathfrak{X} = \mathbb{N}_{22}^*$, s. Abbildung 10.5.17)

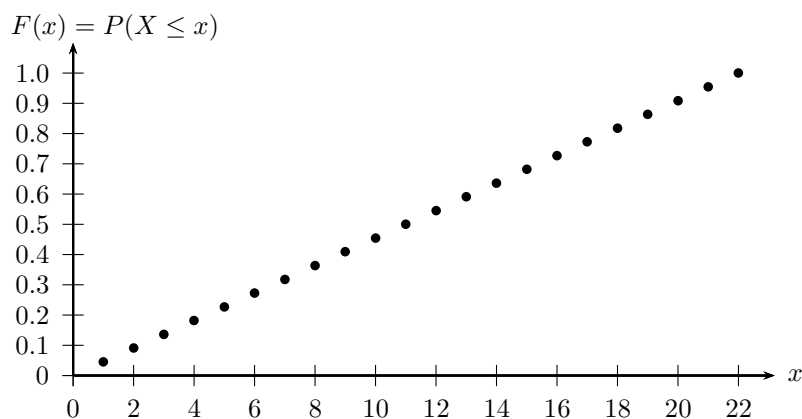


Abbildung 10.5.17: Graphische Darstellung der Verteilungsfunktion einer diskreten, uniformverteilten Zufallsvariable X ($\mathfrak{X} = \mathbb{N}_{22}^*$)

$$\begin{aligned}
 P(X = 3) &= \frac{1}{22}; \\
 P(X = 4) &= \frac{1}{22}, \\
 P(2 \leq X \leq 20) &= \frac{19}{22} \\
 P(X \geq 5) &= 1 - P(X \leq 4) = 1 - \frac{4}{22} = \frac{9}{11} \\
 P(X < 4) &= P(X \leq 3) = \frac{3}{22}
 \end{aligned}$$

18. s. Exceltabelle.

19. (Denksport, fakultativ) Wir können voraussetzen, dass jeder Bezug von 9 Karten ohne Zurücklegen und ohne Reihenfolge gleichwahrscheinlich ist (uniforme Verteilung). Damit beträgt die Wahrscheinlichkeit dafür, genau ein spezifisches Set von Karten zu erhalten: $\frac{1}{\binom{36}{9}}$. Der Bezug von 9 Karten gleicher Farbe ist dabei bei 4 Sets erreicht. Entsprechen gilt: $P = \frac{4}{\binom{36}{9}} = 4.2488 \times 10^{-8}$.
20. (Denksport, fakultativ) Wir können wiederum voraussetzen, jeder Bezug von 9 Karten ohne Zurücklegen und ohne Reihenfolge gleichwahrscheinlich ist (uniforme Verteilung). Damit

beträgt die Wahrscheinlichkeit dafür, genau ein spezifisches Set von Karten zu erhalten: $\frac{1}{\binom{36}{9}}$.

Jetzt müssen wir untersuchen, in wie vielen Sets vier Asse vorkommen. Ein Set, das die 4 Asse enthält kann eine beliebige Menge von 5 weiteren Karten aus den übrigen Karten enthalten. Dies sind $\binom{32}{5}$.

Somit gibt es $\binom{32}{5}$ Sets, die die vier Asse enthalten. Somit beträgt die Wahrscheinlichkeit, vier Asse zu erhalten: $\frac{\binom{32}{5}}{\binom{36}{9}} = 2.139 \times 10^{-3}$

10.6 Lernziele

- Den Begriff der Verteilungsfamilie mit Hilfe eines Beispiels erläutern können.
- Die Begriffe „Bernoulli-, Binomial-, Poisson-Verteilung“ kennen (Abkürzungen verstehen). Begriffe der hypergeometrischen und diskreten uniformen Verteilung kennen (Abkürzung verstehen).
- Werte dieser Verteilungen und der entsprechenden Verteilungsfunktionen berechnen können (von Hand, mit Excel, mit Statistikprogramm).
- Wahrscheinlichkeitsfunktionen und Verteilungsfunktionen der besprochenen Art zeichnen können.
- Die Ausgangsformeln für den Erwartungswert und die Varianz für jede der Verteilungen durch Einsetzen in die entsprechenden Definitionsformeln erzeugen können (die Resultate der Herleitungen können jedoch nachgeschlagen werden).
- Aufgaben von der Art der Übungen lösen können (Auf Grund einer Situationsbeschreibung die richtige Verteilungsfamilie wählen, die Parameter der Verteilung bestimmen und Wahrscheinlichkeiten berechnen können, ohne die Aufgaben, die als fakultativ gekennzeichnet sind).
- Die Verteilung der Summe von unabhängigen, identisch bernoulliverteilten Zufallsvariablen berechnen können.

Kapitel 11

Stetige Wahrscheinlichkeitsräume

Bei Variablen, die metrisch und stetig skaliert sind, sind im betrachteten Intervall theoretisch alle reellen Zahlen als Ausprägungen zulässig und es gibt damit überabzählbar unendlich viele Ausprägungen. Im Allgemeinen würden zwei verschiedene Objekte nie in dieselbe Kategorie fallen (die selbe Ausprägung haben). Misst man z.B. 100 Schrauben genügend genau, wird vermutlich keine Schraube genau gleich lang sein wie eine andere.

Eine diskrete Wahrscheinlichkeitsfunktion ordnet den einzelnen Werten einer Zufallsvariable Wahrscheinlichkeiten zu. Für die Summe der Wahrscheinlichkeiten $f_X(x_i)$, die den Werten x_i der Zufallsvariable zugeordnet sind, muss gelten: $\sum_{i=1}^n f_X(x_i) = 1$, wobei statt n auch ∞ stehen kann. Würden wir nun aber z.B. den überabzählbar unendlich vielen möglichen Ausprägungen x einer uniform verteilten Zufallsvariable positive Wahrscheinlichkeiten $f_X(x)$ zuordnen, würde sich das Problem ergeben, dass schon für eine abzählbar unendliche Teilmenge der Menge dieser Ausprägungen die Summe der Wahrscheinlichkeiten grösser als 1 wäre. Für eine uniforme Wahrscheinlichkeitsfunktion gilt nämlich: $f_X(x) = \frac{1}{k}$ (mit $k > 0$). Bei abzählbar unendlich vielen Werten der Zufallsvariablen ist $\sum_{i=1}^{\infty} f_X(x) > 1$, da mit $f_X(x) = \frac{1}{k}$ bereits

$$\sum_{i=1}^{k+1} \frac{1}{k} = (k+1) \frac{1}{k} = 1 + \frac{1}{k} > 1.$$

z.B. $f_X(x) = 0.02 \implies k = \frac{1}{0.02} = 50$. Für 51 gilt nun: $51 \cdot 0.02 = 1.02$. Somit würden bereits 51 Ausprägungen zu einem Ergebnis führen, das grundlegenden Definitionen der Wahrscheinlichkeitstheorie widerspricht.

Bei abzählbar unendlich vielen Ausprägungen gilt im Falle der Zuordnung einer Konstanten $k > 0$ sogar $\sum_{i=1}^{\infty} \frac{1}{k} = \infty$.

Bemerkung 11.0.1. Man beachte dabei, dass die Summe abzählbar unendlich vieler positiver Zahlen nicht ins Unendliche zu wachsen braucht. So gilt etwa

$$\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{1}{6} \pi^2.$$

Ein weiteres Beispiel haben wir bei der Poissonverteilung gesehen, wo

$$\sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = 1$$

gilt, mit $e^{-\lambda} \frac{\lambda^x}{x!} > 0$ für alle $x \in \mathbb{N}$.

Deshalb können wir bei Zufallsvariablen, die überabzählbar viele Werte annehmen können, nicht den einzelnen Werten der Zufallsvariable eine positive Wahrscheinlichkeit zuordnen. Wir müssen einen anderen Weg verfolgen: wir ordnen Intervallen - statt den Werten der Zufallsvariable - positive Wahrscheinlichkeiten zu. Wir geben dabei nicht mehr an, wie hoch die Wahrscheinlichkeit ist, dass eine Schraube eine bestimmte Länge hat. Wir geben nunmehr die Wahrscheinlichkeit an, dass die Länge einer Schraube in einem spezifischen Intervall liegt. Um entsprechende Festlegungen vorzunehmen, gibt es im wesentlichen zwei Vorgehensweisen. Die erste besteht darin, die Wahrscheinlichkeitsverteilung mit Hilfe einer Verteilungsfunktion F_X zu definieren, wobei wir definieren $P_X([a, b]) = P(X \in [a, b]) = P(a < X \leq b) := F_X(b) - F_X(a)$. Dadurch sind dann Wahrscheinlichkeiten für linkshalboffene Intervalle definiert. Mit Hilfe der Wahrscheinlichkeiten, die linkshalboffenen Intervallen $I \subset \mathfrak{X}$ ($\mathfrak{X}$ = Bild der Zufallsvariable X) zugeordnet sind, sind dann auch die Wahrscheinlichkeiten, die beliebigen Vereinigungsmengen paarweise disjunkter solcher Intervalle und beliebigen Differenzmengen $\mathfrak{X} \setminus I$ zugeordnet sind, bestimmt. Diese erste Methode führen wir nun formeller ein.

11.1 Verteilungsfunktionen; Definition und Beispiele

Um eine Wahrscheinlichkeitsverteilung P_X einer Zufallsvariable X mit überabzählbar vielen möglichen Werten festzulegen, verwenden wir unter anderem Verteilungsfunktionen. Während bei diskreten Verteilungen Verteilungsfunktionen auf einer echten Teilmenge von $\mathbb{R}$ definiert sind (s. Seite 201, Definition 9.1.16), bei den behandelten Beispielen auf $\mathbb{N}$ oder auf Teilmengen von $\mathbb{N}$, definieren wir bei stetigen Verteilungen diese auf $\mathbb{R}$. Sie nehmen wie im diskreten Fall Werte zwischen 0 und 1 an. Sie sind ebenfalls monoton steigend, allerdings nicht unbedingt strikt monoton steigend, können also intervallweise konstante Teilstücke enthalten.

Definition 11.1.1. Ist $F_X : \mathbb{R} \rightarrow [0, 1]$ eine Verteilungsfunktion und ist F stetig, nennen wir die durch F_X definierte Wahrscheinlichkeitsverteilung $P_X([a, b]) := F_X(b) - F_X(a)$ stetig. Die Zufallsvariable X wird in diesem Fall „stetig verteilt“ genannt.

Bemerkung 11.1.2. (a) Intuitiv ist eine Funktion stetig, wenn man ihren Graphen in einem Zug in ein kartesisches Koordinatensystem zeichnen kann.

(b) Geht es z.B. um Schraubenlängen, so sind nur positive Längen möglich. Trotzdem können wir die Verteilungsfunktion auf $\mathbb{R}$ definieren. Wir setzen diese für $x < 0$ mit 0 gleich. Sollten in anderen Intervallen keine Werte der Zufallsvariablen fallen können, setzen wir die Verteilungsfunktion dort konstant.

Mit $P_X([a, b]) := F_X(b) - F_X(a)$ für $b > a$ ist die Wahrscheinlichkeit für beliebige Intervalle festgelegt. So ist z.B. $P(a < X \leq b) = F_X(b) - F_X(a)$. Konkret würde $P(3 < X \leq 5)$ für Schrauben z.B. die Wahrscheinlichkeit ausdrücken, dass eine Schraube eine Länge im Intervall $]3, 5]$ aufweist.

Beispiel 11.1.3. Wir nehmen an, die Verteilung von Schraubenlängen wäre durch die folgende Verteilungsfunktion bestimmt (in Zentimeter, s. Abbildung 11.1.1):

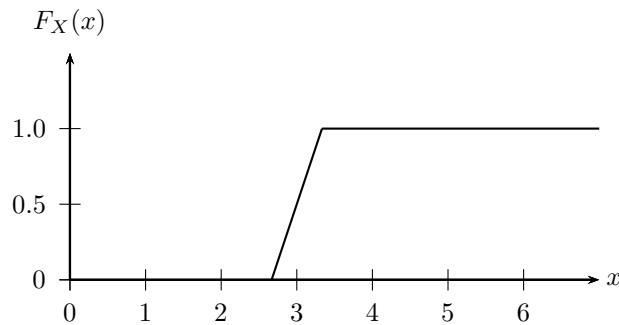
$$F_X(x) = \begin{cases} 1.5x - 4 & \text{für } \frac{8}{3} < x \leq \frac{10}{3} \\ 1 & \text{für } x > \frac{10}{3} \\ 0 & \text{für } x \leq \frac{8}{3} \end{cases}$$

Die Wahrscheinlichkeit $P(X \leq 3)$ (= die Wahrscheinlichkeit, dass eine zufällig gezogene Schraube kleiner oder gleich 3 cm lang ist), beträgt dann: $F_X(3) = 1.5 \cdot 3 - 4 = 0.5$

Die Wahrscheinlichkeit $P(3 < X \leq 3.2)$ (die Wahrscheinlichkeit, dass eine zufällig gezogene Schraube zwischen 3 und 3.2 cm lang ist) beträgt: $F_X(3.2) - F_X(3) = 1.5 \cdot 3.2 - 4 - (1.5 \cdot 3 - 4) = 0.3$

Die Wahrscheinlichkeit $P(X \leq 2)$ (die Wahrscheinlichkeit, dass eine zufällig gezogene Schraube kleiner als 2 cm lang ist) beträgt: $F_X(2) = 0$

Die Wahrscheinlichkeit $P(3.3 < X)$ (die Wahrscheinlichkeit, dass eine zufällig gezogene Schraube

Abbildung 11.1.1: Graphische Darstellung des Beispiels 11.1.3 - Verteilungsfunktion einer stetig verteilten Zufallsvariable X

grösser als 3.3 cm lang ist) beträgt: $1 - F_X(3.3) = 1 - (1.5 \cdot 3.3 - 4) = 0.05$

Die Wahrscheinlichkeit $P(2 < X)$ (die Wahrscheinlichkeit, dass eine zufällig gezogene Schraube grösser als 2 cm ist) beträgt: $1 - F_X(2) = 1 - 0 = 1$

Die Wahrscheinlichkeit $P(5 < X)$ (die Wahrscheinlichkeit, dass eine zufällig gezogene Schraube grösser als 5 cm ist) beträgt: $1 - F_X(5) = 1 - 1 = 0$ $\diamond$

Beispiel 11.1.4. Die Zeit bis zum ersten Ausfall einer Maschine wird als Wert einer Zufallsvariable betrachtet. Deren Verteilung sei durch die folgende Verteilungsfunktion gegeben (s. Abbildung 11.1.2):

$$F_X(x) = \begin{cases} 0.2x & \text{für } 0 < x \leq 5 \\ 1 & \text{für } x > 5 \\ 0 & \text{für } x \leq 0 \end{cases} \quad (\text{in Monaten})$$

Berechnen Sie die Wahrscheinlichkeit dafür, dass die Maschine in den ersten zwei Monaten ausfällt?

Berechnen Sie zudem:

b) $P(3.3 < X \leq 4)$

c) $P(X > 2.2)$

d) Zeichnen Sie die Verteilungsfunktion.

Lösung: a) $P(X \leq 2) = F_X(2) = 0.2 \cdot 2 = 0.4$

b) $P(3.3 < X \leq 4) = F_X(4) - F_X(3.3) = 0.2 \cdot 4 - 0.2 \cdot 3.3 = 0.14$

c) $P(X > 2.2) = 1 - F_X(2.2) = 1 - 0.2 \cdot 2.2 = 0.56$

d)

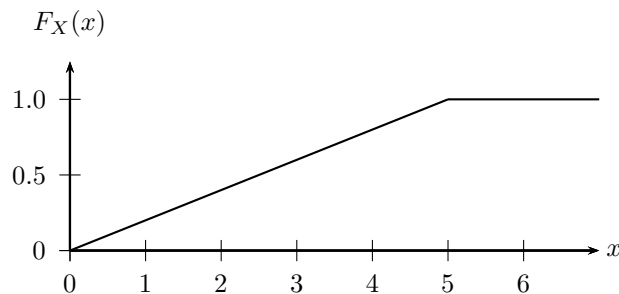


Abbildung 11.1.2: Graphische Darstellung des Beispiels 11.1.4

$\diamond$

Beispiel 11.1.5. Bei einer Produktion werden Güter einer spezifischen Art verwendet. Dabei können nur die Stücke mit einer Länge von 1 bis 14 verwendet werden. Die Verteilung der Länge der Stücke sei durch die folgende Verteilungsfunktion gegeben (s. Abbildung 11.1.3):

$$F_X(x) = \begin{cases} 0.005x^2 & \text{für } 0 < x \leq 14.1421356257 \\ 1 & \text{für } x > 14.1421356257 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Berechnen Sie die Wahrscheinlichkeit, dass X zwischen folgenden x_i und x_j liegt:

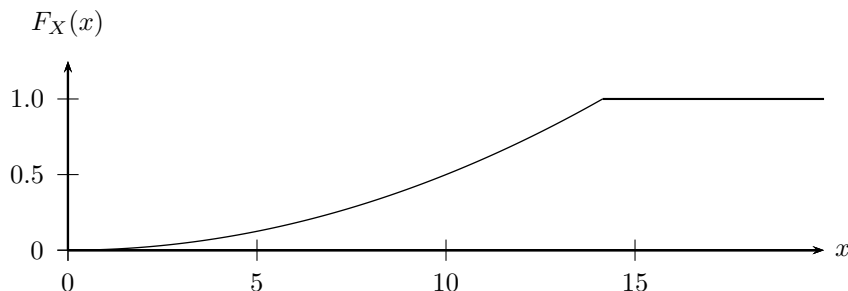


Abbildung 11.1.3: Graphische Darstellung des Beispiels 11.1.5

- a) 7 und 8
- b) -5 und -4
- c) 16 und 17
- d) Wieviel beträgt der Ausschuss?

◇

Lösung 11.1.6. a) $P(7 < X \leq 8) = F_X(8) - F_X(7) = 0.005(8)^2 - 0.005(7)^2 = 0.075$
 b) Zwischen -5 und -4 ist die Verteilungs-Funktion eine konstante Funktion mit $F_X = 0$, wobei gilt $0 - 0 = 0$.
 c) Zwischen 16 und 17 ist die Verteilungs-Funktion eine konstante Funktion mit $F_X = 1$, wobei gilt: $1 - 1 = 0$
 d) $P(X \leq 1) + P(X > 14) = 0.005 \cdot (1)^2 + 1 - 0.005 \cdot (14)^2 = 0.025$. Somit ist mit einem Ausschuss von 2.5% zu rechnen.

Eingangs wurde erwähnt, dass es zwei Möglichkeiten gibt, bei stetigen Wahrscheinlichkeitsräumen Wahrscheinlichkeitsverteilungen von Zufallsvariablen zu definieren. Wir wenden uns nun der zweiten Möglichkeit zu.

11.2 Die Dichtefunktion

Bei der Verteilungsfunktion gibt der Wert $F(x)$ die Wahrscheinlichkeit $P(X \leq x)$ an. Damit wird dem Intervall $]-\infty, x]$ eine Wahrscheinlichkeit zugeordnet. Bei der Dichtefunktion wird jedem Intervall $[a, b]$ unmittelbar eine Wahrscheinlichkeit zugeordnet, indem man dem Intervall die Fläche zwischen der x -Achse und der Dichtefunktion als Wahrscheinlichkeit zuordnet. Entsprechend muss die Fläche unter der Dichtefunktion im Intervall $]-\infty, \infty[$ 1 betragen. Deshalb könnte die folgende Funktion im

Beispiel 11.2.1.

$$f_X : \mathbb{R} \rightarrow \mathbb{R}^+ \\ f_X(x) = \begin{cases} 0 & \text{für } x \leq 2 \\ 0.5 & \text{für } 2 < x \leq 4 \\ 0 & \text{für } x > 4 \end{cases}$$

als Dichtefunktion aufgefasst werden ($\mathbb{R}^+ := [0, \infty[$) (s. Abbildung 11.2.4).

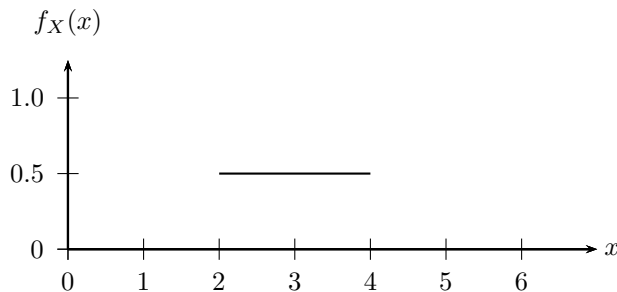


Abbildung 11.2.4: Graphische Darstellung des Beispiels 11.2.1

Die Fläche zwischen 2 und 4 und der Strecke im Abstand 0.5 in diesem Intervall beträgt nämlich: $(4 - 2)0.5 = 1$

Um die Wahrscheinlichkeit auszudrücken, dass ein Wert im Intervall $]a, b]$ liegt, können wir nun die Fläche zwischen f und der x -Achse im Intervall $]a, b]$ berechnen. So beträgt die Wahrscheinlichkeit, dass der Wert einer Zufallsvariable ins Intervall $]2, 3]$ fällt z.B. $(3 - 2)0.5 = 0.5$. Die Wahrscheinlichkeit, dass der Wert einer Zufallsvariable ins Intervall $]2.3, 3.7]$ fällt ist: $(3.7 - 2.3)0.5 = 0.7$.

Wir brauchen dieselbe Abkürzung für Dichtefunktionen und für Bildwahrscheinlichkeitsfunktionen. Der Zusammenhang macht klar, wovon die Rede ist. Es folgt ein weiteres Beispiel:

Beispiel 11.2.2. Eine Dichtefunktion sei durch

$$f(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ 0.5x & \text{für } x \in]0, 2] \\ 0 & \text{für } x > 2 \end{cases}$$

gegeben.

Die Fläche zwischen der Funktion und der x -Achse beträgt 1, da $2 \cdot 1 \cdot \frac{1}{2} = 1$ (Basis $\cdot$ Höhe $\cdot \frac{1}{2}$). Die Wahrscheinlichkeit, dass der Wert einer so verteilten Zufallsvariablen ins Intervall $]a, b]$ fällt - mit $a, b \in [0, 2]$ - ist dann $\frac{b \cdot 0.5b}{2} - \frac{a \cdot 0.5a}{2}$ (z.B. $P(X \in]0.6, 1.7]) = \frac{1.7 \cdot 0.85}{2} - \frac{0.6 \cdot 0.3}{2} = 0.6325$; Es wird ein Dreieck vom anderen abgezogen; s. Abbildung 11.2.5). .

◇

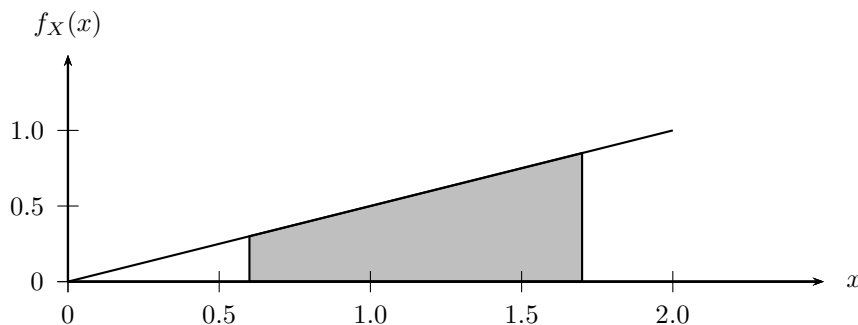


Abbildung 11.2.5: Graphische Darstellung des Beispiels 11.2.2 einer Dichtefunktion; die graue Fläche gibt die Wahrscheinlichkeit dafür an, dass die Zufallsvariable einen Werte zwischen 0.6 und 1.7 annimmt

Bevor wir zur Definition der Dichtefunktion übergehen, erläutern wir intuitiv die Symbolik $\int_b^a f(x)dx$ mit $f \geq 0$, $f : D \rightarrow \mathbb{R}$, $[a, b] \subset D \subset \mathbb{R}$. Sie bezeichnet die Fläche zwischen der x -Achse und der Funktion f im Intervall $[a, b]$, falls diese Fläche definiert ist. Diese Fläche ist jedenfalls definiert, falls f stetig und beschränkt auf D ist oder wenn es eine Zerlegung von D in Intervalle gibt, so dass f auf jedem dieser Intervalle bis auf die Intervallgrenzen stetig und beschränkt ist. f mit positiven Werten ist beschränkt, wenn es eine Schranke $a \in \mathbb{R}$ gibt, so dass $f(x) < a$ für alle $x \in \mathbb{R}$.

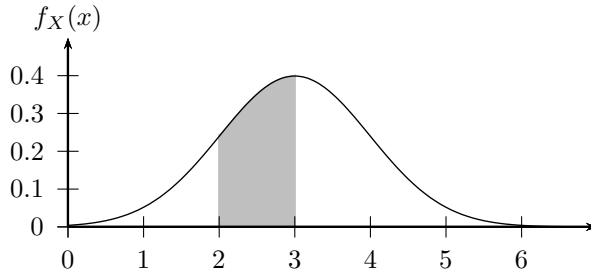


Abbildung 11.2.6: $\int_2^3 f(x)dx$ = die Fläche zwischen der x -Achse und der Funktion f im Intervall $[2, 3]$. Unter der Voraussetzung, dass die Fläche zwischen Kurve und x -Achse mit 1 identisch ist, gibt die graue Fläche die Wahrscheinlichkeit an, dass die Zufallsvariable einen Werte zwischen 2 und 3 annimmt.

Entsprechend bezeichnet $\int_{-\infty}^a f(x)dx$ die Fläche zwischen der Funktion f und der x -Achse im Intervall $] - \infty, a]$ und $\int_{-\infty}^{\infty} f(x)dx$ die Fläche zwischen der Funktion f und der x -Achse in $\mathbb{R}$.

Wir können nun die folgende Definition vornehmen:

Definition 11.2.3. Sei $f : \mathbb{R} \rightarrow \mathbb{R}$. f ist eine Dichtefunktion einer Zufallsvariable X genau dann, wenn

- (1) $f \geq 0$ und
- (2) die Fläche zwischen f und der x -Achse = 1 (d.h. $\int_{-\infty}^{\infty} f dx = 1$).

Die Dichtefunktion entspricht in etwa der Wahrscheinlichkeitsfunktion, nur dass den einzelnen Werten der Zufallsvariable bei Dichtefunktionen, die stetigen Verteilungen entsprechen, keine positive Wahrscheinlichkeit zugeordnet ist, wie wir gleich sehen werden.

Wie bei diskreten Verteilungs- und Wahrscheinlichkeitsfunktionen gibt es einen engen Zusammenhang von stetigen Verteilungsfunktionen und den entsprechenden Dichtefunktionen („ableitbar“ bedeutet, dass man die Steigung einer Funktion in allen Punkten berechnen kann).

Satz 11.2.1. Ist $F_X : \mathbb{R} \rightarrow [0, 1]$ eine ableitbare Verteilungsfunktion, dann gibt es eine Dichtefunktion $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$, so dass $F_X(t) = \int_{-\infty}^t f_X(x)dx$. (wobei $f_X = \frac{d}{dx}F_X$, d.h. f_X ist die Ableitung von F_X). (Ein Beweis würde einiges an Vorarbeit erfordern und damit zu weit führen).

Damit gilt für ableitbare Verteilungsfunktionen auf $\mathbb{R}$, dass es eine Dichtefunktion auf $\mathbb{R}$ gibt, so dass

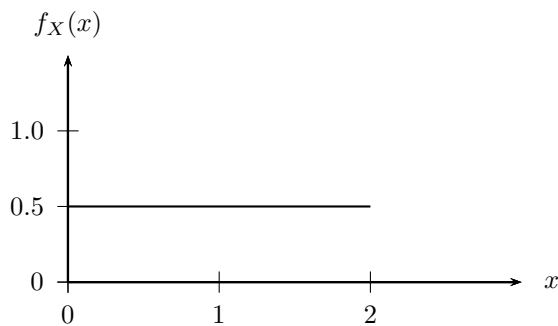
$$P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) = \int_{x_1}^{x_2} f_X(x)dx$$

für $x_1 < x_2 \in \mathbb{R}$.

Satz 11.2.2. Ist X eine stetig verteilte Zufallsvariable, dann ist $P(X = x) = 0$ für alle $x \in \mathbb{R}$.

Statt eines Beweises eine intuitive Begründung. Wir betrachten die Verteilungsfunktion F_X in ableitbaren Bereichen. Damit gibt es laut vorletztem Theorem in diesen Bereichen eine Dichtefunktion f_X , welche die einem Intervall zugeordnete Wahrscheinlichkeit als Fläche zwischen f_X und der x -Achse ausdrückt. Für $\{x\} = [x, x]$ gilt jedoch, dass die Basis dieser Fläche keine Ausdehnung hat. Entsprechend ist die Fläche über diesem Intervall mit 0 identisch. In nicht ableitbaren Punkten von F_X liegen Knickstellen von F_X vor (F_X ist ja als stetig vorausgesetzt!). Damit liegen bei der Dichtefunktion dort Sprungstellen vor. Auch dort ist die Fläche zwischen der Dichtefunktion und der x -Achse über einem Einpunktintervall mit 0 identisch.

Die Zusammenhänge können wir ohne Differential- und Integralrechnung an einem einfachen Beispiel klar machen. Sei $f(x) = 0.5$ für $0 < x \leq 2$ eine Dichtefunktion (mit $f(x) = 0$ für $x \leq 0$ und $f(x) = 0$ für $x > 2$). Die Fläche zwischen der Dichtefunktion und der x -Achse ist dann $0 + 0.5 \cdot 2 + 0 = 1$ (s. Abbildung 11.2.7)

Abbildung 11.2.7: $0, 2]$

Die Wahrscheinlichkeit $P(X \leq 0.6)$ ist die Fläche unter der Dichtefunktion im Intervall $]0, 0.6]$:
Somit gilt: $P(X \leq 0.6) = 0.5 \cdot 0.6 = 0.3$.

$P(0.3 < X \leq 0.7)$ ist die Fläche im Intervall $]0.3, 0.7]$. Somit gilt:

$$P(0.3 < X \leq 0.7) = 0.4 \cdot 0.5 = 0.2$$

$$P(0.4 < X) = 1 - (0.4 \cdot 0.5) = 0.8$$

Die entsprechende Verteilungsfunktion F_X erhalten wir aus der obigen Dichtefunktion wie folgt: Die Fläche unter der Dichtefunktion zwischen 0 und x ist im Intervall $]0, 2]$ $0.5 \cdot x$. Somit ist die entsprechende Verteilungsfunktion $F_X(x) = 0.5x$ für $0 < x \leq 2$ (und $F_X(x) = 0$ für $x \leq 0$, $F_X(x) = 1$ für $x > 2$; die Ableitung von $F_X(x) = 0.5x$ im offenen Intervall $]0, 2[$ ist durch $f = 0.5$ gegeben) (s. Abbildung 11.2.8)

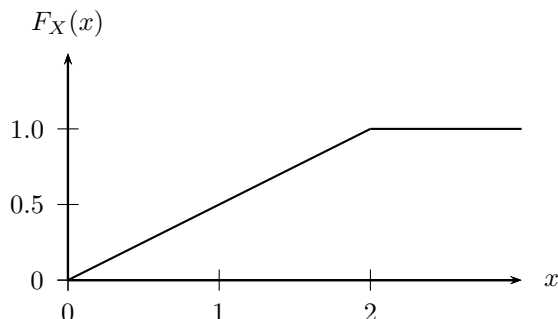


Abbildung 11.2.8: Graphische Darstellung des Beispiels der Verteilungsfunktion $F(x) = 0.5x$ für $x \in [0, 2]$, $f(x) = 1$ sonst, die der Dichtefunktion von Abbildung 11.2.7 entspricht

Wir überprüfen das Ergebnis an den obigen Beispielen:

$$P(X \leq 0.6) = F_X(0.6) = 0.5 \cdot 0.6 = 0.3$$

$$P(0.3 < X \leq 0.7) = F_X(0.7) - F_X(0.3) = 0.5 \cdot 0.7 - 0.5 \cdot 0.3 = 0.2$$

$$P(0.4 < X) = 1 - F_X(0.4) = 1 - 0.4 \cdot 0.5 = 0.8$$

Es ist ersichtlich, dass wir dieselben Ergebnisse erhalten, ob wir nun die Dichtefunktion oder die entsprechende Verteilungsfunktion verwenden. Bei komplizierteren Dichtefunktionen (Logarithmusfunktionen, Polynome höheren Grades, Exponentialfunktionen), können wir die Fläche nicht mehr mit klassischer Geometrie berechnen. Wir brauchen dazu Integralrechnung oder Näherungsverfahren.

Da bei stetigen Verteilungen den einzelnen Werten die Wahrscheinlichkeit 0 zugeordnet wird, spielt es bei diesen keine Rolle, ob wir schreiben: $P(X \leq x)$ oder $P(X < x)$ (respektive: $P(X \geq x)$ oder $P(X > x)$). Im Allgemeinen halten wir jedoch die bisher verwendete Schreibweise bei.

Die Dichtefunktion wird verwendet, weil es

- (1) nicht für alle wichtigen Verteilungsfunktionen geschlossen berechenbare Formeln gibt. So ist z.B. die später zu behandelnde Normalverteilung nur als Dichtefunktion in einer geschlossenen, berechenbaren Formel angebar, während die Werte der Verteilungsfunktion nur mit Hilfe von Näherungszahlen (z.B. in Tabellen; Excel; Taschenrechner) angegeben werden.
- (2) Um die Kennzahlen wie Erwartungswert, Varianz und Standardabweichung für stetige Wahrscheinlichkeitsverteilungen zu berechnen, verwenden wir i.A. die Dichtefunktion.
- (3) Die Dichtefunktion ist recht anschaulich.

Beispiel 11.2.4. Gegeben sei die folgende Dichtefunktion:

$$f(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ 0.2 & \text{für } x \in]0, 5] \\ 0 & \text{für } x > 5 \end{cases}$$

Zeigen Sie:

- a) Die Fläche unter der Dichtefunktion ist mit 1 identisch.
- b) Berechnen Sie mit Hilfe der Dichtefunktion die Wahrscheinlichkeiten:

$$P(X \leq 4)$$

$$P(3 < X \leq 4.5)$$

$$P(0.5 < X \leq 0.6)$$

$$P(X > 3)$$

- c) Berechnen Sie die Verteilungsfunktion, die der Dichtefunktion entspricht und berechnen Sie nach der Verteilungsfunktion:

$$P(X \leq 4)$$

$$P(3 < X \leq 4.5)$$

$$P(0.5 < X \leq 0.6)$$

$$P(X > 3)$$

◇

Lösung 11.2.5. a) $0.2 \cdot 5 = 1$

$$b) P(X \leq 4) = 4 \cdot 0.2 = 0.8$$

$$P(3 < X \leq 4.5) = 1.5 \cdot 0.2 = 0.3$$

$$P(0.5 < X \leq 0.6) = 0.1 \cdot 0.2 = 0.02$$

$$P(X > 3) = 1 - 3 \cdot 0.2 = 0.4 = 2 \cdot 0.2$$

c) Die Fläche als Funktion von $x = 0.2x$. Somit ist $F_X(x) = 0.2x$ im Intervall $]0, 5]$, $F_X(x) = 0$ für $x \leq 0$ und $F_X(x) = 1$ für $x > 5$.

$$P(X \leq 4) = F_X(4) = 0.2 \cdot 4 = 0.8$$

$$P(3 < X \leq 4.5) = F_X(4.5) - F_X(3) = 0.2 \cdot 4.5 - 0.2 \cdot 3 = 0.3$$

$$P(0.5 < X \leq 0.6) = F_X(0.6) - F_X(0.5) = 0.2 \cdot 0.6 - 0.2 \cdot 0.5 = 0.02$$

$$P(X > 3) = 1 - F_X(3) = 1 - 0.2 \cdot 3 = 0.4$$

11.3 Kennzahlen

Der *Median* einer stetigen Verteilung wird definiert als der Punkt, der die Fläche unter der Dichtefunktion in zwei gleiche Hälften teilt. In den Fällen, die für uns von Interesse sind, ist dieser Punkt immer eindeutig bestimmt (d.h. die Dichtefunktion ist im interessierenden Bereich von 0 verschieden). Somit liegt links vom Median 0.5 der Fläche. Dies ist äquivalent mit der folgenden Definition:

Definition 11.3.1. Der Median einer stetig verteilten Zufallsvariable X ist x mit $F_X(x) = 0.5$ für im von 0 und 1 verschiedenen Bereich strikt monoton steigendes F_X .

Betrachten wir ein Beispiel: $F_X(x) = 0.2x$ in $]0, 5]$, sonst 1 für $x > 5$ und 0 für $x \leq 0$. Wir setzen $F_X(x) = 0.2x = 0.5 \implies x = 2.5$. 2.5 ist der Median.

Definition 11.3.2. Das *p-Quantil* einer stetig verteilten Zufallsvariable X ist x mit $F_X(x) = p$ für im von 0 und 1 verschiedenen Bereich strikt monoton steigende Verteilungsfunktionen F_X .

So ist das 0.3-Quantil von $F_X(x) = 0.2x$ in $]0, 5]$, sonst 1 für $x > 5$ und 0 für $x \leq 0$ das x in

$$\begin{aligned} 0.2x &= 0.3 \implies \\ x &= 1.5 \end{aligned}$$

Definition 11.3.3. Der *Erwartungswert* einer stetigen P -verteilten Zufallsvariable X wird wie folgt definiert:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Dabei ist f eine Dichtefunktion der P -verteilten Zufallsvariable X . Die Ähnlichkeit mit der bisherigen Definition des Erwartungswertes springt ins Auge. In Bezug auf diskrete Verteilungen haben wir den Erwartungswert definiert als

$$\sum_{x \in \mathfrak{X}} x P(X = x) = \sum_{x \in \mathfrak{X}} x f_X(x).$$

Das Symbol fürs Integral, das man als Summenbildung einer spezifischen Art auffassen könnte, hat das Symbol für unendliche oder endliche Summenbildung abgelöst. Die Werte der Wahrscheinlichkeitsfunktion ($f_X(x) = P(X = x)$) werden durch die Werte der Dichtefunktion ersetzt.

Bei symmetrischen Dichtefunktionen fällt der Median mit dem Erwartungswert zusammen. Die Dichtefunktion des obigen Beispiels ist symmetrisch. Entsprechend haben wir dort mit dem Median auch den Erwartungswert berechnet.

Definition 11.3.4. Die *Varianz* einer stetig verteilten Zufallsvariable wird definiert als:

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

Auch hier sollte man die Ähnlichkeit mit der Definition der Varianz für diskrete Verteilungen im Auge behalten:

$$\sum_{x \in \mathfrak{X}} (x - E(X))^2 P(X = x) = \sum_{x \in \mathfrak{X}} (x - E(X))^2 f_X(x).$$

Wie üblich ist die Standardabweichung definiert als die Wurzel aus der Varianz.

Übungen

1. In $\{x|0 < F_X(x) < 1\}$ sei $F_X(x) := 0.02x - 5$. Berechnen Sie das Intervall, in dem die Funktion strikt monoton steigt. Berechnen Sie den Median, das 0.3-Quantil und das 3. Quartil. Berechnen Sie den Erwartungswert.
2. In $\{x|0 < F_X(x) < 1\}$ sei $F_X(x) := 0.2x^2 - 5$ gegeben. Berechnen Sie das Intervall, in dem die Funktion strikt monoton steigt. Berechnen Sie den Median, das 0.3-Quantil und das 3. Quartil.

11.3.1 Lösungen

1. $0.02x - 5 = 0 \implies x = 250$; $0.02x - 5 = 1 \implies x = 300$. Das Intervall mit strikt monoton steigender Verteilungsfunktion ist damit $]250, 300[$; Median: $F_X(x) = 0.02x - 5 = 0.5 \implies x = 275$; 0.3-Quantil: $F_X(x) = 0.02x - 5 = 0.3 \implies x = 265$; 3. Quartil: $F_X(x) = 0.02x - 5 = 0.75 \implies x = 287.5$; Erwartungswert = Median (Symmetrie)
2. $0.2x^2 - 5 \implies x_1 = -5$; $x_2 = 5$. Eine Verteilungsfunktion muss monoton steigend sein. Links von 0 ist die vorliegende Funktion sinkend. Entsprechend ist nur der Bereich rechts von 0 in Betracht zu ziehen.
 $0.2x^2 - 5 = 1 \implies x = 5.4772$; Damit ist der strikt monoton steigende Bereich $]5, 5.4772[$.
 Median: $F_X(x) = 0.2x^2 - 5 = 0.5 \implies x = 5.244$ 0.3-Quantil: $F_X(x) = 0.2x^2 - 5 = 0.3 \implies x = 5.1478$ 3. Quartil: $F_X(x) = 0.2x^2 - 5 = 0.75 \implies x = 5.3619$; Der Erwartungswert kann hier nicht mittels Median berechnet werden, da die Dichtefunktion nicht achsensymmetrisch ist (diese ist nämlich $0.4x$ für $x \in]5, 5.4772[$).

11.4 Übersicht stetige Wahrscheinlichkeitsräume

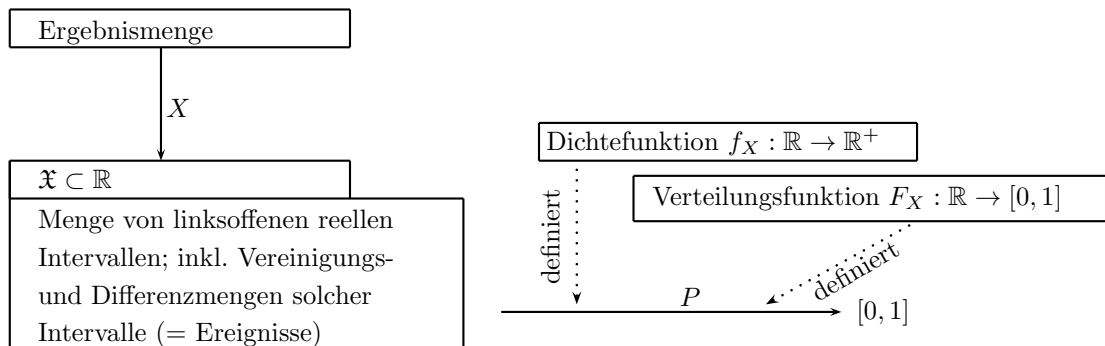


Abbildung 11.4.9: $\cdot X$ = Zufallsvariable

- Definition von P mit Hilfe der Dichtefunktion durch Fläche zwischen Dichtefunktion und x-Achse im Intervall $]a, b]$
- Definition von P mit Hilfe einer stetigen Verteilungsfunktion: $F_X(b)$ drückt die Wahrscheinlichkeit aus, dass ein Wert der Zufallsvariable in das Intervall $] - \infty, b]$ fällt. $F_X(b) - F_X(a)$ drückt Wahrscheinlichkeit, aus, dass der Wert der Zufallsvariablen in das Intervall $]a, b] = [a, b]$ fällt.

11.5 Lernziele

- Die Begründung dafür angeben können, dass bei stetigen Verteilungen den einzelnen Werten der Zufallsvariable die Wahrscheinlichkeit 0 zugeordnet wird.

- Die definierenden Bedingungen einer stetigen Verteilung sowie einer stetig verteilten Zufallsvariable angeben können.
- Die Wahrscheinlichkeiten, welche Intervallen mit Hilfe einer Verteilungsfunktion zugeordnet sind, berechnen können.
- Die Definition der Dichtefunktion angeben können. Die Gründe für die Einführung von Dichtefunktionen kennen.
- Den Zusammenhang von Dichtefunktionen und stetigen Verteilungsfunktionen erläutern können (Bevor der entsprechende Stoff in der Mathematik behandelt wurde, ohne die Integral- und Differentialsymbolik. Es genügt zu wissen, dass man mit Hilfe der Dichtefunktion und der stetigen Verteilungsfunktion jedem Intervall $]a, b]$ dieselbe Wahrscheinlichkeit zuordnet und wie diese Zuordnung im Prinzip erfolgt: bei der Dichtefunktion durch die Fläche zwischen der Funktion und der x -Achse im Intervall $]a, b]$; bei der Verteilungsfunktion durch die Differenz $F_X(b) - F_X(a)$).
- In Beispielen, die mit einfacher Geometrie zu berechnen sind, Wahrscheinlichkeiten berechnen können, die Intervallen mit Hilfe einer Dichtefunktion zugeordnet sind.

Kapitel 12

Stetige Verteilungsfamilien

Wie bei den diskreten Verteilungen verdienen auf Grund ihrer praktischen Bedeutung bei den stetigen Verteilungen bestimmte Familien von Verteilungen einen eigenen Namen. Wir werden einige dieser Familien in der Folge genauer untersuchen.

12.1 Die Familie der stetigen, uniformen Verteilungen

Wir haben bisher bereits einige Beispiele von stetigen, uniformen Verteilungen kennengelernt: in einem Intervall $[a, b]$ wird jedem Intervall $I \subset [a, b]$ derselben Länge dieselbe Wahrscheinlichkeit zugeordnet. Die uniforme Verteilung ist dadurch gekennzeichnet, dass ihre Dichtefunktion im interessierenden Intervall eine konstante Funktion ist mit $f_X(x) \neq 0$ oder dass die Verteilungsfunktion im interessierenden Intervall eine Gerade $F_X(x) = dx + e$ ist (f_X sei die Dichtefunktion und F_X die Verteilungsfunktion der Zufallsvariable X):

Definition 12.1.1. Eine Abbildung $f_X : \mathbb{R} \rightarrow \mathbb{R}$ ist die Dichtefunktion einer stetigen, uniformen Verteilung genau dann, wenn es ein $a, b \in \mathbb{R}$ gibt, so dass $b > a$ und

$$f_X(x) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{1}{b-a} & \text{für } a < x \leq b \\ 0 & \text{für } x > b \end{cases}$$

Für uniform verteilte Zufallsvariablen schreiben wir:

$$X \sim U(a, b).$$

- gelesen als: die Zufallsvariable X ist uniform verteilt mit den Intervallgrenzen a und b oder: die Zufallsvariable X ist uniform verteilt mit den Parametern a und b . Die Ausdrücke in den Klammern werden als „Parameter“ der Verteilung bezeichnet. Die Fläche zwischen der x-Achse und der Dichtefunktion ergibt 1, da diese Fläche in den Bereichen, wo die Dichtefunktion 0 ergibt, 0 ist. Im Intervall $]a, b]$ ist die Fläche ein Rechteck, dessen Basis $b-a$ und dessen Höhe $\frac{1}{b-a}$ beträgt. Die Fläche ist das Produkt dieser zwei Größen: $(b-a)\frac{1}{b-a} = 1$.

Satz 12.1.2. Die Verteilungsfunktion einer uniformen stetigen Verteilung $U(a, b)$ ist gegeben durch :

$$F_X : \mathbb{R} \rightarrow \mathbb{R}$$
$$F_X(x) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{1}{b-a}(x-a) & \text{für } x \in]a, b] \\ 1 & \text{für } x > b, \end{cases}$$

Beweis. Die der Verteilung $U(a, b)$ entsprechende Dichtefunktion ist gemäss Definition 12.1.1

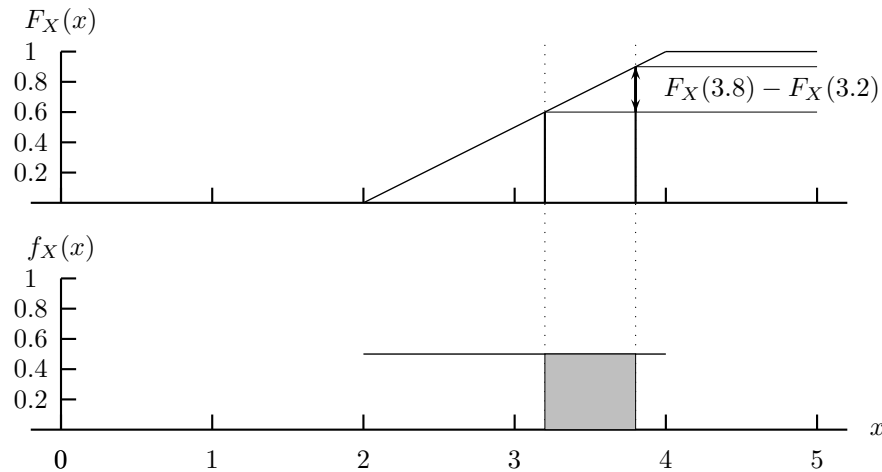
$$f_X = \begin{cases} 0 & \text{für } x \leq a \\ \frac{1}{b-a} & \text{für } a < x \leq b \\ 0 & \text{für } x > b \end{cases}$$

mit $a, b \in \mathbb{R}$ und $a < b$. Gemäss Definition ist $F_X(x) = P(X \leq x)$, wobei $P(X \leq x)$ mit der Fläche zwischen der Dichtefunktion und der x -Achse im Intervall $]-\infty, x]$ identisch ist.

(i) Für $x \in]-\infty, a]$ ist diese Fläche mit 0 identisch, da dort $f_X(x) = 0$. Entsprechend gilt $F_X(x) = P(X \leq x) = 0$ für $x \in]-\infty, a]$.

(ii) Für $x \in]a, b]$ ist diese Fläche mit $(x - a) \cdot \frac{1}{b-a}$ identisch (Im Bereich $]-\infty, a]$ ist die Fläche 0, im Bereich $]a, x]$ ist die Fläche ein Rechteck der Basis $(x - a)$ und der Höhe $\frac{1}{b-a}$). Damit gilt dort $F_X(x) = P(X \leq x) = (x - a) \cdot \frac{1}{b-a}$.

(iii) Für $x \in]b, \infty[$ gilt, $F_X(x) = P(X \leq x) = 1$, denn die Fläche zwischen der Dichtefunktion und der x -Achse im Intervall $]-\infty, b]$ ist gemäss (ii) mit 1 identisch, da $(b - a) \cdot \frac{1}{b-a} = 1$. Zudem ist die Fläche zwischen der Dichtefunktion und der x -Achse im Intervall $]b, x]$ mit 0 identisch, da dort $f_X(x) = 0$. Entsprechend gilt $F_X(x) = P(X \leq x) = 1$ für $x \in]b, \infty[$. $\square$



Abbildungung 12.1.1: Dichtefunktion (unten) und Verteilungsfunktion einer uniform verteilten Zufallsvariable mit den Parametern 2 und 4. Die Strecke $F_X(3.8) - F_X(3.2)$ ($=P(3.8 < X \leq 3.2)$) ist mit der Fläche $f_X(3.2)(3.8 - 3.2)$ identisch

Die uniforme Verteilungsfunktion wird in der „Wirklichkeit“ selten angetroffen. Uniform verteilte Zufallszahlen können wir aber dazu verwenden, um anders verteilte Zufallszahlen zu gewinnen. Zudem stellen sie ein besonders einfaches Beispiel für stetig verteilte Zufallsvariablen dar.

Satz 12.1.3. *Der Median einer uniform verteilten Zufallsvariable $X \sim U(a, b)$ ist:*

$$\text{median}_X = \frac{a + b}{2}$$

Beweis. Sei F_X die Verteilungsfunktion von $X \sim U(a, b)$. Laut obigem Theorem gilt dann: $F_X(x) = \frac{1}{b-a}(x - a)$ im Intervall $]a, b]$. Für den Median gilt: $F_X(x) = \frac{1}{b-a}(x - a) = 0.5$. Wir lösen nach x auf: $(x - a) = 0.5(b - a) \implies x = 0.5(b - a) + a = \frac{b+a}{2}$. $\square$

Der Erwartungswert $E(X)$ ist mit dem Median identisch, da die Dichtefunktion symmetrisch bezüglich zu y parallelen Geraden durch $\frac{a+b}{2}$ ist.

Die Varianz $V(X)$ einer uniform verteilten Zufallsvariable $X \sim U(a, b)$ ist:

$$\int_a^b f_X(x)(x - E(X))^2 dx = \frac{(b-a)^2}{12}$$

Denn:

$$\begin{aligned} V(X) &= \int_a^b \left(\frac{1}{b-a} (x - E(X))^2 \right) dx = \frac{1}{b-a} \int_a^b (x - E(X))^2 dx \\ &= \frac{1}{b-a} \int_a^b (x^2 - 2xE(X) + E^2(X)) dx \\ &= \frac{x^3}{3} - x^2 E(X) + x E^2(X) \Big|_a^b \frac{1}{b-a} \\ &= \left(\frac{1}{3} b^3 - b^2 E(X) + b E^2(X) - \frac{1}{3} a^3 + a^2 E(X) - a E^2(X) \right) \frac{1}{b-a} \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Beweis. der letzten Gleichung:

Da $\frac{1}{b-a} (b-a) = 1$ und $E(X) = \frac{b+a}{2}$ erhalten wir:

$$\begin{aligned} &\left(\frac{1}{3} b^3 - b^2 \frac{a+b}{2} + b \left(\frac{a+b}{2} \right)^2 - \frac{1}{3} a^3 + a^2 \frac{a+b}{2} - a \left(\frac{a+b}{2} \right)^2 \right) \frac{1}{b-a} = \\ &\left(\frac{1}{3} b^3 - b^2 \frac{a+b}{2} + b \frac{a^2+2ab+b^2}{4} - \frac{1}{3} a^3 + a^2 \frac{a+b}{2} - a \frac{a^2+2ab+b^2}{4} \right) \frac{1}{b-a} = \\ &\left(\frac{1}{3} b^3 - \frac{ab^2+b^3}{2} + \frac{a^2b+2ab^2+b^3}{4} - \frac{1}{3} a^3 + \frac{a^3+a^2b}{2} - \frac{a^3+2a^2b+ab^2}{4} \right) \frac{1}{b-a} = \\ &\left(\frac{4b^3}{12} - \frac{6ab^2+6b^3}{12} + \frac{3a^2b+6ab^2+3b^3}{12} - \frac{4a^3}{12} + \frac{6a^3+6a^2b}{12} - \frac{3a^3+6a^2b+3ab^2}{12} \right) \frac{1}{b-a} = \\ &\left(\frac{b^3}{12} - \frac{b^2a}{4} + \frac{ba^2}{4} - \frac{a^3}{12} \right) \frac{1}{b-a} = \\ &(b^3 - 3b^2a + 3ba^2 - a^3) \frac{1}{12(b-a)} = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12} \end{aligned}$$

□

Satz 12.1.4. Für $X \sim U(0, 1)$ und $x \in]0, 1]$ gilt: $P(X \leq x) = x$

Beweis. Wir erhalten laut Definition die Dichtefunktion $f_X(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ 1 & \text{für } 0 < x \leq 1 \\ 0 & \text{für } x > 1 \end{cases}$

f_X ist eine Dichtefunktion, da $(1-0) \cdot 1 = 1$.

Wir betrachten das interessierende Intervall $]0, 1]$

$$P(0 < X \leq x) = P(X \leq x) = F_X(x) \stackrel{\text{siehe Satz 12.1.2}}{=} \frac{1}{1-0} (x-0) = x$$

□

12.1.1 Übungen

1. (a) Berechnen Sie eine uniforme Dichtefunktion mit $f(x) > 0$ für $x \in]0, 20]$, sonst $f(x) = 0$.
- (b) Geben Sie die entsprechende Verteilungsfunktion an.
- (c) Berechnen Sie mit Hilfe der Dichtefunktion die folgenden Wahrscheinlichkeiten:
 - $P(X \leq 19)$
 - $P(18 < X \leq 19)$
 - $P(18.999 < X \leq 19)$
 - $P(0.1 < X \leq 1)$
 - $P(X > 5)$
 - $P(X > 0.1)$

- (d) Berechnen Sie dieselben Wahrscheinlichkeiten mit Hilfe der Verteilungsfunktion.
 - (e) Berechnen Sie den Median
 - (f) Berechnen Sie den Erwartungswert.
 - (g) Berechnen Sie die Varianz und die Standardabweichung.
2. Berechnen Sie für eine uniforme Verteilung mit $f(x) > 0$ für $x \in]-20, 20]$ (sonst $f(x) = 0$).
- (a) die Dichtefunktion.
 - (b) Geben Sie die entsprechende Verteilungsfunktion an.
 - (c) Berechnen Sie mit Hilfe der Dichtefunktion die folgenden Wahrscheinlichkeiten:
 $P(X \leq 19)$
 $P(-18 < X \leq 19)$
 $P(18.999 < X \leq 19)$
 $P(-1.1 < X \leq -1)$
 $P(X > 5)$
 $P(X > -0.1)$
 - (d) Berechnen Sie dieselben Wahrscheinlichkeiten mit Hilfe der Verteilungsfunktion.
 - (e) Berechnen Sie den Median
 - (f) Berechnen Sie den Erwartungswert.
 - (g) Berechnen Sie die Standardabweichung und die Varianz.

12.1.2 Lösungen

1. (a) Die Fläche im Intervall muss 1 geben: Somit gilt: $20 \cdot x = 1 \implies x = 0.05$. Die Dichtefunktion ist somit fürs Intervall: $f_X(x) = 0.05$.
- (b) Die Fläche unter der Dichtefunktion beträgt (als Funktion von x) $0.05x$. Somit ist die Verteilungsfunktion im Intervall: $F_X(x) = 0.05x$.
- (c) $P(X \leq 19) = 19 \cdot f_X(19) = 19 \cdot 0.05 = 0.95$
 $P(18 < X \leq 19) = 1 \cdot 0.05 = 0.05$
 $P(18.999 < X \leq 19) = 0.001 \cdot 0.05 = 0.00005$
 $P(0.1 < X \leq 1) = 0.9 \cdot 0.05 = 0.045$
 $P(X > 5) = 15 \cdot 0.05 = 0.75 = 1 - (5 \cdot 0.05)$
 $P(X > 0.1) = 1 - (0.1 \cdot 0.05) = 0.995$
- (d) $P(X \leq 19) = F_X(19) = 0.05 \cdot 19 = 0.95$
 $P(18 < X \leq 19) = F_X(19) - F_X(18) = 19 \cdot 0.05 - 18 \cdot 0.05 = 0.05$
 $P(18.999 < X \leq 19) = F_X(19) - F_X(18.999) = 19 \cdot 0.05 - 18.999 \cdot 0.05 = 0.00005$
 $P(0.1 < X \leq 1) = F_X(1) - F_X(0.1) = 1 \cdot 0.05 - 0.1 \cdot 0.05 = 0.045$
 $P(X > 5) = 1 - F_X(5) = 1 - 5 \cdot 0.05 = 0.75$
 $P(X > 0.1) = 1 - F_X(0.1) = 1 - (0.1 \cdot 0.05) = 0.995$
- (e) der Median x ist: $F_X(x) = 0.05x = 0.5 \implies x = 10$
- (f) der Erwartungswert = Median, da uniforme Dichten symmetrisch sind.
- (g) $V(X) = \frac{(20-0)^2}{12} = 33.333$
 $\sqrt{V(X)} = \sqrt{33.333} = 5.7735$

Wir stellen fest, dass im Intervall $[E(X) - \sqrt{V(X)}, E(X) + \sqrt{V(X)}] = [10 - 5.7735, 10 + 5.7735]$ die folgende Fläche unter der Dichtefunktion liegt: $0.05 \cdot 2 \cdot 5.7735 = 0.57735$.
 (s. Abbildung [12.1.2](#))

2. Man erhält:

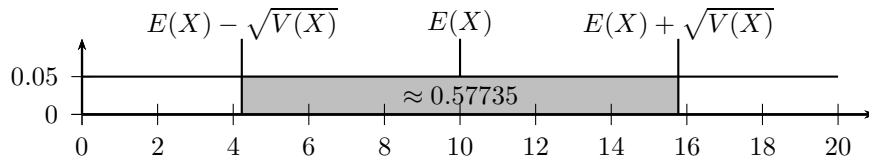


Abbildung 12.1.2: Dichtefunktion von $U(0,20)$ mit der Fläche, die durch $E(X) \pm \sqrt{V(X)}$ bestimmt ist

- (a) Die Fläche beträgt $40 \cdot h = 1 \implies h = \frac{1}{40} = 0.025 \implies f_X(x) = 0.025$
 (b) Die Funktion verläuft durch die Punkte $(-20,0)$ und $(20,1)$. Wir berechnen eine Gerade durch diese Punkte:

$$0 = -20a + b$$

$$1 = 20a + b$$

$$a = \frac{1}{40}, b = \frac{1}{2} \implies F_X(x) = 0.025x + 0.5 \text{ (fürs Intervall }]-20, 20])$$

- (c) $P(X \leq 19) = 39 \cdot 0.025 = 0.975$
 $P(-18 < X \leq 19) = 37 \cdot 0.025 = 0.925$
 $P(18.999 < X \leq 19) = 0.001 \cdot 0.025 = 2.5 \times 10^{-5}$
 $P(-1.1 < X \leq -1) = 0.1 \cdot 0.025 = 0.0025$
 $P(X > 5) = 15 \cdot 0.025 = 0.375$
 $P(X > -0.1) = 20.1 \cdot 0.025 = 0.5025$
 (d) $P(X \leq 19) = F_X(19) = 0.025 \cdot 19 + 0.5 = 0.975$
 $P(-18 < X \leq 19) = F_X(19) - F_X(-18) = 0.025 \cdot 19 + 0.5 - (0.025 \cdot (-18) + 0.5) = 0.925$
 $P(18.999 < X \leq 19) = F_X(19) - F_X(18.999) =$
 $0.025 \cdot 19 + 0.5 - (0.025 \cdot (18.999) + 0.5) = 2.5 \times 10^{-5}$
 $P(-1.1 < X \leq -1) = F_X(-1) - F_X(-1.1) = 0.025 \cdot (-1) + 0.5 - (0.025 \cdot (-1.1) + 0.5) = 0.0025$
 $P(X > 5) = 15 \cdot 0.025 = 1 - F_X(5) = 1 - (0.025 \cdot (5) + 0.5) = 0.375$
 $P(x > -0.1) = 1 - F_X(-0.1) = 1 - (0.025 \cdot (-0.1) + 0.5) = 0.5025$
 (e) Der Median x von $F_X = 0.025x + 0.5 = 0.5 \implies x = 0$
 (f) der Erwartungswert = der Median.
 (g) $V(X) = \frac{(20+20)^2}{12} = 133.33$
 $\sqrt{V(X)} = \sqrt{133.3333} = 11.547$

Wir stellen fest, dass im Intervall $[E(X) - \sqrt{V(X)}, E(X) + \sqrt{V(X)}] = [0 - 11.547, 0 + 11.547]$ die folgende Fläche unter der Dichtefunktion liegt: $0.025 \cdot 2 \cdot 11.547 = 0.57735$

Es handelt sich um dieselbe Zahl wie unter Übung 1. Die Fläche, die von Standardabweichung aus der Fläche unter der Dichtefunktion herausgeschnitten wird, ist bei uniformen Verteilungsfunktionen immer mit 0.57735 identisch, da $2\sqrt{V(X)} \frac{1}{b-a} = 2\sqrt{\frac{(b-a)^2}{12}} \frac{1}{b-a} = 2 \frac{(b-a)}{\sqrt{12}} \frac{1}{b-a} = \frac{2}{\sqrt{12}} = 0.57735$. Grob geschätzt beträgt somit bei uniformen Verteilungen die Wahrscheinlichkeit, dass der Wert von X im Bereiche der Standardabweichung um den Erwartungswert liegt etwas weniger als $\frac{3}{5}$.

12.2 Familie der Exponentialverteilungen

Weitere, einfache Verteilungsmodelle stellen Exponentialverteilungen dar. Sie dienen etwa dazu, Ereignisse in der Zeit zu modellieren (z.B. die Lebensdauer von Glühlampen, von Maschinen, von Gegenständen, die wegen Ausfall dem Produzenten zurückgeschickt werden).

Definition 12.2.1. F_X ist die Verteilungsfunktion einer Exponentialverteilung genau dann, wenn

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } 0 \leq x; \lambda > 0 \\ 0 & \text{für } x < 0 \end{cases}$$

Abkürzung: $X \sim \text{Exp}(\lambda)$

λ ist der *Parameter* der Verteilungsfamilie.

Auf die Erwähnung von „0 für $x < 0$ “ verzichten wir künftig gewöhnlich.

Bemerkung 12.2.2. Exponentialverteilungen sind durch ihre „Gedächtnislosigkeit“ gekennzeichnet. Wir definieren „Wahrscheinlichkeitsrate“ wie folgt: $\frac{P([a, a+t])}{P([a, \infty[)}$. Es ist das Verhältnis der Wahrscheinlichkeit, dass die Zufallsvariable einen Wert in $]a, a+t]$ zur Wahrscheinlichkeit, dass die Zufallsvariable einen Wert in $]a, \infty[$ annimmt. Die Wahrscheinlichkeitsraten für das Eintreffen von Ereignissen sind für beliebige Intervalle der gleichen Grösse - unabhängig von deren Lage auf der Zeitachse - bezüglich einer Exponentialverteilung mit dem Parameter λ konstant, d.h.

$$\frac{P([a, a+t])}{P([a, \infty[)} = \frac{P([a_1, a_1+t])}{P([a_1, \infty[)}$$

(für $a, a_1 \in \mathbb{R}^+$). Es gilt nämlich:

$$\frac{P([a, a+t])}{P([a, \infty[)} = \frac{1 - e^{-\lambda(a+t)} - (1 - e^{-\lambda a})}{1 - (1 - e^{-\lambda a})} = 1 - e^{-\lambda t}.$$

Damit ist $\frac{P([a, a+t])}{P([a, \infty[)}$ nur von der Länge t des Intervalls abhängig, nicht jedoch vom Ort a , wo dieses Intervall beginnt. Man kann beweisen, dass nur die Exponentialverteilung dieser Bedingung genügt. $\diamond$

Für $\lambda = 0.5$ erhalten wir die folgende graphische Darstellung des Graphen der Verteilungsfunktion (s. Abbildung 12.2.3):

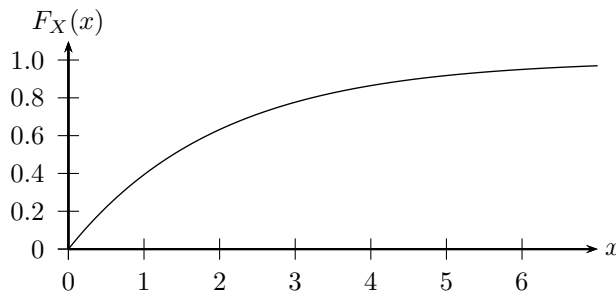


Abbildung 12.2.3: Graphische Darstellung der Verteilungsfunktion F der Exponentialverteilung mit dem Parameter $\lambda = 0.5$

Wir zeigen, dass $F_X(x) = 1 - e^{-\lambda x}$ eine Verteilungsfunktion ist, indem wir beweisen, dass gilt:

Satz 12.2.3. (1) $1 - e^{-\lambda x}$ ist monoton steigend auf $]0, \infty[$.

(2) $0 \leq 1 - e^{-\lambda x} \leq 1$ ($\lambda > 0$)

(3) $\lim_{x \rightarrow -\infty} (1 - e^{-\lambda x}) = 0$; $\lim_{x \rightarrow \infty} (1 - e^{-\lambda x}) = 1$.

Beweis. (1) Sei $x_2 > x_1$

$$1 - e^{-\lambda x_2} > 1 - e^{-\lambda x_1} \iff e^{-\lambda x_2} < e^{-\lambda x_1} \iff \frac{1}{e^{\lambda x_2}} < \frac{1}{e^{\lambda x_1}} \iff x_2 > x_1 \text{ (oder mit Differenzieren:)}$$

$\frac{d}{dx}(1 - e^{-\lambda x}) = \lambda e^{-\lambda x} > 0$ für $\lambda > 0$ auf $\mathbb{R}$).

(2) Wir bestimmen die Nullstelle:

$$1 - e^{-\lambda x} = 0 \iff -e^{-\lambda x} = -1 \iff e^{-\lambda x} = 1 \iff -\lambda x \ln e = \ln 1 = 0 \iff -\lambda x = 0 \iff x = 0$$

Somit verläuft die Kurve durch den Ursprung.

(3) Laut Definition ist $F_X(x) = 0$ für $x \in]-\infty; 0[$. Entsprechend gilt $\lim_{x \rightarrow -\infty} F_X(x) = 0$;

$$\lim_{x \rightarrow \infty} (1 - e^{-\lambda x}) = \lim_{x \rightarrow \infty} 1 - \lim_{x \rightarrow \infty} \frac{1}{e^{\lambda x}} = 1 + 0 = 1. \quad \square$$

Je grösser λ , desto langsamer steigt die Kurve für wachsende x gegen 1, denn in $1 - e^{-\lambda x}$ ist $e^{-\lambda x} = \frac{1}{e^{\lambda x}}$ kleiner, wenn λ grösser ist. Entsprechend wird von 1 weniger abgezählt. Es folgen dazu ein paar graphische Beispiele (s. Abbildungen 12.2.4 - 12.2.7):

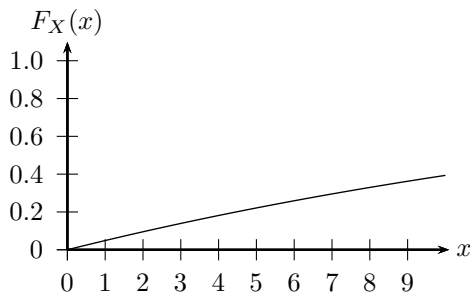


Abbildung 12.2.4: Graphische Darstellung der Verteilungsfunktion F der Exponentialverteilung mit dem Parameter $\lambda = 0.05$

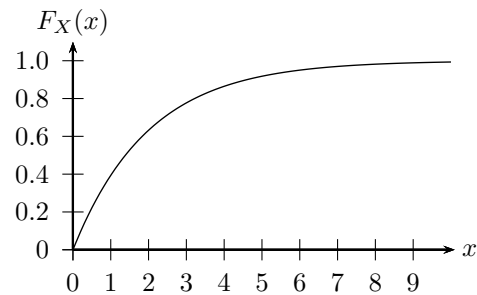


Abbildung 12.2.5: Graphische Darstellung der Verteilungsfunktion F der Exponentialverteilung mit dem Parameter $\lambda = 0.5$

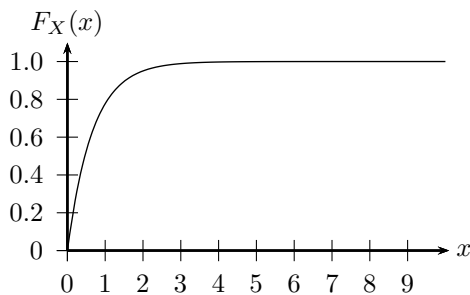


Abbildung 12.2.6: Graphische Darstellung der Verteilungsfunktion F der Exponentialverteilung mit dem Parameter $\lambda = 1.5$

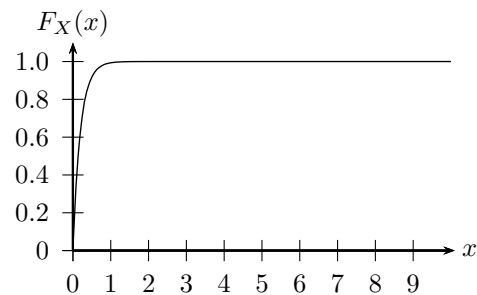


Abbildung 12.2.7: Graphische Darstellung der Verteilungsfunktion F der Exponentialverteilung mit dem Parameter $\lambda = 5$

Bei Exponentialverteilungen ist die Wahrscheinlichkeit für das Eintreffen von kleinen x_i bedeutend grösser als für das Eintreffen von grossen x_i , wobei die Bedeutung von „klein“ vom Parameter abhängt.

Die Verteilungsfunktion der Exponentialverteilung ist stetig. Wir können entsprechend die Dichtefunktion einer Exponentialverteilung berechnen:

$$\frac{d}{dx} F_X(x) = \frac{d}{dx} (1 - e^{-\lambda x}) = \lambda e^{-\lambda x}$$

(für $x \geq 0$; 0 für $x < 0$, da $\frac{d}{dx} 0 = 0$)

Für $\lambda = 0.5$ erhalten wir den folgende Graphen der Dichtefunktion (s. Abbildung 12.2.8):

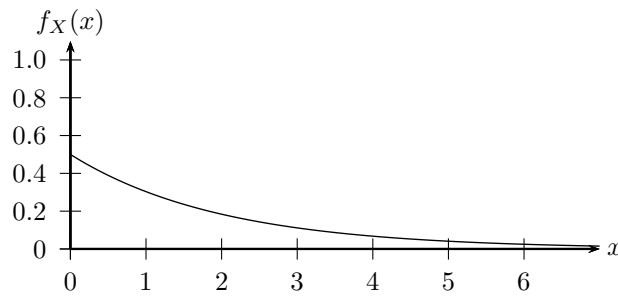


Abbildung 12.2.8: Graphische Darstellung der Dichtefunktion f der Exponentialverteilung mit dem Parameter $\lambda = 0.5$

Der Parameter λ gibt an, wie schnell sich die Dichtefunktion der x -Achse nähert: Je größer der Parameter, desto mehr Fläche liegt nahe der f -Achse (s. Abbildungen 12.2.9 - 12.2.12, der Ordinatenabschnitt ist jeweils λ , da $f_X(0) = \lambda e^{-\lambda \cdot 0} = \lambda$).

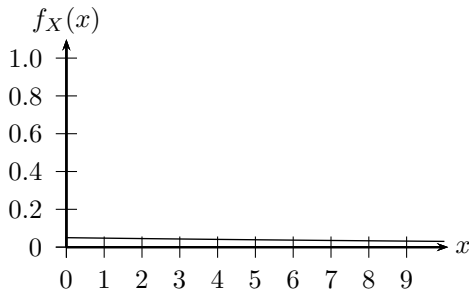


Abbildung 12.2.9: Graphische Darstellung der Dichtefunktion f der Exponentialverteilung mit dem Parameter $\lambda = 0.05$

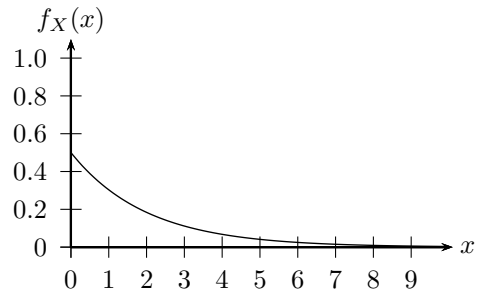


Abbildung 12.2.10: Graphische Darstellung der Dichtefunktion f der Exponentialverteilung mit dem Parameter $\lambda = 0.5$

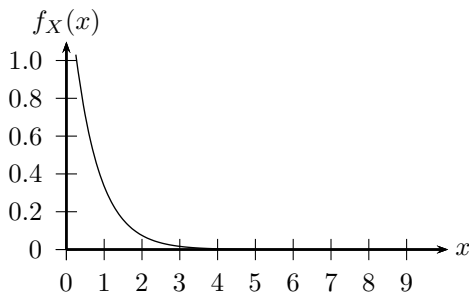


Abbildung 12.2.11: Graphische Darstellung der Dichtefunktion f der Exponentialverteilung mit dem Parameter $\lambda = 1.5$

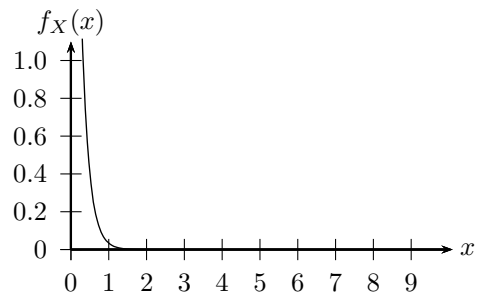


Abbildung 12.2.12: Graphische Darstellung der Dichtefunktion f der Exponentialverteilung mit dem Parameter $\lambda = 5$

Auch hier gibt die Fläche unter der Dichtefunktion die Wahrscheinlichkeit des Eintreffens der Werte von Zufallsvariablen an. Die Fläche unter der Dichtefunktion im Intervall $]0, \infty[$ ist gleich 1. Mit Hilfe der Dichtefunktion können wir im Prinzip die Wahrscheinlichkeiten für beliebige Teilintervalle von $]0, \infty[$ bestimmen. Dazu bräuchte man allerdings die Integralrechnung, wobei das Integral der Dichtefunktion die Verteilungsfunktion ist (Stammfunktion). Damit wird die Flächenberechnung auf die selbe Berechnungsart zurückgeführt, die wir oben in den Übungen angewendet haben.

Der Median $median_X$ der Exponentialverteilung ist:

Satz 12.2.4.

$$median_X = -\frac{\ln 0.5}{\lambda}$$

Beweis. Es gilt $\text{median}_X = x$ so dass $1 - e^{-\lambda x} = 0.5$. Wir isolieren x :

$$\begin{aligned} 1 - e^{-\lambda x} = 0.5 &\iff -e^{-\lambda x} = -0.5 \\ &\iff e^{-\lambda x} = 0.5 \\ &\iff -\lambda x \ln e = \ln 0.5 \\ &\iff x = -\frac{\ln 0.5}{\lambda} \approx \frac{0.69315}{\lambda} \end{aligned}$$

□

Die Dichtefunktion der Exponentialverteilung ist nicht symmetrisch. Entsprechend fallen der Erwartungswert und der Median nicht zusammen. Für Exponentialverteilungen gilt:

$$E(X) = \int_0^{\infty} (x (\lambda e^{-\lambda x})) dx = \frac{1}{\lambda}.$$

Der Median ist immer kleiner als der Erwartungswert, da für $\lambda > 0$ gilt:

$$\frac{0.69315}{\lambda} < \frac{1}{\lambda} \iff 0.69315 < 1.$$

Beispiel 12.2.5. $\lambda = 0.5 \implies E(X) = \frac{1}{0.5} = 2$; Median: $\frac{0.69315}{0.5} = 1.3863$ (s. Abbildung 12.2.13).

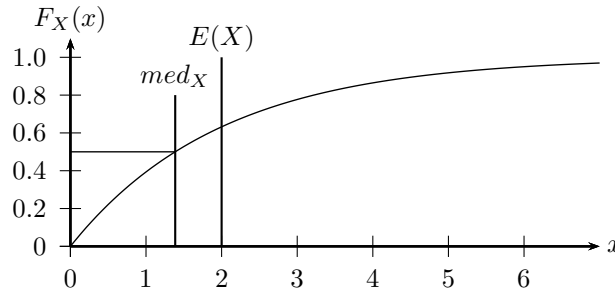


Abbildung 12.2.13: Lage von $E(X)$ und med_X bei der Exponentialverteilung mit dem Parameter $\lambda = 0.5$

◇

Die Varianz einer Exponentialverteilung ist:

$$V(X) = \int_0^{\infty} \left((\lambda e^{-\lambda x}) \left(x - \frac{1}{\lambda} \right)^2 \right) dx = \frac{1}{\lambda^2}$$

Somit gilt:

$$\sqrt{V(X)} = E(X) = \frac{1}{\lambda}.$$

Die Standardabweichung schneidet damit links des Erwartungswertes das Intervall $[0, E(X)]$, ab. Rechts des Erwartungswertes $2E(X)$. Das gesamte Intervall umfasst somit: $F_X(2E(X)) = 1 - e^{-\lambda 2E(X)} = 1 - e^{-\frac{\lambda^2}{\lambda}} = 1 - e^{-2} = 0.86466$. Die Wahrscheinlichkeit dafür, dass die Werte der Zufallsvariable im Intervall

$$[E(X) - \sqrt{V(X)}, E(X) + \sqrt{V(X)}] = [0, 2E(X)]$$

um den Erwartungswert herum liegen, beträgt somit rund 85%.

12.2.1 Übung

1. Gegeben sei die Exponentialverteilung $F_X(x) = 1 - e^{-\lambda x}$ mit $\lambda = 0.8$. Es gehe um die Zeit bis zum ersten Ausfall einer neuen Maschine (in Monaten).
 - (a) Berechnen Sie $P(X \leq 5)$ (= Wahrscheinlichkeit, dass eine neue Maschine in den ersten 5 Monaten zum ersten Mal ausfällt)
 - (b) Berechnen Sie $P(3 < X \leq 6)$ (=Wahrscheinlichkeit, dass eine neue Maschine zwischen dem Ende des 2. Monats und dem Ende des 6. Monat ausfällt)
 - (c) Berechnen Sie $P(100 < X \leq 100.1)$ (=Wahrscheinlichkeit, dass eine neue Maschine in den ersten 3 Tagen des 100. Monats zum ersten Mal ausfällt, Monat à 30 Tage).
 - (d) Berechnen Sie $P(3 < X)$ (Wahrscheinlichkeit, dass eine neue Maschine nach den ersten 3 Monaten zum ersten Mal ausfällt).
 - (e) Berechnen Sie $P(0.01 < X \leq 1)$
2. Berechnen Sie für die Exponentialverteilung (Modell für die Lebensdauer von Glühbirnen in Monaten) mit $\lambda = 0.9$:
 - (a) den Median
 - (b) den Erwartungswert
 - (c) die Varianz
 - (d) die Standardabweichung
 - (e) $P(X \leq 0.8)$
 - (f) $P(0.4 < X \leq 0.8)$
 - (g) $P(X \leq -0.3)$
 - (h) $P(2 < X)$
 - (i) Geben Sie die Dichtefunktion an.
 - (j) Erstellen Sie eine Zeichnung

12.2.2 Lösung

1.
 - (a) $P(X \leq 5) = F_X(5) = 1 - e^{-0.8 \cdot 5} = 0.98168$
 - (b) $P(3 < X \leq 6) = F_X(6) - F_X(3) = 1 - e^{-0.8 \cdot 6} - (1 - e^{-0.8 \cdot 3}) = 0.082488$.
 - (c) $P(100 < X \leq 100.1) = F_X(100.1) - F_X(100) = 1 - e^{-0.8 \cdot 100.1} - (1 - e^{-0.8 \cdot 100}) = 1.3876 \times 10^{-36}$
 - (d) $P(3 < X) = P(X > 3) = 1 - F_X(3) = 1 - (1 - e^{-0.8 \cdot 3}) = 0.090718$.
 - (e) $P(0.01 < X \leq 1) = F_X(1) - F_X(0.01) = 1 - e^{-0.8 \cdot 1} - (1 - e^{-0.8 \cdot 0.01}) = 0.5427$
2. Man erhält:
 - (a) der Median: $\frac{0.69315}{\lambda} = \frac{0.69315}{0.9} = 0.77017$
 - (b) den Erwartungswert: $\frac{1}{\lambda} = \frac{1}{0.9} = 1.1111$
 - (c) die Varianz: $\frac{1}{\lambda^2} = \frac{1}{0.9^2} = 1.2346$
 - (d) die Standardabweichung = $\sqrt{1.2346} = 1.1111$
 - (e) $P(X \leq 0.8) = F_X(0.8) = 1 - e^{-0.9 \cdot 0.8} = 0.51325$
 - (f) $P(0.4 < X \leq 0.8) = F_X(0.8) - F_X(0.4) = 1 - e^{-0.9 \cdot 0.8} - (1 - e^{-0.9 \cdot 0.4}) = 0.21092$
 - (g) $P(X \leq -0.3) = 0$ (da $P(X \geq 0) = 1$)
 - (h) $P(2 < X) = 1 - (1 - e^{-0.9 \cdot 2}) = 0.1653$
 - (i) Dichtefunktion: $f_X(x) = \lambda e^{-\lambda x} = 0.9e^{-0.9x}$

12.3 Lernziele

- Die Familien der stetigen, uniformen Verteilungen und der Exponentialverteilungen definieren können (Formeln müssen nicht auswendig gewusst werden). Abkürzungen verstehen: $U(a, b)$, $Exp(\lambda)$.
- Mit Hilfe von Formeln für diese Verteilungsfamilien die Kennzahlen Erwartungswert, Varianz und Standardabweichung berechnen können.
- Wahrscheinlichkeiten für diese Verteilungsfamilien berechnen können.
- Verteilungs- und Dichtefunktionen von stetigen uniformen Verteilungen zeichnen können.
- Verteilungs- und Dichtefunktionen von Exponentialverteilungen zeichnen können.
- Wissen, für welche Situationen Exponentialverteilungen angemessen sein können (s. Beispiele).
- Übungen der obigen Art lösen können.

Kapitel 13

Schätzen von stetigen Wahrscheinlichkeitsverteilungen durch Daten

Um das Verständnis für stetige Verteilungen zu vertiefen, untersuchen wir, bevor wir uns weiteren stetigen Verteilungsfamilien zuwenden, den Zusammenhang von empirischen und theoretischen Verteilungsfunktionen. In der Tat kann man theoretische Verteilungsfunktionen durch empirische Verteilungsfunktionen schätzen, wie der Beweis des folgenden Satzes nachweist. Da durch eine theoretische Verteilungsfunktion eine Wahrscheinlichkeitsverteilung bestimmt ist, kann man mittels empirischer Verteilungsfunktion auch Wahrscheinlichkeitsverteilungen schätzen, laut der *i.i.d.* Zufallsvariablen verteilt sind, als deren Werte wir Daten interpretieren.

Als erstes führen wir ein paar Redewendungen ein. Betrachten wir ein Datum x_i als Wert einer P -verteilten Zufallsvariable X , so sagen wir künftig auch kürzer, das Datum sei P -verteilt. Betrachten wir n Daten x_i als Werte von n (unabhängigen) P -verteilten Zufallsvariablen X_i , so sprechen wir von (unabhängigen) P -verteilten Daten (Beachten Sie, dass in diesem Falle alle X_i dieselbe Verteilung P aufweisen). Statt von der Verteilungsfunktion einer P -verteilten Zufallsvariable sprechen wir oft kürzer von der Verteilungsfunktion von P .

Der folgende Satz ist für das folgende von fundamentaler Bedeutung. Es handelt sich um eine schwache Version eines Satzes, der wegen seiner zentralen Bedeutung oft „Hauptsatz der Statistik“ genannt wird.

Satz 13.0.1. *Sei*

- $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum ($\mathcal{A}$ ist das Mengensystem, auf das W -Mass P definiert ist) und $X_i : \Omega \rightarrow \mathbb{R}$ abzählbar unendlich viele, identisch nach P verteilte unabhängige Zufallsvariablen.
- $\mathbf{x} := (x_1, x_2, \dots, x_n, \dots)$ ein unendlicher Vektor von Werten der abzählbar unendliche vielen Zufallsvariablen $(X_1, X_2, \dots, X_n, \dots)$, d.h. $\mathbf{x} \in \mathbb{R}^\infty$ ($\mathbb{R}^\infty$ ist das unendlichfache kartesische Produkt von $\mathbb{R}$).
-

$$F_n : \mathbb{R} \times \mathbb{R}^\infty \rightarrow [0, 1]$$

$$F_n(t, \mathbf{x}) := \frac{1}{n} |\{x_j : x_j \leq t \text{ mit } j \leq n\}|$$

die empirische Verteilungsfunktion der n ersten Daten x_i von $\mathbf{x}$ (Treppenform; t wird von F_n die relative Häufigkeit der ersten n Daten von $\mathbf{x}$ zugeordnet, die kleiner-gleich t sind), und

- F die theoretische Verteilungsfunktion von X_i .

Dann gilt für beliebig kleine $k > 0$:

$$P(\{\mathbf{x} : |F_n(t, \mathbf{x}) - F(t)| > k\}) = 0$$

für n gegen Unendlich. (formal: $\lim_{n \rightarrow \infty} P(\{\mathbf{x} : |F_n(t, \mathbf{x}) - F(t)| > k\}) = 0$)

Beweis. (i) Laut Schwachem Gesetz der grossen Zahlen gilt $P(\{\mathbf{x} : |\bar{X}_n(\mathbf{x}) - E(X_i)| > k\}) = 0$ für n gegen ∞ für $\bar{X}_n : \mathbb{R}^\infty \rightarrow \mathbb{R}$; $\bar{X}_n(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^n x_i$ (X_i unabhängig und identisch verteilt; $\mathbf{x}$ definiert wie oben, x_i sind die ersten n Daten von $\mathbf{x}$).

(ii) Der Wert einer Zufallsvariable X_i fällt entweder ins Intervall $]-\infty, t]$ oder nicht und zwar fällt der Wert der Zufallsvariable X_i mit einer Wahrscheinlichkeit von $F(t)$ in dieses Intervall, da F die Verteilungsfunktion von X_i ist. Man kann also die bernoulli-verteilten Zufallsvariablen Y_i betrachten, mit $p = F(t)$ und $Y_i(x_i) = 1$ für $x_i \in]-\infty, t]$ und $Y_i(x_i) = 0$ sonst.

Für jedes n gilt: $\bar{Y}_n(\mathbf{x}) = F_n(t, \mathbf{x})$, denn $\bar{Y}_n(\mathbf{x})$ ist im Falle bernoulli-verteilter Zufallsvariablen die relative Häufigkeit der ersten n Daten von $\mathbf{x}$, die kleiner-gleich t sind.

Zudem gilt für Bernoulli-Verteilungen, wie gezeigt, $E(Y_i) = p$ und damit $E(Y_i) = F(t)$.

(iii) Durch Einsetzen ins schwache Gesetz der grossen Zahlen erhält man: $P(\{\mathbf{x} : |\bar{Y}_n(\mathbf{x}) - E(Y_i)| > k\}) = 0$ für n gegen ∞ .

Durch Ersetzen auf Grund der Ergebnisse von (ii) in diese Formulierung des schwachen Gesetzes der grossen Zahlen erhält man: $P(\{\mathbf{x} : |F_n(t, \mathbf{x}) - F(t)| > k\}) = 0$ für n gegen ∞ $\square$

Bemerkung 13.0.2. • Der Satz besagt etwas locker ausgedrückt: je mehr Daten wir haben, die als Werte von unabhängigen, P -verteilten Zufallsvariablen zu betrachten sind, desto näher liegen mit steigender Wahrscheinlichkeit die Punkte der empirischen Verteilungsfunktion bei der theoretischen Verteilungsfunktion von P .

- Die n -Tupel der Daten nennen wir im folgenden auch „Stichprobe“. Eine Stichprobe ist damit dasselbe wie ein Datensatz, wobei wir hier den Zufallsaspekt betonen. In der beschreibenden Statistik haben wir Daten als Skalenwerte aufgefasst, die wir durch Messung von Objekten einer Grundmenge erhalten. Hier fassen wir Daten als Werte von Zufallsvariablen auf. Diese zwei Sichtweisen sind verträglich. In der Tat erhalten wir konkrete Daten auch hier durch Messung von Objekten. Diese betrachten wir hier einfach als Werte von Zufallsvariablen. $\diamond$

Wir können den obigen Satz an einem Beispiel klar machen:

Beispiel 13.0.3. Wir betrachten Stichproben ungleicher Grösse ($n = 15; 30; 200$), wobei die Daten Werte exponentialverteilter Zufallsvariablen $X_i \sim \text{Exp}(2)$ sind (s. Abbildungen 13.0.1 - 13.0.3).

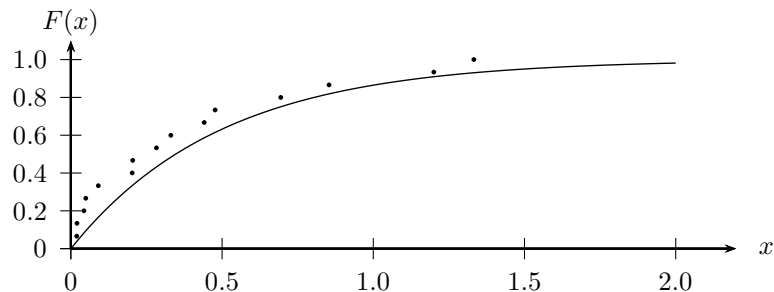


Abbildung 13.0.1: Empirische Verteilungsfunktion von mit Zufallsgenerator erzeugten exponentialverteilten Daten und dazugehörige theoretische Verteilungsfunktion; $n = 15$, $\lambda = 2$

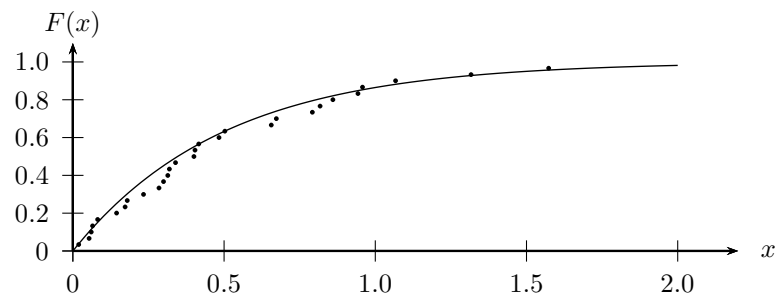


Abbildung 13.0.2: Empirische Verteilungsfunktion von mit Zufallsgenerator erzeugten exponentiellverteilten Daten und dazugehörige theoretische Verteilungsfunktion; $n = 30$, $\lambda = 2$

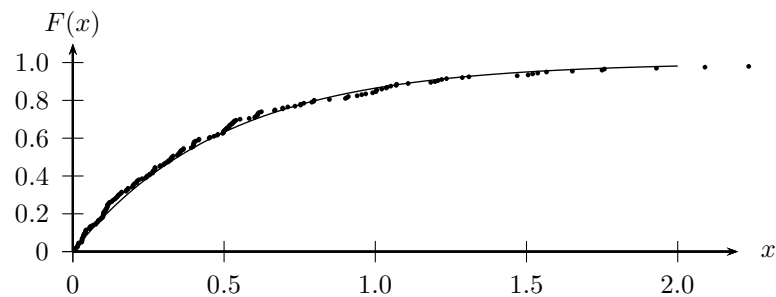


Abbildung 13.0.3: Empirische Verteilungsfunktion von mit Zufallsgenerator erzeugten exponentiellverteilten Daten und dazugehörige theoretische Verteilungsfunktion; $n = 200$, $\lambda = 2$

◇

Im Beispiel sehen wir, dass bei grösserer Stichprobe die Abweichung der empirischen Verteilungsfunktion von der theoretischen Verteilungsfunktion immer kleiner wird. Zu beachten ist allerdings, dass bei der Reproduktion sich die Tendenz nur mit hoher Wahrscheinlichkeit ergibt, nicht jedoch mit Sicherheit (je mehr Daten, desto höher die Wahrscheinlichkeit, dass die beiden Funktionen nahe beieinander liegen)!

Können wir nun davon ausgehen, dass Daten als Werte von identisch und unabhängig verteilten Zufallsvariablen gelten können, können wir auf Grund von Satz 13.0.1 die empirische Verteilungsfunktion verwenden, um die Verteilung zu schätzen, laut der die Daten verteilt sind. Je mehr Daten wir verwenden, desto besser ist die Schätzung.

Beispiel 13.0.4. Die folgenden metrisch skalierten Daten liegen vor. Wir gehen davon aus, dass sie als Werte von identisch und unabhängig verteilten Zufallsvariablen gelten können (s. Tabelle 13.0.1):

Daten
0.3
0.5
0.7
0.8
1.1
1.2
1.5
1.7
2
2.2

Tabelle 13.0.1: Beispieldaten für die Schätzung einer Verteilung

Die Anzahl der Daten ist für eine Schätzung zu knapp. Um die Berechnung und Zeichnung von Hand zu erlauben, arbeiten wir aber zuerst mit wenigen Daten. Die empirische Verteilungsfunktion F_n ist dann (x_i ist das i -te Datum des geordneten Datensatzes): $F_n(x_i) = \frac{i}{n}$ (Punkteform; s. Abbildung 13.0.4):

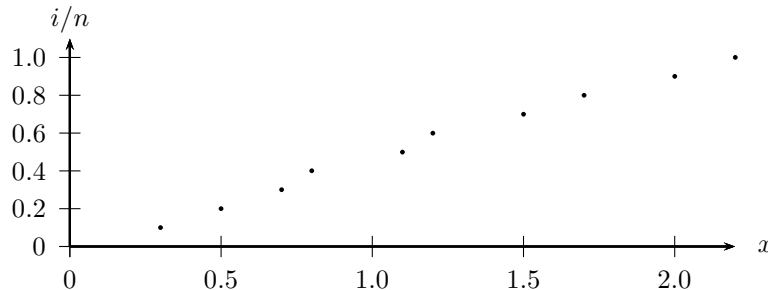


Abbildung 13.0.4: Empirische Verteilungsfunktion der Daten von Tabelle 13.0.1

Dies ist die bildliche Darstellung des folgenden Graphen: $\{(0.3, \frac{1}{10}), (0.5, \frac{2}{10}), (0.7, \frac{3}{10}), (0.8, \frac{4}{10}), (1.1, \frac{5}{10}), (1.2, \frac{6}{10}), (1.5, \frac{7}{10}), (1.7, \frac{8}{10}), (2, \frac{9}{10}), (2.2, \frac{10}{10})\}$. Die Punkte scheinen mehr oder weniger auf einer Geraden zu liegen. Entsprechend wäre hier wohl als Modell eine uniforme Verteilung angesagt. Aus der Familie der uniformen Verteilungen gilt es nun, die optimale Verteilung zu schätzen. Dabei sollte kein Datum ausserhalb des geschätzten Intervalls liegen, da die Wahrscheinlichkeit dafür 0 ist. Es gibt verschiedene Schätz-Möglichkeiten:

1. Man kann zeigen, dass bei der uniformen Verteilung das erste und der letzte Datum die plausibelsten Schätzer für a und b des zu schätzenden Intervalls $[a, b]$ ergeben (sogenannte Maximum-Likelihood-Schätzer). Bei dieser Methode berechnen wir für die Verteilungsfunktion auf dem Intervall $[0.3, 2.2]$ eine Gerade durch $(0.3, 0)$ und $(2.2, 1)$ - also durch die Punkte $(x_1, 0)$ und $(x_n, 1)$ des geordneten Datenvektors mit n Daten.
2. Bei wenig Daten ist es allerdings besser, die Verteilungsfunktions-Gerade G durch $(x_1, \frac{1}{n})$ und $(x_n, 1)$ des geordneten Datenvektors mit n Daten zu berechnen - im Beispiel durch $(0.3, \frac{1}{10})$ und $(2.2, 1)$. $\hat{a}$ wird dann mit der Nullstelle der Geraden G identifiziert und $\hat{b}$ mit x_n , im Beispiel mit 2.2, d.h. mit dem grössten Datum. Asymptotisch sind die Verfahren 1 und 2 identisch.
3. Bei den beiden erwähnten Verfahren haben das erste und das letzte Datum einen recht grossen Einfluss auf den Verlauf der Geraden. Man könnte also versucht sein, eine lineare Regression durch die Punkte der empirischen Verteilungsfunktion zu rechnen. Dadurch können allerdings Daten ausserhalb des geschätzten Intervalls zu liegen kommen. Das Verfahren ist also nur sinnvoll, wenn keine Daten ausserhalb des geschätzten Intervalls liegen. Man könnte zwar $\hat{a} = \min\{x_1, x'_1\}$ setzen, wobei x'_1 die Nullstelle der Regressionsgerade ist, und $\hat{b} = \max\{x_n, x'_n\}$ mit x'_n der x -Koordinate des Schnittpunktes der Regressions-Geraden mit der Konstanten 1. Dann müsste man die Gerade neu berechnen, womit sie nicht mehr die optimale Gerade durch die Punkte der empirischen Verteilungsfunktion ist (für die Bedeutung der Hütchen s. Seite 245). Müssten beide Punkte angepasst werden, erhalten wir die Lösung von 1.
4. Man könnte versucht sein, a und b aus den Formeln $E(X) = \frac{a+b}{2}$ und $V(X) = \frac{(b-a)^2}{12}$ zu schätzen, indem man $E(X)$ mit dem Mittelwert der Daten und $V(X)$ mit der Varianz der Daten schätzt. Man erhält zwei Gleichungen mit zwei Unbekannten, was einfach zu lösen ist. Das Verfahren unterschätzt allerdings die Breite des Intervalls $[a, b]$ und führt dazu, dass es Daten ausserhalb dieses Intervalls hat.

Um die ersten zwei Verfahren zu vergleichen, rechnen wir sie durch. Zu beachten ist, dass der Datensatz klein ist, wodurch sich recht grosse Unterschiede ergeben:

1. Die Steigung ist $\frac{1}{2.2-0.3} = 0.52632$. Mit dem Punkt $(0.3, 0)$ gilt $0 = 0.52632 \cdot 0.3 + b$ und damit $b = -0.15790$. Damit gilt

$$\hat{F}(x) = \begin{cases} 0 & x \leq 0.3 \\ 0.52632x - 0.15790 & x \in]0.3, 2.2] \\ 1 & x > 2.2 \end{cases}$$

2. Die geschätzte Steigung der Geraden ist $\hat{a} = \frac{\frac{9}{10}}{2.2-0.3} = 0.4736$. Wir haben damit $\frac{1}{10} = 0.47368 \cdot 0.3 + \hat{b} \Rightarrow \hat{b} = -0.042104$. Damit erhalten wir die folgende geschätzte theoretische Verteilungsfunktion:

$$\hat{F}(x) = \begin{cases} 0 & x \leq 0.08887 \\ 0.47368x - 0.042104 & x \in]0.08887, 2.2] \\ 1 & x > 2.2 \end{cases}$$

da für $0.47368x - 0.042104 = 0$ gilt: $x = 0.08887$.

Um zu zeigen, dass das 4. Verfahren nicht geeignet ist, rechnen wir es durch: Im Beispiel erhält man mit $\bar{x} = 1.2$ und $s^2 = 0.41111111$

$$\begin{aligned} 1.2 &= \frac{a+b}{2} \\ 0.41111111 &= \frac{(b-a)^2}{2} \end{aligned}$$

mit den Lösungen $a = 0.74662$ und $b = 1.6534$.

Für das zweite Verfahren erhalten wir die Abbildung 13.0.5

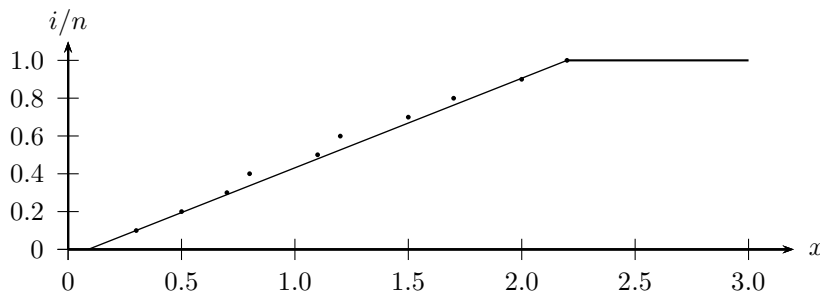


Abbildung 13.0.5: Empirische Verteilungsfunktion der Daten von Tabelle 13.0.1 mit aus Daten geschätzter Verteilungsfunktion

Wir haben somit ein Modell für unsere konkreten Daten, das relativ gut zu passen scheint. Wir können annehmen, dass unsere Daten Werte uniform verteilter stetiger Zufallsvariablen X_i sind, wobei $X_i \sim U(0.0888, 2.2)$. Mit anderen Worten: wir haben durch unsere Daten ein Verteilungsmodell geschätzt, so dass die Daten vermutlich gemäss dieser Verteilung verteilt sind. Auf Grund dieser Hypothese können wir nun beliebige Wahrscheinlichkeiten berechnen: z.B. $\hat{P}(X < 0.7) = \hat{F}(0.7) = 0.47368 \cdot 0.7 - 0.042104 = 0.28947$
 $\hat{P}(0.3 < X \leq 0.8) = \hat{F}(0.8) - \hat{F}(0.3) = 0.47368 \cdot 0.8 - 0.042104 - (0.47368 \cdot 0.3 - 0.042104) = 0.23684$ $\diamond$

Beispiel 13.0.5. Für 11 Maschinen aus einer spezifischen Produktion hat man die folgenden Lebensdauern gemessen (in Jahren; s. Tabelle 13.0.2):

Daten
0.3
0.5
0.8
1.5
2
2.5
3.5
4
6
8
11

Tabelle 13.0.2: Beispieldaten für die Schätzung einer Verteilung

Wir erstellen einen Punkteplot der entsprechenden empirischen Verteilungsfunktion: $\{(0.3, \frac{1}{11}), (0.5, \frac{2}{11}), (0.8, \frac{3}{11}), (1.5, \frac{4}{11}), (2, \frac{5}{11}), (2.5, \frac{6}{11}), (3.5, \frac{7}{11}), (4, \frac{8}{11}), (6, \frac{9}{11}), (8, \frac{10}{11}), (11, \frac{11}{11})\}$.

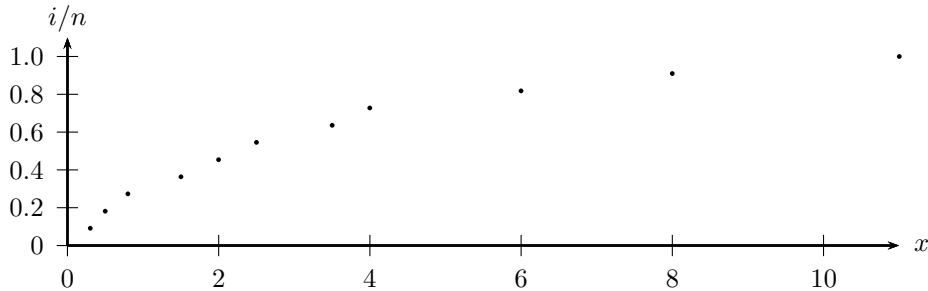


Abbildung 13.0.6: Empirische Verteilungsfunktion der Daten der Tabelle 13.0.2

Die Daten scheinen zu einer Exponentialverteilung zu passen. Wir müssen nun aus der Familie der Exponentialverteilungen die Verteilungsfunktion picken, die am besten passt. Es geht darum einen geeigneten Parameter λ zu schätzen. Der Mittelwert eines Datensatzes ist ein guter Schätzer für den Erwartungswert $E(X)$ einer Verteilung (Erwartungstreue). Damit setzen wir schätzungsweise für $E(X)$: $\widehat{E(X)} = \frac{0.3+0.5+0.8+1.5+2+2.5+3.5+4+6+8+11}{11} = 3.6455$. Dies erlaubt uns eine Schätzung des Parameters λ der Verteilung, da $\frac{1}{\lambda} = E(X)$. Wir erhalten für λ den geschätzten Wert $\hat{\lambda} : \frac{1}{\hat{\lambda}} = \widehat{E(X)} = 3.6455 \implies \hat{\lambda} = \frac{1}{3.6455} = 0.27431$. Die geschätzte theoretische Verteilungsfunktion lautet deshalb:

$$\hat{F}_X(x) = 1 - e^{-0.27431x}$$

Zusammen mit der theoretischen Exponentialverteilung erhalten wir ein Bild, das von Auge die Abschätzung der Angemessenheit der Modellierung erlaubt (s. Abbildung 13.0.7):

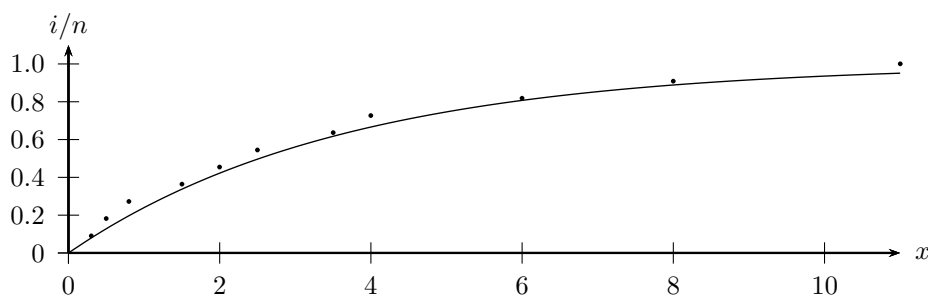


Abbildung 13.0.7: Empirische Verteilungsfunktion der Daten der Tabelle 13.0.2 mit aus Daten mittels Mittelwert geschätzter Verteilungsfunktion

Wir könnten ebenfalls den empirischen Median verwenden, um λ zu schätzen. Der empirische Median der obigen Daten ist mit 2.5 identisch. Wir nehmen eine Schätzung des theoretischen Medians durch den empirischen Median vor und erhalten: $\frac{0.69315}{\hat{\lambda}} = 2.5 \implies \hat{\lambda} = 0.27726$

Wir erhalten die folgende Anpassung (s. Abbildung 13.0.8):

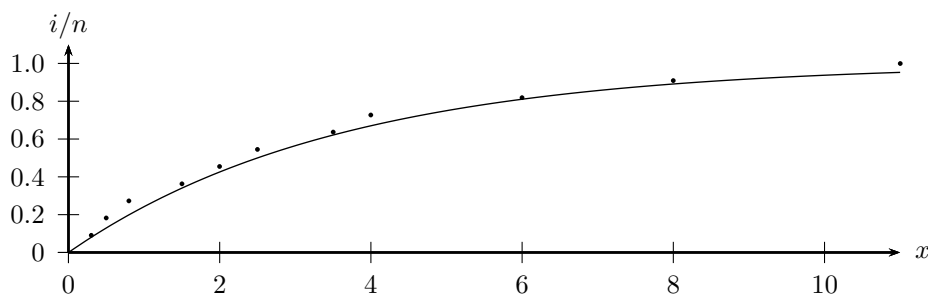


Abbildung 13.0.8: Empirische Verteilungsfunktion der Daten der Tabelle mit aus Daten mittels Median geschätzter Verteilungsfunktion

Die Schätzung des Modells mit Hilfe des Medians führt im konkreten Fall zu einer mit der vorangegangenen Schätzung vergleichbaren Anpassung an die Punkte. Wir werden in der Folge jedoch jeweils den Mittelwert verwenden, um λ zu schätzen.

Auch im Fall der Exponentialverteilungen können wir das Modell dazu verwenden, um Wahrscheinlichkeiten zu berechnen: gelingt eine gute Anpassung einer theoretischen Verteilung an die Daten (d.h. eine gute Schätzung des zugrundeliegenden Modells mit Hilfe von Daten), kann die theoretische Verteilung dazu verwendet werden, um beliebige Wahrscheinlichkeiten zu berechnen. So wäre die Wahrscheinlichkeit, dass eine Maschine eine Lebensdauer von mehr als 10 Jahren hat, im obigen Beispiel je nach gewähltem Parameter näherungsweise:

$$\hat{P}(X > 10) = 1 - (1 - e^{-0.27726 \cdot 10}) = 0.062499$$

$$\hat{P}(X > 10) = 1 - (1 - e^{-0.27431 \cdot 10}) = 0.06437$$

Will eine Firma darüber nachdenken, wie lange die Garantiedauer zu sein hätte, damit sich die Kosten in Grenzen halten, können die folgenden Wahrscheinlichkeiten berechnet werden: $\hat{P}(X \leq 1) = 1 - e^{-0.32 \cdot 1} = 0.27385$ (27% der Maschinen müssten durchschnittlich in Jahresfrist ersetzt werden; es wäre wohl nötig, die Qualität der Maschinen zu verbessern!). Man könnte die Meinung vertreten, dieses Resultat könnte man auch einfacher haben, denn der Anteil der Ausfälle im ersten Jahr an den gesamten Ausfällen beträgt: $\frac{3}{11} = 0.27273$. Dagegen ist einzuwenden, dass diese Übereinstimmung zufällig erfolgt, da die Daten zufällig verteilt sind. Betrachten wir z.B. $\hat{P}(4 < X \leq 5.9) = \hat{F}_X(5.9) - \hat{F}_X(4) = 1 - e^{-0.32 \cdot 5.9} - (1 - e^{-0.32 \cdot 4}) = 0.12666$. Berechnen wir nun den Anteil der Fälle, so erhalten wir $\frac{0}{11} = 0$. Obwohl die Wahrscheinlichkeit dafür, dass ein Ergebnis ins Intervall $[4, 5.9]$ fällt, 0.126 beträgt, fiel bei der konkreten Zufallsstichprobe kein Ergebnis in das besagte Intervall. Eine angemessene Berechnung der entsprechenden Wahrscheinlichkeiten durch die relativen Häufigkeiten pro Intervall würde sich erst bei einer sehr grossen Stichprobe ergeben.

Allerdings ist sie in diesem Fall mühsamer als mit dem entsprechenden Modell. Die Schätzung mittels Verteilungsfunktion ist zudem präziser als die Schätzung mittels relativer Häufigkeiten, da bei der ersten Methode alle Daten in die Schätzung einfließen.

$\hat{P}(X \leq 2) = 1 - e^{-0.32 \cdot 2} = 0.47271$ (47% der Maschinen müssten durchschnittlich in zwei Jahren ersetzt werden). $\diamond$

13.0.1 Übungen

1. Gegeben sind die folgenden, stetig und metrisch skalierten Daten: 3.483624465, 4.28712291, 3.555873009, 1.807750757, 2.654784157, 11.90916005, 2.665970716, 1.479815386, 0.08793726, 2.227560483, 5.569712412, 0.204311269.

Zeichnen Sie einen Punkteplot der empirischen Verteilungsfunktion und versuchen Sie eine geeignete theoretische Verteilung an die Daten anzupassen.

Berechnen Sie $\hat{P}(X \leq 2.5)$, $\hat{P}(2 < X \leq 5)$ und $\hat{P}(4 < X)$.

2. Zeichnen Sie für die folgenden Daten (mit metrischer, stetiger Skala gemessen) eine empirische Verteilungsfunktion (mit Hilfe von Excel):

2.15; 2.411; 2.58; 2.81; 3.05; 3.198; 3.388; 3.601

Schlagen Sie ein „vernünftiges“ Modell für diese Daten vor. Berechnen Sie dann mit Hilfe Ihres Modells die folgenden Wahrscheinlichkeiten:

$$\hat{P}(2.2 < X \leq 2.9)$$

$$\hat{P}(X \leq 0.4)$$

$$\hat{P}(X \leq 3)$$

$$\hat{P}(2.8 < X)$$

3. Zeichnen Sie den Punkteplot der empirischen Verteilungsfunktion. Passen Sie eine passende theoretische Verteilungsfunktion an die Daten an (= schätzen). Berechnen Sie die geschätzte Wahrscheinlichkeit, dass eine Maschine im ersten Jahr ausfällt. Berechnen Sie die geschätzte Wahrscheinlichkeit, dass eine Maschine im 3. Jahr ausfällt. Wie lange muss die Garantiedauer sein, damit in ihr nur 5% der verkauften Produkte zurückkommen?: 4.8 5.1 8.2
1.2 9.2 2.7 3.5 11.7 3.6 2.8 20.9 11.7 2.5 1 7.9 0.8
5 3.3 2.9
4. Siehe für die Daten und die Aufgabenstellung Excel-Tabelle
5. Siehe für die Daten und die Aufgabenstellung Excel-Tabelle

13.0.2 Lösungen

1. Wir zeichnen den Graphen der Funktion: $\{(3.483624465, \frac{8}{12}), (4.28712291, \frac{10}{12}), (3.555873009, \frac{9}{12}), (1.807750757, \frac{4}{12}), (2.654784157, \frac{6}{12}), (11.90916005, \frac{12}{12}), (2.665970716, \frac{7}{12}), (1.479815386, \frac{3}{12}), (0.08793726, \frac{1}{12}), (2.227560483, \frac{5}{12}), (5.569712412, \frac{11}{12}), (0.204311269, \frac{2}{12})\}$ (s. Abbildung 13.0.9)

Die Daten scheinen exponentialverteilt zu sein. Wir berechnen den Mittelwert:

$\bar{x} = 3.3278$ und erhalten die Schätzung von λ : $\hat{\lambda} = \frac{1}{3.3278} = 0.3005$ Wir erhalten die Anpassung (s. Abbildung 13.0.10):

Die Anpassung scheint nicht besonders gut zu sein. Die Daten wurden mit Hilfe des Computers erstellt (exponentialverteilte Zufallszahlen mit $\lambda = 0.5$). Sie gehören somit zu einer Kurve, von der sie noch mehr abweichen (s. Abbildung 13.0.11). Bei so wenigen Daten passen die beiden Funktionen nicht unbedingt gut zueinander!

$$\hat{P}(X \leq 2.5) = 1 - e^{-0.3005 \cdot 2.5} = 0.52822$$

$$\hat{P}(2 < X \leq 5) = 1 - e^{-0.3005 \cdot 5} - (1 - e^{-0.3005 \cdot 2}) = 0.32569$$

$$\hat{P}(4 < X) = 1 - (1 - e^{-0.3005 \cdot 4}) = 0.30059$$

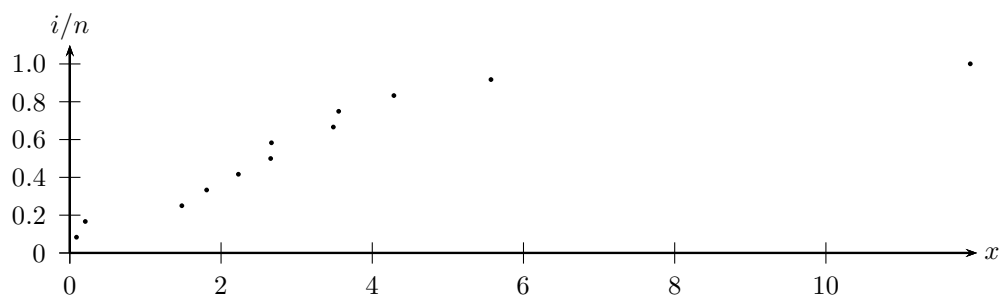


Abbildung 13.0.9: Empirische Verteilungsfunktion der Daten der Tabelle der Übung 1

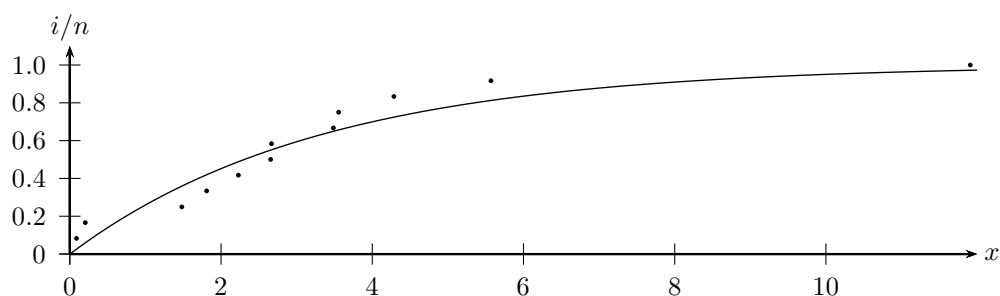


Abbildung 13.0.10: Empirische Verteilungsfunktion der Daten der Tabelle; $\lambda = 0.3005$ geschätzt mit Hilfe des Mittelwertes

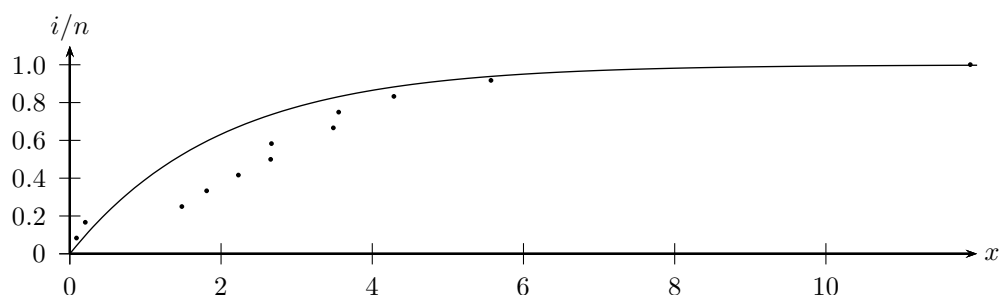


Abbildung 13.0.11: Empirische Verteilungsfunktion der Daten der Tabelle; $\lambda = 0.5$

2. Es gilt den Graphen der folgenden Funktion zu zeichnen:

$$\{(2.15, \frac{1}{8}), (2.411, \frac{2}{8}), (2.58, \frac{3}{8}), (2.81, \frac{4}{8}), (3.05, \frac{5}{8}), (3.198, \frac{6}{8}), (3.388, \frac{7}{8}), (3.601, \frac{8}{8})\}$$

Mit Excel müssen wir zwei Spalten füllen: eine Spalte mit den Daten, die der Grösse nach geordnet sind. Die zweite Spalte mit $\frac{i}{8}$. (i von 1 bis 8 in aufsteigender Reihenfolge). Dann lassen wir vom Programm das folgende Punktediagramm zeichnen (s. Abbildung 13.0.12):

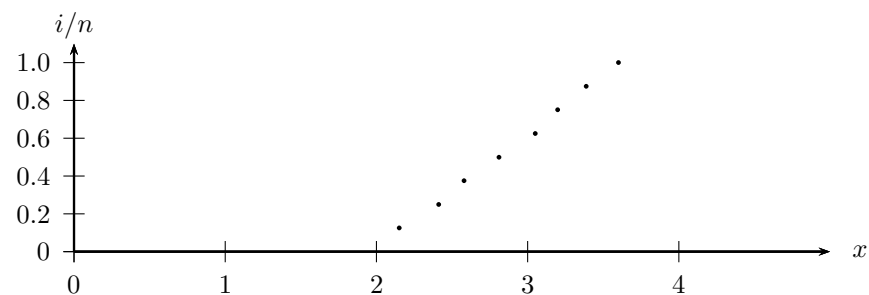


Abbildung 13.0.12: Empirische Verteilungsfunktion der Daten der Übung 2

Offenbar liegen die Punkte mehr oder weniger auf einer Geraden. Wir nehmen deshalb an, die Daten könnten durch eine uniform verteilte, stetige Zufallsvariable modelliert werden.

Wir wählen wiederum den obersten und den untersten Punkt, um eine Gerade zu berechnen: $(2.15, \frac{1}{8})$ und $(3.601, 1)$:

$$1 = 3.601a + b$$

$$\frac{1}{8} = 2.15a + b$$

Wir erhalten als Lösungen: $a = 0.60303$, $b = -1.1715$ und damit die geschätzte Verteilungsfunktion

$$F_X(x) = \begin{cases} 0 & \text{für } x < 1.9427 \\ 0.60303x - 1.1715 & \text{für } 2 \leq x \leq 3.601 \\ 1 & \text{für } x > 3.601 \end{cases}$$

da $0 = 0.60303x - 1.1715 \implies x = 1.9427$.

Anders ausgedrückt: Ein passendes Modell für die Daten ist $X \sim U(1.9427, 3.601)$ (s. Abbildung 13.0.13).

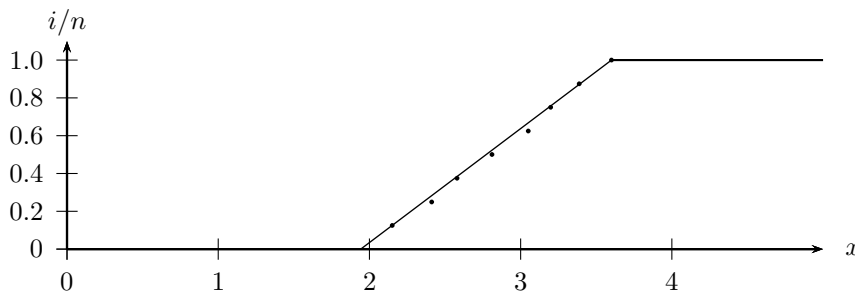


Abbildung 13.0.13: Empirische Verteilungsfunktion der Daten der Übung 2 mit aus Daten geschätzter Verteilungsfunktion

Wir berechnen die Wahrscheinlichkeiten:

$$\hat{P}(2.2 < X \leq 2.9) = \hat{F}_X(2.9) - \hat{F}_X(2.2) = 0.60303 \cdot 2.9 - 1.1715 - (0.60303 \cdot 2.2 - 1.1715) = 0.42212$$

$$\hat{P}(X \leq 0.4) = 0$$

$$\hat{P}(X \leq 3) = \hat{F}_X(3) = 0.60303 \cdot 3 - 1.1715 = 0.63759$$

$$\hat{P}(2.8 < X) = 1 - \hat{F}_X(2.8) = 1 - (0.60303 \cdot 2.8 - 1.1715) = 0.48302$$

3. Lösung siehe Excel-Tabelle

4. Lösung siehe Excel-Tabelle

5. Lösung siehe Excel-Tabelle

13.1 Lernziele

- Modelle aus bisher behandelten stetigen Verteilungsfamilien an geeignete Daten anpassen (schätzen) können und mit Hilfe der geschätzten Verteilungsfunktion geschätzte Wahrscheinlichkeiten berechnen können.
- Die Begründung kennen, wieso man mit Hilfe einer empirischen Verteilungsfunktion eine Verteilung schätzen kann, laut der die einzelnen Daten verteilt sind und wissen wie der Zusammenhang in Abhängigkeit von der Stichprobengröße aussieht (Inhalt des Hauptsatzes der Statistik wiedergeben können).

Kapitel 14

Die Familie der Normalverteilungen

14.1 Einleitung

Eine weitere stetige Verteilungsfamilie stellt die der Normalverteilungen dar. Sie spielt in der Statistik eine zentrale Rolle. Die Werte der Verteilungsfunktion von Normalverteilungen sind nicht durch eine Formel berechenbar. Nur die Dichtefunktion kann mit Hilfe einer Formel berechnet werden.

Definition 14.1.1. $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$ ist die Dichtefunktion einer normalverteilten Zufallsvariable genau X dann, wenn

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

mit $\sigma > 0; \mu \in \mathbb{R}$.

σ und μ sind die Parameter der Verteilung.

Ein Nachweis, dass f_X wirklich Dichtefunktion einer Zufallsvariable ist, besteht im Beweis von

$$\int_{-\infty}^{+\infty} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right) dx = 1.$$

Um auszudrücken, dass eine stetige Zufallsvariable normal verteilt ist, schreiben wir:

$$X \sim N(\mu, \sigma^2).$$

gebräuchlich ist auch die Darstellung $X \sim N(\mu, \sigma)$. Ist bekannt, dass eine Verteilung zur Familie der Normalverteilungen gehört, ist durch die Parameter die Verteilung genau bestimmt. Die Normalverteilung wird auch „Gaussche Verteilung“ genannt, die Dichtefunktion der Normalverteilung „Gaussche Glockenkurve“ (Johann Carl Friedrich Gauß, 1777 - 1855, deutscher Mathematiker und Astronom). Ohne Beweis halten wir fest:

Satz 14.1.2. Für $X \sim N(\mu, \sigma^2)$ gilt $E(X) = \mu$ und $V(X) = \sigma^2$.

Beispiel 14.1.3. a) Wir berechnen die Funktionswerte $f_X(x_i)$ der Werte $x_i = 2, 4$ der Dichtefunktion von $X \sim N(4, 9)$ mit dem Taschenrechner und der obigen Formel.

$$\frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{2-4}{3}\right)^2} = 0.10648; \quad \frac{1}{3\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{4-4}{3}\right)^2} = 0.13298$$

b) Wir berechnen für $x_i = -10, -9, \dots, 16, 17$ die Funktionswerte $f_X(x_i)$ der Dichtefunktion von $X \sim N(4, 9)$ mit Hilfe von Excel und tragen die Punkte in einem Koordinatensystem ab. Wir verbinden die Punkte mit einer glatten Kurve. Mit EXCEL werden die Werte der Dichtefunktion in $x = 2$ der Normalverteilung $N(4, 9)$ berechnet mit (unter $f(x)$, Statistik, NORM.VERT(x;Mittelwert;Standabwn;kumuliert). Bei „Mittelwert“ ist der Erwartungswert einzugeben, bei „Stabw“ die Standardabweichung, bei „kumuliert“ die Zahl 0 (s. Abbildung 14.1.1).

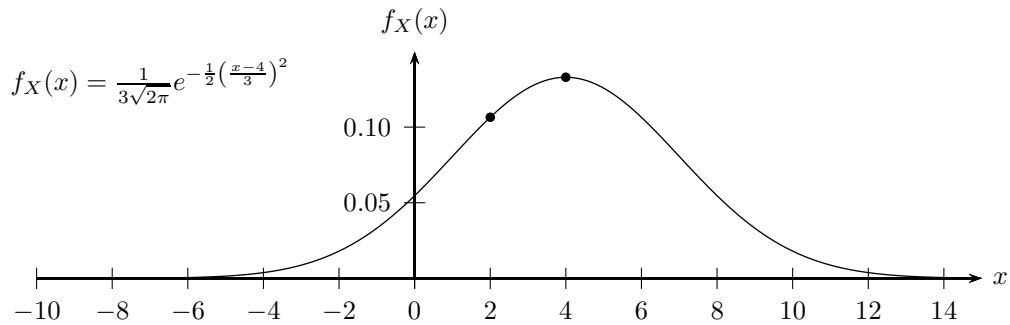


Abbildung 14.1.1: Dichtefunktion der Normalverteilung mit $\sigma = 3$ und $\mu = 4$ sowie den Punkten $(2, f_X(2))$ und $(4, f_X(4))$

14.2 Eigenschaften

Die Dichtefunktionen der Normalverteilungen sind um μ (= Erwartungswert) symmetrische Funktionen. Damit gilt $P(X \leq \mu) = 0.5$ und der Median fällt mit dem Erwartungswert zusammen. Durch Veränderung von μ können wir die Kurve nach links oder nach rechts verschieben. Je größer μ , desto mehr wird die Kurve nach rechts verschoben (s. Abbildungen 14.2.2 bis 14.2.5).

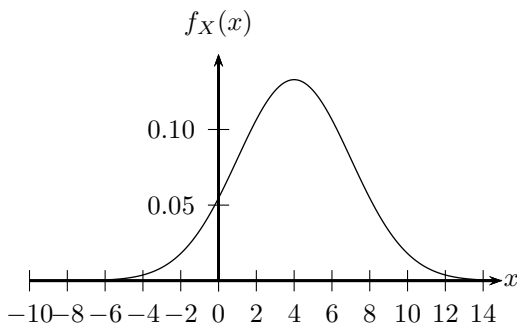


Abbildung 14.2.2: Dichtefunktion der Normalverteilung mit $\sigma = 3$ und $\mu = 4$

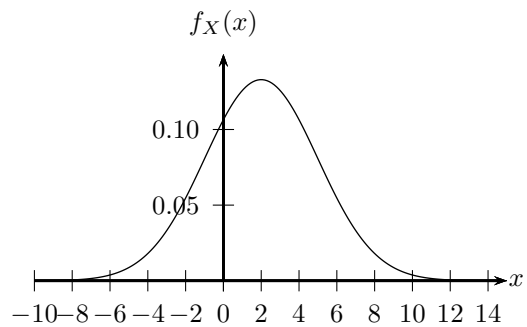
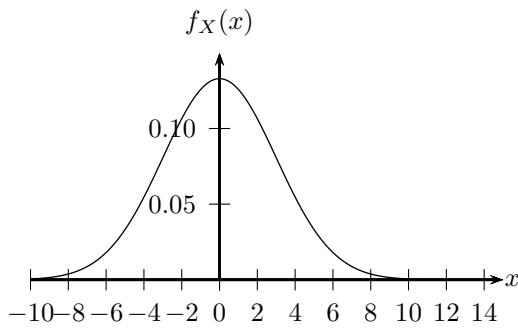
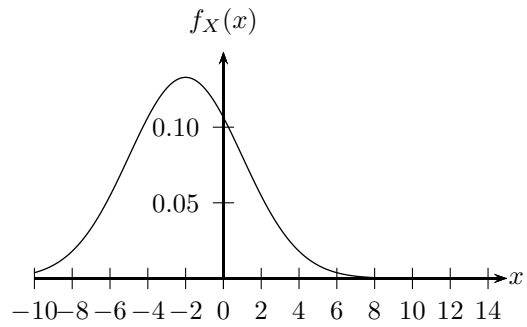
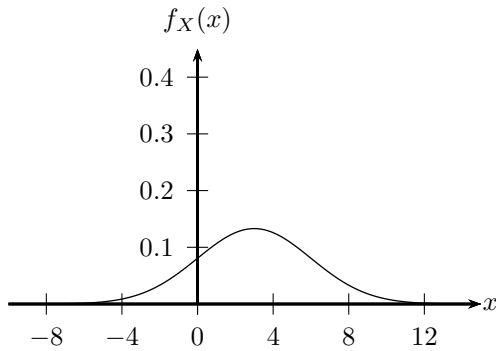
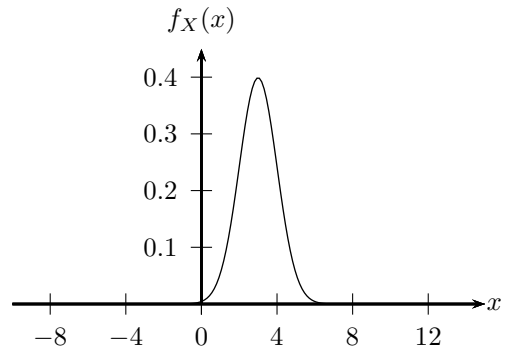


Abbildung 14.2.3: Dichtefunktion der Normalverteilung mit $\sigma = 3$ und $\mu = 2$

Abbildung 14.2.4: Dichtefunktion der Normalverteilung mit $\sigma = 3$ und $\mu = 0$ Abbildung 14.2.5: Dichtefunktion der Normalverteilung mit $\sigma = 3$ und $\mu = -2$

Die Varianz σ^2 gibt an, wie steil oder wie flach die Dichtefunktion einer Normalverteilung ist: Je kleiner die Varianz ist, desto steiler läuft die Kurve beim Erwartungswert zusammen (s. Abbildungen 14.2.6 und 14.2.7).

Abbildung 14.2.6: Dichtefunktion der Normalverteilung mit $\sigma = 3$ und $\mu = 3$ Abbildung 14.2.7: Dichtefunktion der Normalverteilung mit $\sigma = 1$ und $\mu = 3$

Die Standardabweichung schneidet beidseitig des Erwartungswertes ca. 0.68 ab (s. Abbildung 14.2.8), d.h:

$$P(\mu - \sigma < X \leq \mu + \sigma) \approx 0.68 \text{ (ca. 68\%)}$$

Zudem gilt (s. Abbildung 14.2.9):

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) \approx 0.955 \text{ (ca. 95\%)}$$

$$P(\mu - 3\sigma < X \leq \mu + 3\sigma) \approx 0.9994 \text{ (ca. 99.9\%)}$$

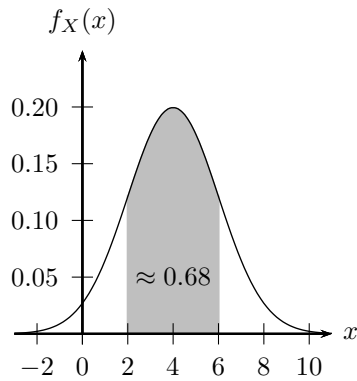


Abbildung 14.2.8: Fläche zwischen der Dichtefunktion der Normalverteilung und der x-Achse im Intervall $[\mu - \sigma, \mu + \sigma]$ beträgt ca. 0.68; im Beispiel gilt $\mu = 4$ und $\sigma = 2$

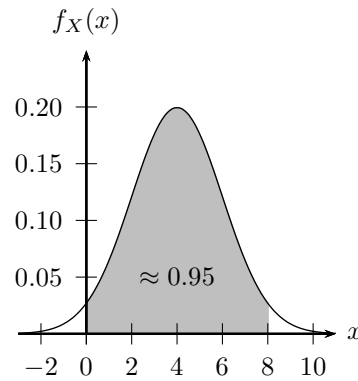


Abbildung 14.2.9: Fläche zwischen der Dichtefunktion der Normalverteilung und der x-Achse im Intervall $[\mu - 2\sigma, \mu + 2\sigma]$ beträgt ca. 0.95; im Beispiel gilt $\mu = 4$ und $\sigma = 2$

Ähnlich wie Daten (s. Seite 63, beschreibende Statistik) können auch Zufallsvariablen standardisiert werden.

Definition 14.2.1. Sei eine Zufallsvariable X gegeben. Wir nennen die Zufallsvariable

$$Z := \frac{X - E(X)}{\sqrt{V(X)}}$$

standardisiert.

Für standardisierte Zufallsvariablen gilt der Satz:

Satz 14.2.2. Wenn $Z := \frac{X - E(X)}{\sqrt{V(X)}}$ für eine Zufallsvariable X , dann gilt:

$$E(Z) = 0$$

$$V(Z) = 1$$

Beweis.

$$\begin{aligned} E(Z) &= E\left(\frac{X - E(X)}{\sqrt{V(X)}}\right) \\ &\stackrel{\text{Satz 9.2.10}}{=} \frac{1}{\sqrt{V(X)}} E(X - E(X)) \\ &\stackrel{\text{Satz 9.5.13}}{=} \frac{1}{\sqrt{V(X)}} (E(X) - E(E(X))) \\ &\stackrel{\text{Satz 9.2.9}}{=} \frac{1}{\sqrt{V(X)}} (E(X) - E(X)) \\ &= 0 \end{aligned}$$

$$\begin{aligned}
V(Z) &= V\left(\frac{X - E(X)}{\sqrt{V(X)}}\right) \\
&\stackrel{\text{Satz 9.5.21}}{=} \frac{1}{V(X)} V(X - E(X)) \\
&\stackrel{\text{Satz 9.5.22}}{=} \frac{1}{V(X)} V(X) \\
&= 1
\end{aligned}$$

□

Standardisieren wir eine normalverteilte Zufallsvariable, erhalten wir eine Zufallsvariable X , für die gilt: $X \sim N(0, 1)$. Wir nennen sie „standardnormalverteilt“. Für die Formel der entsprechenden Dichtefunktion f_X gilt ($f_X : \mathbb{R} \rightarrow \mathbb{R}^+$) (s. Abbildung 14.2.10):

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2} = f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

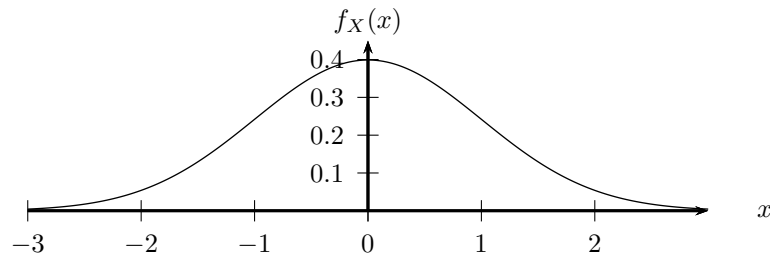


Abbildung 14.2.10: Dichtefunktion der Standardnormalverteilung ($\sigma = 1$ und $\mu = 0$)

Ohne Beweis halten wir fest:

Satz 14.2.3. Für unabhängige Zufallsvariablen $X_i \sim N(\mu_i, \sigma_i^2)$ ist die Statistik Summe normalverteilt, und zwar gilt:

$$T := \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Satz 14.2.4. Ist $X \sim N(\mu, \sigma^2)$, dann ist $aX + b \sim N(a\mu + b, a^2\sigma^2)$ für $a \in \mathbb{R} \setminus \{0\}$ und $b \in \mathbb{R}$.

Die Sätze implizieren, dass die Statistik Mittelwert von *i.i.d.* Zufallsvariablen $X_i \sim N(\mu, \sigma^2)$ normalverteilt ist mit den Parametern μ und $\frac{\sigma^2}{n}$. Denn dann gilt gemäss Satz 14.2.3

$$T := \sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$$

und gemäss Satz 14.2.4

$$\bar{X} = \frac{1}{n}T \sim N\left(\frac{1}{n}n\mu, \frac{1}{n^2}n\sigma^2\right) = N\left(\mu, \frac{\sigma^2}{n}\right).$$

14.3 Berechnung der Werte von $P(a < X \leq b)$

Die Werte $P(a < X \leq b)$ für $X \sim N(0, 1)$ werden mit Näherungsverfahren berechnet. In der Praxis benutzen wir zur Bestimmung von Wahrscheinlichkeiten bei normalverteilten Zufallsvariablen Rechner, Computerprogramme oder Tabellen. Um das Gefühl für die Normalverteilung und das,

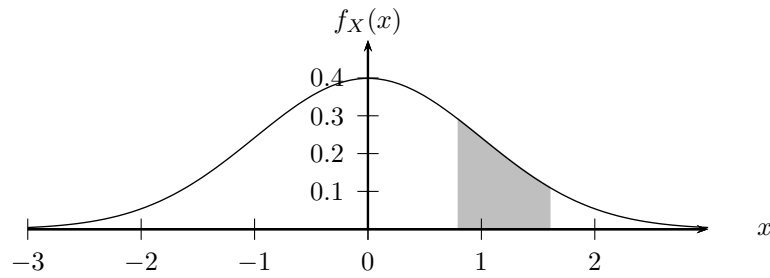


Abbildung 14.3.11: Zu berechnende Fläche zwischen der Dichtefunktion der Standardnormalverteilung und der x-Achse im Intervall $[0,8, 1,6]$

was sie ausdrückt, zu verbessern, berechnen wir trotzdem näherungsweise eine Teilfläche unter der Dichtefunktion der Standardnormalverteilung. Wir wollen die Fläche im Intervall $]0,8, 1,6] =]a, b]$ unter der Dichtefunktion der Standardnormalverteilung berechnen (s. Abbildung 14.3.11).

Ein praktisches Näherungsverfahren stellt die „Trapezregel“ dar. Wir können die Fläche zwischen den Punkten $(a, 0)$, $(a, f(a))$, $(b, f(b))$, $(b, 0)$ in ein Viereck und ein Dreieck aufteilen (= Trapez; siehe Abbildung 14.3.12), wobei das Dreieck im Fall $f(a) = f(b)$ zur Strecke wird. Die Fläche dieses Trapezes stimmt näherungsweise mit der gesuchten Fläche unter der Dichtefunktion im Intervall $]a, b]$ überein.

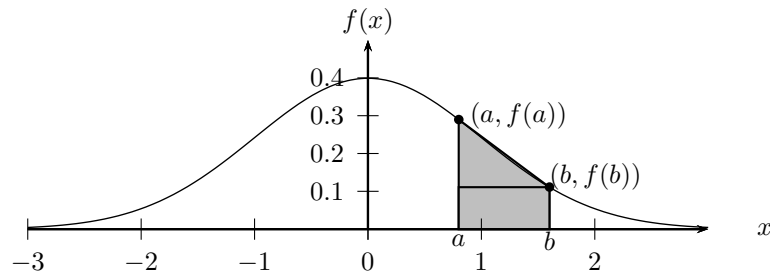


Abbildung 14.3.12: Trapez (= Rechteck + Dreieck), dessen Fläche näherungsweise mit der gesuchten Fläche von Abbildung 14.3.11 übereinstimmt. Die Fläche des Rechtecks beträgt Basis $\cdot$ Höhe: $(b - a) \cdot f(b) = (1,6 - 0,8) \cdot f(1,6)$; Die Fläche des Dreiecks beträgt Basis $\cdot$ Höhe $\cdot \frac{1}{2}$: $(b - a) \cdot (f(a) - f(b)) \cdot \frac{1}{2}$

Die Fläche des Trapezes ist für $f(a) > f(b)$ und $b > a$ (= Rechteck + Dreieck; s. Abbildung 14.3.12):

$$f(b)(b - a) + \frac{1}{2}(f(a) - f(b))(b - a)$$

Es gilt:

$$\begin{aligned} & f(b)(b - a) + \frac{1}{2}(f(a) - f(b))(b - a) \\ &= (b - a) \left(f(b) + \frac{1}{2}(f(a) - f(b)) \right) \\ &= (b - a) \left(f(b) + \frac{1}{2}f(a) - \frac{1}{2}f(b) \right) \\ &= (b - a) \left(\frac{1}{2}f(a) + \frac{1}{2}f(b) \right) \\ &= \frac{b - a}{2}(f(a) + f(b)) \end{aligned}$$

Diese Formel gilt allgemein: für $f(a) < f(b)$ erhalten wir nämlich dasselbe Resultat, wie Sie

selber nachrechnen können. Für $f(a) = f(b)$ erhalten wir:

$$\begin{aligned} f(b)(b-a) + \frac{1}{2}(f(a) - f(b))(b-a) \\ = f(b)(b-a), \end{aligned}$$

da $f(a) - f(b) = 0$. Dasselbe Resultat erhalten wir aus

$$\begin{aligned} \frac{b-a}{2}(f(a) + f(b)) \\ = \frac{b-a}{2}2f(b) \\ = (b-a)f(b), \end{aligned}$$

denn $f(a) + f(b) = 2f(b)$ für $f(a) = f(b)$.

In unserem Beispiel können wir berechnen:

$$\begin{aligned} \frac{b-a}{2}(f(a) + f(b)) \\ = \frac{1.6 - 0.8}{2} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(0.8)^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1.6)^2} \right) \\ = 0.16024 \end{aligned}$$

Wir vergleichen das Resultat mit dem Resultat von Excel: $F_X(1.6) - F_X(0.8) \approx 0.157056044$

Die Näherung ist um weniger als $4^\circ/\text{oo}$ neben dem Excel-Resultat, wobei im vorliegenden Fall die Kurve der Dichtefunktion wenig von der Hypotenuse des Dreiecks abweicht. Die Näherung könnte bis auf Rundungsfehler *beliebig* verbessert werden, indem wir die Basis der zu berechnende Fläche in kleinere Intervalle einteilen und jeweils die Flächen darüber mit der Trapezmethode berechnen. Dies wäre vor allem in den Intervallen angemessen, wo die Dichtefunktion der Standardnormalverteilung stärker gekrümmt ist und damit durch die Hypotenuse des Dreiecks weniger gut angenähert wird als im Beispiel.

Wir würden also berechnen:

$$F = \sum_{i=0}^{n-1} \left(\frac{x_{i+1} - x_i}{2} (f(x_{i+1}) + f(x_i)) \right)$$

mit $x_0 = a$ und $x_n = b$ und $x_i \in]b, a[$ für $1 \leq i \leq n-1$ und $x_{i+1} > x_i$. Im Allgemeinen wählt man dabei zwischen den x_i gleich lange Abstände, wodurch man die Formel zusätzlich vereinfachen könnte.

Wir berechnen die näherungsweise Fläche A im betrachteten Intervall $[0.8, 1.6]$ bei einer zusätzlichen Unterteilung des Intervalls in $]0.8, 1.2] \cup]1.2, 1.6] =]0.8, 1.6]$ (s. Abbildung 14.3.13)

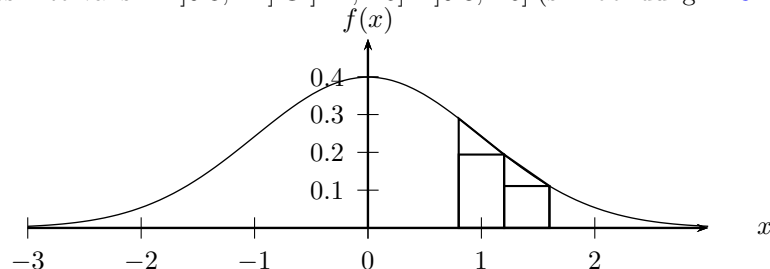


Abbildung 14.3.13: Zwei Trapeze, so dass die Summe derer Fläche näherungsweise mit der gesuchten Fläche von Abbildung 14.3.11 übereinstimmt

$$\begin{aligned}
A_1 + A_2 &= A \\
A_1 &\approx \frac{(1.6-1.2)}{2} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1.6)^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1.2)^2} \right) \approx 6.1021 \times 10^{-2} \\
A_2 &\approx \frac{(1.2-0.8)}{2} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1.2)^2} + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(0.8)^2} \right) \approx 9.6776 \times 10^{-2} \\
A &\approx 6.1021 \times 10^{-2} + 9.6776 \times 10^{-2} = 0.1578
\end{aligned}$$

Die Näherung ist schon bedeutend besser geworden. Es ist einfach, z.B. mit Excel noch bessere Näherungen zu berechnen, indem die Unterteilung durch zusätzliche Intervalle verfeinert wird.

In der Praxis werden wir die Werte $P(a < X \leq b)$ bei normalverteilten Zufallsvariablen nicht auf diese Art berechnen. Wir verwenden Programme, Rechner oder Tabellen (*Mit Excel*: Statistikfunktionen, =NORM.VERT($x, \mu, \sigma, 1$) liefert $P(x \leq X)$ für $X \sim N(\mu, \sigma^2)$). Dabei wird unter x der Wert x eingegeben, unter „Mittelwert“ der Erwartungswert, unter „Standardabweichung“ die Standardabweichung, unter „kumuliert“ 1 (0 liefert den Wert der Dichtefunktion von x). Statt $\Phi(x)$ durch =Norm.vert($x, 0, 1$) zu berechnen, kann man diesen Wert auch durch „=standnormvert(x)“ erhalten).

Mit Tabellen: In Tabellen werden nur die Werte für die Standardnormalverteilung geliefert. Daten, die normalverteilt sind, können jedoch leicht in Daten transformiert werden, die standardnormalverteilt sind. Dies geschieht durch die Standardisierung. Wir möchten $P(X \leq x_i)$ berechnen mit $X \sim N(\mu, \sigma^2)$. Dann können wir $P(Z \leq z_i)$ berechnen mit $z_i = \frac{x_i - \mu}{\sigma}$ und $Z \sim N(0, 1)$.

14.4 Verteilungsfunktion

Die Verteilungsfunktion der Normalverteilung ist definiert durch:

$$F_X(y) = \int_{-\infty}^y \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \right) dx$$

Sie ordnet jedem $y \in \mathbb{R}$ die Fläche zwischen der Dichtefunktion und der x -Achse im Intervall $]-\infty, y]$ zu. Für die Standardnormalverteilung definieren wir:

$$\Phi(z) = \int_{-\infty}^z \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right) dx$$

Es wird das spezielle Zeichen Φ für die Verteilungsfunktion der Standardnormalverteilung eingeführt, da wir diese oft brauchen werden. Die näherungsweise berechneten Punkte von Φ erlauben es, eine punktweise Zeichnung der Verteilungsfunktion der Standardnormalverteilung zu erstellen, wobei die Punkte durch eine Kurve verbunden werden können. Der Plot der Punkte

$$\left(-3 + i\frac{1}{4}, \Phi(-3 + i\frac{1}{4}) \right)$$

für $i \in \mathbb{N}_{24}$ ergibt die folgende Graphik (s. Abbildung 14.4.14).

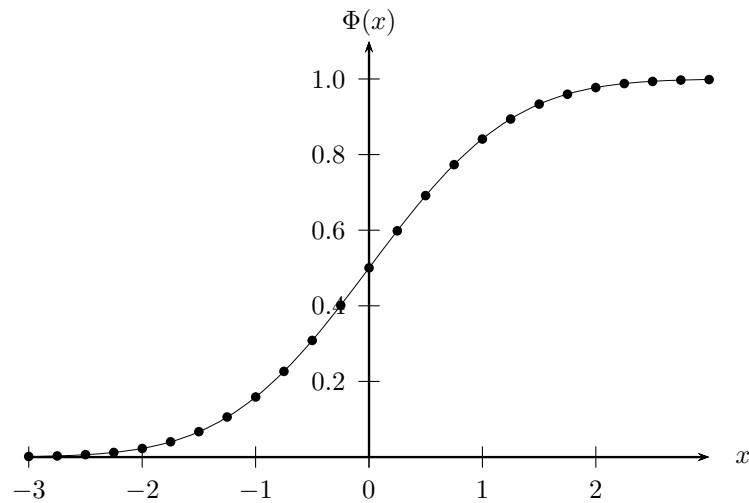


Abbildung 14.4.14: Graphische Darstellung der Verteilungsfunktion der Standardnormalverteilung

Zeichnen wir die punktweise empirische Verteilungsfunktion von standardisierten Daten (s. S. 63), so können wir davon ausgehen, dass sie normalverteilt sind, wenn die Punkte ungefähr auf der obigen Kurve liegen. Die Kurve ist punktsymmetrisch in $(0, 0.5)$ und steigt stark nahe bei 0 und schwächer für Punkte, die weiter weg von 0 liegen. Ohne eine Standardisierung vorzunehmen, sind die Daten normalverteilt, wenn Sie auf einer Kurve liegen, die bis auf einen Streckungsfaktor die obige Form hat. Die Verteilung schätzen wir, indem wir die Parameter aus den Daten schätzen (Erwartungswert durch den Mittelwert der Daten, Varianz durch die Varianz der Daten).

14.5 Zusammenhang von Normalverteilung und Binomialverteilung

Es besteht ein enger Zusammenhang zwischen den Binomialverteilungen und der Standardnormalverteilung. Wir können die Wahrscheinlichkeitsfunktion der Binomialverteilungen durch eine Dichtefunktion auf $\mathbb{R}$ darstellen: wir definieren in den Intervallen $[x_i - 0.5, x_i + 0.5]$ die Dichtefunktion f jeweils durch die konstante Funktion $f_i(x) = P(X = x_i)$, wobei $X \sim B(n, p)$, und setzen für alle übrigen $x \in \mathbb{R}$: $f_X(x) = 0$, d.h. $f_X(x) = 0$ in $]-\infty, -0.5]$ und in $]n + 0.5, \infty[$. Für $n = 5$ und $p = 0.5$ erhalten wir z.B. die folgende „Treppenfunktion“ (s. Abbildung 14.5.15):

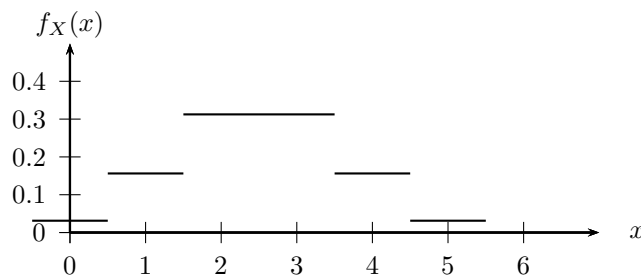


Abbildung 14.5.15: Dichtefunktion, die aus der Wahrscheinlichkeitsfunktion der Binomialverteilung mit $n = 5$ und $p = 0.5$ durch Fortführen von $P(X = x_i)$ auf $[x_i - 0.5, x_i + 0.5]$ entsteht ($f_X(x) = 0$ in $]-\infty, -0.5]$ und in $]5.5, \infty[$). Zu beachten ist, dass im Beispiel $P(X = 2) = P(X = 3)$ ist.

Für die Flächen über $[x_i - 0.5, x_i + 0.5]$ gilt, dass sie mit $P(X = x_i)$ identisch sind, da die Basis jeweils 1 ergibt. Entsprechend ist die Summe all dieser Flächen mit 1 identisch, da sie mit der

Summe der Werte der Wahrscheinlichkeitsfunktion der Binomialverteilung identisch ist. Entsprechend ist die definierte Funktion f eine Dichtefunktion (s. für die Fläche unter der Dichtefunktion von Abbildung 14.5.15 die graue Fläche von Abbildung 14.5.16).

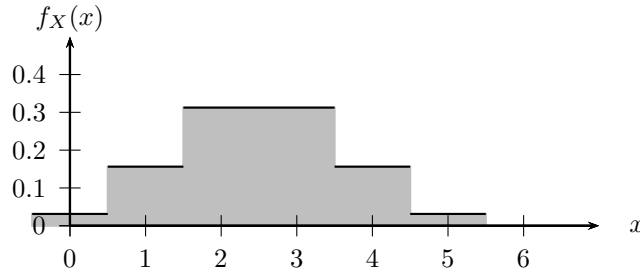


Abbildung 14.5.16: Fläche (grau) unter der Dichtefunktion von Abbildung. Zu beachten ist, dass im Beispiel $P(X = 2) = P(X = 3)$ ist

Wir transformieren nun die Dichtefunktion von Abbildung 14.5.15, indem wir wie folgt vorgehen: Wir führen eine Standardisierung auf die Intervallgrenzen $x_i \pm 0.5$ durch, d.h. wir berechnen $\frac{(x_i \pm 0.5) - np}{\sqrt{np(1-p)}}$. Die neue Dichtefunktion h wird dann auf $\left[\frac{(x_i - 0.5) - np}{\sqrt{np(1-p)}}, \frac{(x_i + 0.5) - np}{\sqrt{np(1-p)}} \right]$ durch h_i definiert, so dass gilt:

$$P(X = x_i) = h_i \cdot \left(\frac{(x_i + 0.5) - np}{\sqrt{np(1-p)}} - \frac{(x_i - 0.5) - np}{\sqrt{np(1-p)}} \right)$$

Damit ist die Fläche über $\left[\frac{(x_i - 0.5) - np}{\sqrt{np(1-p)}}, \frac{(x_i + 0.5) - np}{\sqrt{np(1-p)}} \right]$ mit $P(X = x_i)$ identisch. In $\left[-\infty, \frac{(0 - 0.5) - np}{\sqrt{np(1-p)}} \right]$ und $\left[\frac{(n + 0.5) - np}{\sqrt{np(1-p)}}, \infty \right]$ wird $h = 0$ gesetzt. Wir erhalten fürs Beispiel der Abbildung 14.5.15 die folgenden Daten (s. Tabelle 14.5.1):

x_i	$P(X = x_i)$	Intervallgrenzen	standardisierte Intervallgrenzen	Höhen h_i
		-0.5	-2.683281573	
0	0.03125	0.5	-1.788854382	0.034938562
1	0.15625	1.5	-0.894427191	0.174692811
2	0.3125	2.5	0	0.349385621
3	0.3125	3.5	0.894427191	0.349385621
4	0.15625	4.5	1.788854382	0.174692811
5	0.03125	5.5	2.683281573	0.034938562

Tabelle 14.5.1: Berechnung der Höhen über den standardisierten Intervallen fürs Beispiel der Abbildung 14.5.15

Die derart berechnete neue Dichtefunktion h zeichnen wir nun zusammen mit der Dichtefunktion der Standardnormalverteilung in ein Koordinatennetz. Wir erhalten fürs Beispiel $n = 5$ und $p = 0.5$ (s. Abbildung 14.5.17):

Wir betrachten noch die Zeichnung für $n = 10$ und $n = 20$, sowie $p = 0.5$ (s. Abbildungen 14.5.18 und 14.5.19):

Wir stellen fest, dass mit zunehmendem n die durch Standardisierung gewonnene Dichtefunktion sich der Standardnormalverteilung nähert - oder anders formuliert: die durch Standardisierung gewonnene Dichtefunktion *konvergiert* gegen die Dichtefunktion der Standardnormalverteilung. Dies führt uns zum folgenden Satz, wobei wir folgendes festlegen: h ist die Grenzfunktion einer Folge von Funktionen h_n (symbolisch: $\lim_{n \rightarrow \infty} h_n = h$) genau dann, wenn $h(x) = \lim_{n \rightarrow \infty} h_n(x)$ für alle $x \in \mathbb{R}$.

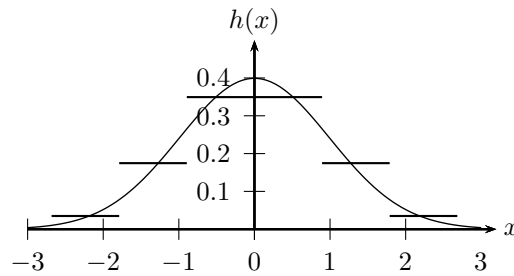


Abbildung 14.5.17: Durch z-Transformation der Intervalle und Konstanthaltung der Wahrscheinlichkeiten aus der Dichtefunktion f (Beispiel der Abbildung 14.5.15) entstandene Dichtefunktion h ($n = 5, p = 0.5$)

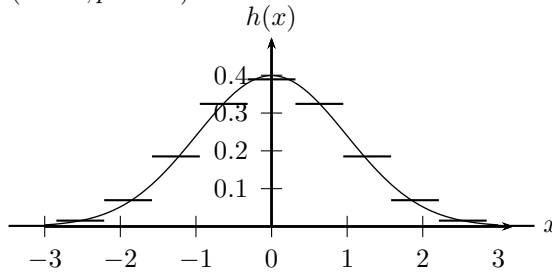


Abbildung 14.5.18: Durch z-Transformation der Intervalle und Konstanthaltung der Wahrscheinlichkeiten aus der Dichtefunktion der Binomialverteilung mit $p = 0.5, n = 10$ entstandene Dichtefunktion h

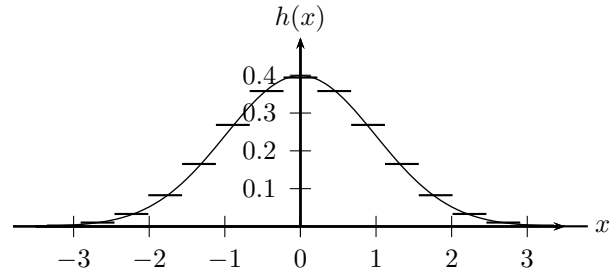


Abbildung 14.5.19: Durch z-Transformation der Intervalle und Konstanthaltung der Wahrscheinlichkeiten aus der Dichtefunktion der Binomialverteilung mit $p = 0.5, n = 20$ entstandene Dichtefunktion h

Satz 14.5.1. Für $X \sim B(n, p)$ sei

$$h_n(x) = \begin{cases} \frac{P(X=x_i)}{\frac{(x_i+0.5)-np}{\sqrt{np(1-p)}} - \frac{(x_i-0.5)-np}{\sqrt{np(1-p)}}} & \text{für } x \in \left[\frac{(x_i-0.5)-np}{\sqrt{np(1-p)}}, \frac{(x_i+0.5)-np}{\sqrt{np(1-p)}} \right] \\ & \text{und } 0 \leq i \leq n \\ 0 & \text{sonst} \end{cases}$$

dann gilt:

(a) $\lim_{n \rightarrow \infty} h_n = f$ (f ist die Dichtefunktion der Standardnormalverteilung)

(b) Für $X \sim B(n, p)$ und $n \rightarrow \infty$ gilt: $P(X \leq x) = \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$ (Satz von de Moivre-Laplace, Abraham de Moivre, 1667 - 1754, französischer Mathematiker, für Pierre-Simon Laplace s. oben).

Ein Beweis des Satzes ist anspruchsvoll (s. Homepage <http://math.logik.ch>) und unterbleibt hier.

Auf Grund des Zusammenhangs der Standardnormalverteilung und der Familie der Binomialverteilungen können wir die Standardnormalverteilung verwenden, um näherungsweise Werte der Binomialverteilung zu berechnen, sofern n genügend gross ist. Die Näherungen sind zudem besser, wenn p nahe bei 0.5 liegt - liegt p in der Nähe von 0 oder 1, muss n um etliches grösser sein, um dies wettzumachen. Als Faustregel gilt, dass die Näherungen genügend exakt sind, wenn $np(1-p) \geq 9$. Bei der Berechnung der Wahrscheinlichkeiten der Binomialverteilung mit Hilfe der Standardnormalverteilung verwenden wir die folgenden Näherungsformel:

für $X \sim B(n, p)$ gilt:

$$P(X \leq x) \approx \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$$

Dabei ist $\frac{x-np}{\sqrt{np(1-p)}}$ die Standardisierung von x , da np der Erwartungswert und $\sqrt{np(1-p)}$ die

Standardabweichung der entsprechenden Binomialverteilung ist.

Traditionell und ohne Computer war eine Berechnung von $P(a < X \leq b)$ und $X \sim B(n, p)$ für grosse n und grosse Differenzen $b-a$ mühsam. Deshalb war früher das eingeführte Näherungsverfahren nützlicher als heute. Allerdings verwenden wir auch heute noch die Normalverteilung zur Berechnung von Wahrscheinlichkeiten der Binomialverteilung, wenn auf Grund der Grösse der Zahlen Computer austeigen.

Beispiel 14.5.1. Man möchte $P(X \leq 12)$ für $X \sim B(50, 0.5)$ berechnen. Es gilt $50 \cdot 0.5 \cdot (1 - 0.5) = 12.5 \geq 9$.

Wir erhalten den folgenden exakten Wert: $P(X \leq 12) = 0.000152932$

Mit der Standardnormalverteilung erhalten wir:

$$\frac{12-50 \cdot 0.5}{\sqrt{50 \cdot 0.5^2}} = -3.676955262$$

$$\text{und } \Phi\left(\frac{12-50 \cdot 0.5}{\sqrt{50 \cdot 0.5^2}}\right) = \Phi(-3.676955262) = 0.000118017$$

Die Differenz der beiden Werte ist klein.

◇

14.5.1 Übungen

1. Berechnen Sie folgende Wahrscheinlichkeiten für $X \sim N(5, 36)$:

- (a) $P(X \leq 7)$
- (b) $P(5 < X \leq 8)$
- (c) $P(6 < X \leq 9)$
- (d) $P(X \geq 8.5)$
- (e) $P(X \leq 3)$
- (f) $P(2 < X \leq 4)$

2. Berechnen Sie folgende Wahrscheinlichkeiten

- (a) für $X \sim N(3, 8)$: $P(X \leq 7)$
- (b) für $X \sim N(6, 8)$: $P(5 < X \leq 8)$
- (c) für $X \sim N(10, 80)$: $P(6 < X \leq 9)$
- (d) für $X \sim N(9, 70)$: $P(X < 8.5)$
- (e) für $X \sim N(5, 16)$: $P(2 < X \leq 8)$
- (f) für $X \sim N(5, 16)$: $P(X \leq 5)$

3. Berechnen Sie mit Excel und dem oben angegebenen Näherungsverfahren die Fläche zwischen der Dichtefunktion der Standardnormalverteilung und der x -Achse im Intervall $]0.5, 1]$, indem Sie zuerst das gesamte Intervall betrachten, dann eine Zweiteilung und eine Vierteilung des Intervalls vornehmen. Vergleichen Sie die Lösungen mit den Ergebnissen, die Sie mit Hilfe des =NORM.VERT-Befehls von Excel berechnen.

4. Berechnen Sie mit Hilfe der Normalverteilung die folgenden Werte der folgenden Binomialverteilungen (wobei Sie jeweils die Voraussetzung für die Näherung überprüfen):

- (a) $P(X > 23)$ für $X \sim B(40, 0.4)$
- (b) $P(X \geq 30)$ für $X \sim B(45, 0.5)$
- (c) $P(X \leq 18)$ für $X \sim B(50, 0.3)$
- (d) $P(X < 2)$ für $X \sim B(40, 0.45)$

- (e) $P(X \geq 24)$ für $X \sim B(100, 0.44)$
5. Schätzen Sie für die Daten der Excel-Tabelle „11bNormalverteilung Uebung Anpassung an Daten.xls“ Verteilungsfunktionen und berechnen Sie die Wahrscheinlichkeiten $P(X \leq 15)$ für die drei Datensätze.
 6. Finanzmathematisches Beispiel: Der Gewinn G (in Franken) einer Kapitalanlage K sei normalverteilt (der Erwartungswert von $K + G$ liege 5% ob der Einlage (Einlage $K = 100\%$), die Standardabweichung sei die Differenz zwischen dem Erwartungswert und dem Wert, der 7% (Erwartungswert = 100%) unterhalb des Erwartungswertes liegt. Sei $K = 1000$.
 - (a) Berechnen Sie die Wahrscheinlichkeit, dass die Kapitalanlage einen Verlust einführt.
 - (b) Berechnen Sie die Wahrscheinlichkeit, dass die Kapitalanlage einen Gewinn abwirft.
 - (c) Berechnen Sie die Wahrscheinlichkeit, dass die Kapitalanlage einen Gewinn von über 10% abwirft.
 - (d) Berechnen Sie die Wahrscheinlichkeit, dass die Kapitalanlage einen Gewinn von über 12% abwirft.
 - (e) Berechnen Sie die Wahrscheinlichkeit, dass die Kapitalanlage einen Gewinn zwischen 2% und 7% der Anlage abwirft.
 7. Die Gewinne zweier Kapitalanlagen seien normalverteilt. Der Gewinn der Anlage K_1 weise einen Erwartungswert von 6% von K_1 auf und eine Standardabweichung von 8% von K_1 . Der Gewinn der Anlage K_2 weise einen Erwartungswert von 8% von K_2 auf und eine Standardabweichung von 12% von K_2 . Berechnen Sie den Erwartungswert und die Standardabweichung des Gewinns der Gesamtanlage (Unabhängigkeit der Zufallsvariablen; $K_1 = 1000, K_2 = 2000$). Geben Sie den neuen Erwartungswert und die neue Standardabweichung in Prozenten von $K_1 + K_2$ an.
 8. Eine Kapitalanlage habe für den Gewinn einen Erwartungswert von 60 Sfr. und eine Varianz von 80 Sfr².
 - (a) Berechnen Sie den Erwartungswert und die Varianz des Gewinns dieser Kapitalanlage in Euro (Kurs: 1.5).
 - (b) Berechnen Sie den Erwartungswert, die Varianz und die Standardabweichung des Gewinns dieser Kapitalanlage in Prozenten (für sFr. und Euro, $K = 1000$ sFr.)
 9. a) Die Korrelation r der Gewinne zweier Kapitalanlage betrage -0.6 (so verhält sich z.B. die Entwicklung des Goldpreises im Allgemeinen umgekehrt zur Entwicklung von Aktien; $K_1 = 1000, K_2 = 2000$). Berechnen Sie den Erwartungswert und die Standardabweichung der Gewinne der beiden Kapitalanlagen, wenn der Erwartungswert μ_1 der Anlage K_1 7% des Wertes der Kapitalanlage 1 beträgt und der Erwartungswert μ_2 der Anlage K_2 8% der Kapitalanlage 2 ist (Die Standardabweichungen sind $s_1 = 10\%, s_2 = 12\%$).
 b) Vergleichen Sie das Risiko dieser Gesamtanlage mit einer Aufspaltung der Gesamtanlage auf zwei Anlagen, deren Gewinne unkorreliert sind.
 c) Wenn die Gewinne völlig korreliert sind ($r = 1$).
 10. Ein Risikomass für Aktien ist die sogenannte „value at risk“ (VaR). Man geht davon aus, dass die künftigen Werte von Aktien normalverteilt sind. Die VaR_α ist das $(1 - \alpha)$ -Quantil der Verteilung. Man betrachtet einen Wert unter diesem Quantil als hinreichend unwahrscheinlich, so dass man mit genügend grosser Sicherheit mindestens diesen Wert erhält. Berechnen Sie die $\text{VaR}_{0.975}$, wenn folgendes W-Modell der Aktie unterlegt ist ($\mu = 60; \sigma = 10$).
 11. Eine gewisse Berühmtheit erlangte in der Finanzwelt die sogenannte Black-Scholes-Formel. Sie dient dazu, Europäische Optionen zu bewerten - will man Geld in Optionen anlegen, muss man wissen, wieviel diese Wert sind. Man kauft Optionen, wenn diese unter dem Wert gehandelt werden, und man verkauft sie, wenn über ihren Wert gestiegen sind. Unter

gewissen Voraussetzungen, die wir hier nicht anführen, gilt für den Wert einer Call-Option (die Herleitung der Formel ist anspruchsvoll und verlangt Kenntnisse der stochastischen Integration und der Theorie stochastischer Prozesse):

$$V_0 = S_0 \Phi \left(\frac{\ln \left(\frac{S_0}{k} \right) + \left(r + \frac{1}{2} \sigma^2 \right) T}{\sigma \sqrt{T}} \right) - k e^{-rT} \Phi \left(\frac{\ln \left(\frac{S_0}{k} \right) + \left(r - \frac{1}{2} \sigma^2 \right) T}{\sigma \sqrt{T}} \right)$$

mit dem aktuellen Wert des Basiswertes S_0 , dem Ausübungspreis k , dem Ausübungszeitpunkt T , der Volatilität σ (Standardabweichung der Aktie, gewöhnlich in % des Wertes der Aktie pro Jahr gegeben, in der Formel als Anteil zu verwenden), dem (beinahe) risikolosen Zinssatz r (= Zins für Staatsanleihen der Güte AAA). Berechnen Sie den Wert der Call-Option für eine Aktie mit dem aktuellen Wert $S_0 = 60$; $k = 62$; $\sigma = 0.2$ (entspricht einer Volatilität von 20%); $r = 0.05$; $T = 0.5$ (ein halbes Jahr).

14.5.2 Lösungen

1. (a) $P(X \leq 7) = 0.630558659818$
 (b) $P(5 < X \leq 8) = P(X \leq 8) - P(X \leq 5) = 0.191462461274$
 (c) $P(6 < X \leq 9) = F_X(9) - F_X(6) = 0.747507533 - 0.566183837 = 0.181323695$
 (d) $P(X \geq 8.5) = 1 - F_X(8.5) = 0.2798344636$
 (e) $P(X \leq 3) = F_X(3) = 0.369441340182$
 (f) $P(2 < X \leq 4) = F_X(4) - F_X(2) = 0.125278628663$
2. Man erhält:
 - (a) $X \sim N(3, 8) : P(X \leq 7) = F_X(7) = 0.921350396475$
 - (b) $X \sim N(6, 8) : P(5 < X \leq 8) = F_X(8) - F_X(5) = 0.398413133991$
 - (c) $X \sim N(10, 80) : P(6 < X \leq 9) = F_X(9) - F_X(6) = 0.128129223246$
 - (d) $X \sim N(9, 70) : P(X < 8.5) = F_X(8.5) = 0.47617$
 - (e) $X \sim N(5, 16) : P(2 < X \leq 8) = F_X(8) - F_X(2) = 0.546745295246$
 - (f) $X \sim N(5, 16) : P(X \leq 5) = 0.5$
3. Folgende Formel muss eingegeben werden:
 $=(((B1-A1)/2)*((1/((2*3.1415926535)^(1/2)))*EXP(-(1/2)*(B1)^2)+$
 $(1/((2*3.1415926535)^(1/2)))*EXP(-(1/2)*(A1)^2)))$
 (Statt „ $((1/((2*3.1415926535)^(1/2)))*EXP(-(1/2)*(B1)^2)$ “ könnte man aber auch kürzer
 „ $(NORM.VERT(B1;0;1;0)$ “ und statt „ $(1/((2*3.1415926535)^(1/2)))*EXP(-(1/2)*(A1)^2)$ “
 kürzer
 „ $(NORM.VERT(A1;0;1;0)$ “ eingeben).
 In die Spalten B und A setzen wir die Intervallgrenzen b und a und wir erhalten:
 Beim gesamten Intervall: 0.148509013
 Bei Zweiteilung: 0.149538864
 Bei Vierteilung: 0.149796428
 Mit Excelbefehl NORM.VERT: 0.149882273
4. Man erhält:
 - (a) Die Voraussetzung: $0.4 \cdot 0.6 \cdot 40 = 9.6 \geq 9$ ist erfüllt.
 $1 - \Phi \left(\frac{23 - (40 \cdot 0.4)}{\sqrt{40 \cdot 0.6 \cdot 0.4}} \right) = 1 - \Phi(2.2592) = 0.0119354734175$
 Der genaue Wert ist: 0.008341770891.

- (b) b) Die Voraussetzung: $0.5 \cdot 0.5 \cdot 45 = 11.25 \geq 9$ ist erfüllt.

$$1 - \Phi\left(\frac{29 - 45 \cdot 0.5}{\sqrt{45 \cdot 0.5 \cdot 0.5}}\right) = 1 - \Phi(1.9379) = 0.0263177119897$$

Der genaue Wert ist: 0.017848901776

- (c) c) Die Voraussetzung: $0.3 \cdot 0.7 \cdot 50 = 10.5 \geq 9$ ist erfüllt.

$$\Phi\left(\frac{18 - 50 \cdot 0.3}{\sqrt{50 \cdot 0.3 \cdot 0.7}}\right) = \Phi(0.92582) = 0.822730234184$$

Der genaue Wert ist: 0.859440123612

- (d) Die Voraussetzung: $0.45 \cdot 0.55 \cdot 40 = 9.9 \geq 9$ ist erfüllt.

$$\Phi\left(\frac{1 - 40 \cdot 0.45}{\sqrt{40 \cdot 0.45 \cdot 0.55}}\right) = \Phi(-5.4030) = 3.2768 \cdot 10^{-8}$$

Der genaue Wert ist: $1.38831751973 \cdot 10^{-9}$

- (e) Die Voraussetzung: $0.44 \cdot 0.56 \cdot 100 = 24.64 \geq 9$ ist erfüllt.

$$1 - \Phi\left(\frac{23 - 100 \cdot 0.44}{\sqrt{100 \cdot 0.44 \cdot 0.56}}\right) = 1 - \Phi(-4.2306) = 0.99998834656$$

Der genaue Wert ist: 0.999989953116

5. Siehe Excel-Tabelle „11bNormalverteilung Ubung Anpassung an Daten.xls“
(<http://math.logik.ch>) (unter: Mathematik und Statistik - Statistik - Wahrscheinlichkeitstheorie II)

6. Man erhält:

- (a) Der Erwartungswert ist gemäss Angaben 5% über der Einlage. Bei einer Einlage $K = 1000$ ergibt dies: $1000 \cdot (1 + 0.05) = 1050$. Die Standardabweichung ist gemäss Angaben die Differenz zwischen dem Erwartungswert und 93% des Erwartungswerts. Bei einem Erwartungswert von $K = 1050$ ergibt dies $1050 - (1050 \cdot 0.93) = 73.5$.

Die Wahrscheinlichkeit, dass die Kapitalanlage einen Verlust einfährt, ist damit - da gemäss Voraussetzung die Zufallsvariable normalverteilt ist - $P(X \leq 1000)$ mit $X \sim N(1050, 73.5^2)$.

Wir erhalten: $P(X \leq 1000) = 0.24817$

- (b) Die Wahrscheinlichkeit, dass die Kapitalanlage einen Gewinn abwirft, ist: $P(X > 1000) = 1 - 0.24817 = 0.75183$

- (c) 10% von 1000 ist 100. Damit ist die Wahrscheinlichkeit zu berechnen, dass $P(X > 1100)$ für $X \sim N(1050, 73.5^2)$. Es ist $P(X > 1100) = 1 - P(X \leq 1100) = 0.248166091$.

Man könnte dies auch wie folgt - ohne zu rechnen - sehen: Die Wahrscheinlichkeit, dass die Wahrscheinlichkeit mehr als 10% abwirft, liegt 5% über dem Erwartungswert. Da die Normalverteilung symmetrisch um den Erwartungswert ist, gilt $P(X \leq 1000) = P(X > 1100)$. Siehe folgende Abbildung 14.5.20:

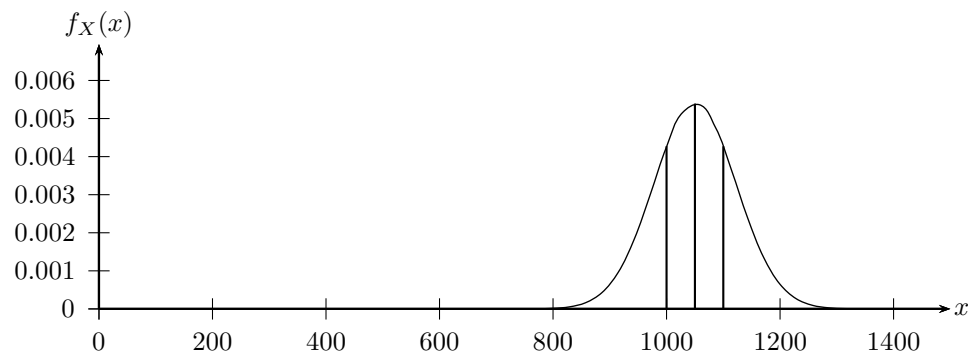


Abbildung 14.5.20: Dichtefunktion der Normalverteilung mit $\mu = 1050$ und $\sigma = 73.5$; Beispiel für Nutzung der Symmetrieeigenschaft der Normalverteilung

- (d) Die Wahrscheinlichkeit, dass die Kapitalanlage einen Gewinn von mehr als 12% abwirft ist: $P(X > 1000(1 + 0.12)) = P(X > 1120)$ mit $X \sim N(1050, 73.5^2)$

Wir erhalten: $P(X \leq 1120) = 0.82955$

$P(X > 1120) : 1 - 0.82955 = 0.17045$

- (e) $P(X \leq 1000(1 + 0.07)) - P(X \leq 1000(1 + 0.02)) = 0.26565$

Bemerkungen: (a) Bei der Anwendung von Verteilungsmodellen auf Gewinne ergibt sich das Problem, dass Anlagemärkte im Allgemeinen recht volatil sind. Zwar kann man problemlos überprüfen, welche Verteilungen Gewinne in der Vergangenheit hatten. Es ist aber schwierig, die Vergangenheit auf die Zukunft zu verlängern (= extrapolieren).

(b) Wir könnten auch mit dem Modell $X \sim N(50, 73.5^2)$ rechnen. Wir haben dann eine Zufallsvariable, welche Gewinne oder Verluste annimmt - ohne das investierte Kapital. Der Erwartungswert bei einem Kapital von 1000 ist dann 50 (5% von 1000). Die Standardabweichung ist die selbe wie oben. Es ergibt sich dann ein Verlust, wenn $X \leq 0$, ein Gewinn wenn $X > 0$, etc.

7. $\mu_1 = 60; \sigma_1 = 80; \mu_2 = 160; \sigma_2 = 240$

Gemäss Satz 9.5.11 (s. Unterkapitel „Statistiken“ im Kapitel „Zufallsvariablen“) gilt:

$$E(X_1 + X_2) = E(X_1) + E(X_2) \implies \mu_{1+2} := \mu_1 + \mu_2 = 60 + 160 = 220$$

In Prozenten bei einem Kapital von 3000: $\frac{220 \cdot 100}{3000} = 7.3333\%$

Die Varianz zweier unabhängiger Zufallsvariablen ist gemäss Satz 9.5.34 (s. Unterkapitel „Statistiken“ im Kapitel „Zufallsvariablen“, die Kovarianz ist wegen der Unabhängigkeit 0): $V(X_1 + X_2) = V(X_1) + V(X_2)$

Die Standardabweichung (= Risikomass) von $K_1 + K_2$ ist damit:

$$\sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{80^2 + 240^2} = \sqrt{64000} = 80\sqrt{10}$$

und in Prozenten: $\frac{80\sqrt{10} \cdot 100}{3000} = \frac{8}{3}\sqrt{10} = 8.4327\%$

Wir sehen: Prozentual liegt der Erwartungswert des Gewinns in % zwischen den beiden Anlagen für sich betrachtet ($6\% < 7.3333\% < 8\%$). Dies gilt auch für die Standardabweichung ($8\% < 8.4327\% < 12\%$). Das Risiko für die Gesamtinvestition liegt zwischen den Risiken der gemischten Teilinvestitionen, was allerdings auch für die Erwartungswerte gilt (Das Ganze unter den Voraussetzungen: Unabhängigkeit der Zufallsvariablen).

8. a) Gemäss Satz 9.2.10 gilt $E(aX) = aE(x)$ (s. Unterkapitel „Erwartungswert“ im Kapitel „Zufallsvariablen“), gilt damit für den Erwartungswert in Euro: $\frac{1}{1.5} \cdot 60 = 40$
Für die Varianz gilt gemäss Satz 9.5.21 $V(bX) = b^2V(X)$ (s. Unterkapitel „Statistiken“ im Kapitel „Zufallsvariablen“): $V(\frac{1}{1.5}X) = (\frac{1}{1.5})^2 \cdot 80 = 35.556$

b) In Prozenten Schweizerfranken: Der Erwartungswert des Gewinns der Anlage ist 6% von der Kapitalanlage. Die Varianz beträgt 8%. Die Standardabweichung ist

$$\frac{\sqrt{80} \cdot 100}{1000} = 0.89443\% \text{ der Kapitalanlage.}$$

In Prozenten Euro: Die Anlage hat in Euro einen Wert von $\frac{1}{1.5} \cdot 1000 = 666.67$. Der Erwartungswert des Gewinns der Anlage ist $\frac{40 \cdot 100}{666.67} = 6\%$ vom Wert der Kapitalanlage. Die Standardabweichung beträgt $\sqrt{35.556} = 5.9629$ Euro und in %: $\frac{5.9629 \cdot 100}{666.67} = 0.89443\%$ der Anlage (man sieht: prozentual verändern sich der Erwartungswert und die Standardabweichung in Prozenten nicht, in welcher Währung wir diese auch berechnen).

9. a) Zur Erinnerung: Korrelation $R = \frac{KOV(X,Y)}{S_1 S_2}$ mit den Standardabweichungen S_1 und S_2 . Gemäss Satz 9.5.34 (s. Unterkapitel „Statistiken“ im Kapitel „Zufallsvariablen“) gilt: $V(X_1 + X_2) = V(X_1) + V(X_2) + 2KOV(X_1, X_2)$

Aus $-0.6 = \frac{KOV(X,Y)}{100 \cdot 240}$ können wir $KOV(X, Y)$ berechnen:

$$KOV(X, Y) = -0.6 \cdot 100 \cdot 240 = -14400$$

und damit

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2KOV(X_1, X_2) = 100^2 + 240^2 - 2 \cdot 14400 = 38800$$

Damit beträgt die Standardabweichung des Gewinns der Gesamtanlage: $\sqrt{38800} = 196.98$

$$(= \frac{196.98 \cdot 100}{3000} = 6.566\%).$$

$$b) V(X_1 + X_2) = V(X_1) + V(X_2) = 100^2 + 240^2 = 67600$$

Damit beträgt die Standardabweichung des Gewinns der Gesamtanlage: $\sqrt{67600} = 260$
 $(= \frac{260 \cdot 100}{3000} = 8.6667\%)$

Das Beispiel zeigt: durch die Investition in Anlagen mit negativ korrelierten Gewinnen kann man das Risiko vermindern, ohne dass sich der Erwartungswert verändert.

$$c) 1 = \frac{KOV(X,Y)}{100 \cdot 240} \implies KOV(X,Y) = 100 \cdot 240 = 24000$$

und damit

$$V(X_1 + X_2) = V(X_1) + V(X_2) + 2KOV(X_1, X_2) =$$

$$100^2 + 240^2 + 2 \cdot 24000 = 115600$$

Standardabweichung: $\sqrt{115600} = 340.0$

In % von 3000: $\frac{100 \cdot 340}{3000} = 11.333\%$. Damit steigt das Risiko, wenn die Anlagen positiv korrelieren.

$$10. \text{ Mit } =\text{norminv}(0.025; 60; 10) = 40.40036015 = \text{VaR}_{0.975}$$

Mit einer Wahrscheinlichkeit von 0.975 hat die Aktie am Ende der betrachteten Periode mindestens einen Wert von 40.40036015.

$$11. \text{ Mit } S_0 = 60; k = 62; \sigma = 0.2; r = 0.05; T = 0.5 \text{ und}$$

$$\frac{\ln\left(\frac{S_0}{k}\right) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} = \frac{\ln\left(\frac{60}{62}\right) + \left(0.05 + \frac{1}{2} \cdot 0.2^2\right)0.5}{0.2 \cdot \sqrt{0.5}} = 0.015628$$

$$\frac{\ln\left(\frac{S_0}{k}\right) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} = \frac{\ln\left(\frac{60}{62}\right) + \left(0.05 - \frac{1}{2} \cdot 0.2^2\right)0.5}{0.2 \cdot \sqrt{0.5}} = -0.12579$$

$$\Phi(0.015628) = 0.506234416$$

$$\Phi(-0.12579) = 0.449949079$$

$$V_0 = S_0 \Phi\left(\frac{\ln\left(\frac{S_0}{k}\right) + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) - k e^{-rT} \Phi\left(\frac{\ln\left(\frac{S_0}{k}\right) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right)$$

$$= 60 \cdot 0.506234416 - 62 \cdot e^{-0.05 \cdot 0.5} \cdot 0.449949079$$

$$= 3.1660$$

14.6 Lernziele

- Die Wahrscheinlichkeiten bezüglich Normalverteilungen
 - mit Excel berechnen können
 - Die Funktionswerte der Dichtefunktion ausrechnen können (die Formel muss man nicht auswendig kennen).
- Die Idee des Näherungsverfahrens für die Berechnung von Flächen unter der Dichtefunktion beschreiben können.
- Wissen, dass die Dichtefunktion der Normalverteilungen eine bezüglich der y-Achse symmetrische Kurve ist und dass deshalb der Erwartungswert und der Median zusammenfallen.
- Die Abkürzung $X \sim N(\mu, \sigma^2)$ verstehen.
- Die Abkürzung Φ verstehen (Abkürzung für die Verteilungsfunktion der Standardnormalverteilung).

- Wahrscheinlichkeiten $P(X \leq x)$ und $P(X > x)$ für $X \sim B(n, p)$ mit Hilfe der Normalverteilung näherungsweise berechnen können.
- Wissen, dass ein enger Zusammenhang zwischen der Familie der Binomialverteilungen und der Standardnormalverteilung besteht.
- Normalverteilungen beim Vorliegen entsprechender Daten schätzen können.
- Übungen der obigen Art machen können.

Kapitel 15

P-P-Plots und Q-Q-Plots

Bisher haben wir eine Möglichkeit gesehen, um graphisch zu überprüfen, ob Daten als Werte Zufallsvariablen betrachtet werden können, die auf eine spezifische Art verteilt sind. Wir zeichneten die empirische und die theoretische Verteilungsfunktion ins gleiche Koordinatennetz und überprüften die Anpassung von Auge. Wir betrachten nun eine weitere, häufig verwendete Methode. Um zu überprüfen, ob n Daten eine bestimmte Verteilung haben, können wir auch sogenannte P-P-Plots (Probability-Probability-Plots) oder Q-Q-Plots (Quantil-Quantil-Plots) zeichnen.

15.1 P-P-Plots

Wir tragen die folgenden Punkte in ein Koordinatennetz ein:

$$(h(x_i), F(x_i))$$

$h(x_i)$ sind die Werte der empirischen Verteilungsfunktion der x_i ($h(x_i)$ ist der Anteil der Daten, die kleiner gleich x_i sind). Damit gilt für das i -te Datum x_i eines geordneten Datensatzes:

$$h(x_i) = \frac{i}{n}.$$

$F(x_i)$ sind die Werte der theoretischen Verteilungsfunktion, d.h.

$$F(x_i) = P(X \leq x_i).$$

Es gilt: Die Punkte $(h(x_i), F(x_i))$ befinden sich im ersten Quadranten, und zwar im Quadrat $[0, 1] \times [0, 1]$.

Passt die theoretische Verteilung zur Verteilung der Daten, so dürfen die Punkte $(h(x_i), F(x_i))$ nicht zu stark um die $y = x$ -Achse streuen. Insbesondere darf keine systematische Abweichungen vorliegen (z.B. Schlangenform). P-P-Plots sind für beliebige Verteilungen erstellbar, gebräuchlich sind sie vor allem für die Normalverteilung.

Beispiel 15.1.1. *Wir wollen überprüfen, ob die folgenden Daten normalverteilt sind (Schraubenlängen, s. Tabelle 15.1.1):*

Schraubenlängen
0.04992132
0.07626878
0.09694301
0.24706014
0.33711828
0.48861922
0.51154522
0.5559821
0.58895487
0.9406725

Tabelle 15.1.1: Auf Normalverteilung zu prüfende Daten - Schraubenlängen (bereits geordnet)

Wir berechnen die Wahrscheinlichkeiten $F(x_i) = P(X \leq x_i)$ mit $X \sim N(\mu, \sigma^2)$, wobei wir den Erwartungswert und Standardabweichung aus den Daten schätzen, indem wir diese mit dem empirischen Mittelwert und der empirischen Standardabweichung identifizieren. Wir erhalten den aus den Daten geschätzten Erwartungswert $\widehat{E(X)} = \hat{\mu} = 0.389308544 = \bar{x}$ und die geschätzte Varianz $\widehat{V(X)} = \hat{\sigma}^2 = 0.282461202^2 = \text{empirische Varianz}$; $n = 10$. Somit ist $h(x_i) = \frac{i}{10}$. Wir erhalten die folgenden Punkte (s. Tabelle 15.1.2):

Geordnete Daten	$h(x_i) = \frac{i}{n}$	$P(X \leq x_i)$
0.04992132	0.1	0.114771729
0.07626878	0.2	0.133875298
0.09694301	0.3	0.150319419
0.24706014	0.4	0.307270074
0.33711828	0.5	0.426704848
0.48861922	0.6	0.637427306
0.51154522	0.7	0.667403845
0.5559821	0.8	0.722430127
0.58895487	0.9	0.760157636
0.9406725	1	0.974530832

Tabelle 15.1.2: Berechnung von $h(x_i)$ und $F(x_i)$ für die Zeichnung des P-P-Plots; Vermutung der Normalverteilung; Aus den Daten geschätzter Erwartungswert $\widehat{E(X)} = \hat{\mu} = 0.389308544 = \bar{x}$ und Varianz $\widehat{V(X)} = \hat{\sigma}^2 = 0.282461202^2 = \text{empirische Varianz}$; $n = 10$; $h(x_i) = \frac{i}{10}$.

Wir tragen die entsprechenden Punkte in ein Koordinatensystem ein (samt der Achse $y = x$ und wir erhalten; s. Abbildung 15.1.1):

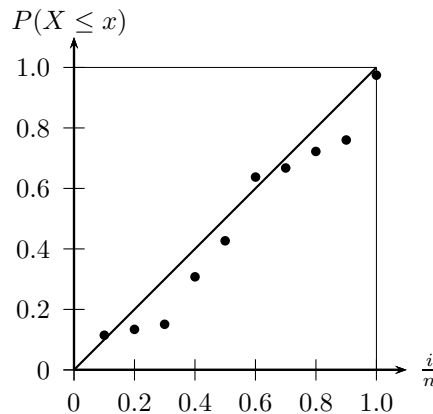


Abbildung 15.1.1: P-P-Plot der Daten von Tabelle 15.1.1 bei Vermutung von Normalverteilung mit geschätzten Parametern $\widehat{E(X)} = \hat{\mu} = 0.389308544 = \bar{x}$ und $\widehat{V(X)} = \hat{\sigma}^2 = 0.282461202^2$

Die Daten sind vermutlich nicht normalverteilt - es handelt sich in der Tat um uniform verteilte Zufallszahlen. Allerdings muss beachtet werden, dass zu einer relativ zuverlässigen Beurteilung zehn Daten wenig sind. Der obige, kleine Datensatz ist als Beispiel gedacht, um die Berechnung mit manuellen Mitteln nachvollziehbar zu machen. $\diamond$

Bemerkung 15.1.2. Beachten Sie, dass empirische Verteilungsfunktionen und P-P-Plots gewöhnlich zur Lösung recht unterschiedlicher Aufgabenstellungen verwendet werden. Die empirische Verteilungsfunktion dient dazu, eine Verteilungsfamilie zu bestimmen, zu der die Verteilung der Daten gehören könnte. Auf Grund der Form der empirischen Verteilungsfunktion wählen wir aus bekannten Verteilungsfamilien eine passende aus. Die P-P-Plots hingegen brauchen wir um zu prüfen, ob die Daten zu einer spezifischen Verteilungsfamilie passen. Wir wollen z.B. überprüfen, ob die Daten exponentialverteilt sind oder nicht. Solche Überprüfungen verwenden wir später oft, da für statistische Tests häufig spezifische Verteilungen von Daten verlangt sind. Führt der P-P-Plot zur Verwerfung der Aussage, dass die Daten zu einer spezifischen Verteilung passen, so wissen wir mit Hilfe des P-P-Plots nicht, zu welcher Verteilungsfamilie die Daten passen könnten. Erst eine empirische Verteilungsfunktion könnte hier weiterhelfen. Muss man also auf Grund von Daten eine theoretische Verteilung schätzen, so darf man nicht mit P-P-Plots beginnen - dies würde auf reines Raten hinauslaufen. Man muss zuerst eine empirische Verteilungsfunktion zeichnen, dann eine Verteilungsfamilie bestimmen. Anschliessend schätzt man die konkrete theoretische Verteilung, indem man die Parameter schätzt. Zuletzt kann man mit einem P-P-Plot überprüfen, ob die theoretische Verteilung zu den Daten passt oder nicht. $\diamond$

15.2 Trendbereinigte P-P-Plots (fakultativ)

Die trendbereinigten P-P-Plots zeigen die Differenzen zwischen den beobachteten und den erwarteten Werten. Wenn die Stichprobe aus der vermuteten Verteilung stammt, sollten die Punkte relativ nahe in einem horizontalen Streifen um 0 liegen. Es sollte kein Muster zu erkennen sein. Wir tragen die Punkte

$$\left(\frac{i}{n}, h(x_i) - F(x_i)\right)$$

in einer graphischen Darstellung ab. Für die Daten von Tabelle 15.1.1 erhalten wir die folgenden Punkte $\left(\frac{i}{n}, h(x_i) - F(x_i)\right)$ (s. Tabelle 15.2.3):

$\frac{i}{n}$	$F(x_i)$	$\frac{i}{n} - F(x_i)$
0.1	0.114771729	-0.014771729
0.2	0.133875298	0.066124702
0.3	0.150319419	0.149680581
0.4	0.307270074	0.092729926
0.5	0.426704848	0.073295152
0.6	0.637427306	-0.037427306
0.7	0.667403845	0.032596155
0.8	0.722430127	0.077569873
0.9	0.760157636	0.139842364
1	0.974530832	0.025469168

Tabelle 15.2.3: Zahlenpaare für die graphische Darstellung des trendbereinigten P-P-Plots der Daten von Tabelle 15.1.1

Wir erhalten die folgende graphische Darstellung (s. Abbildung 15.2.2):

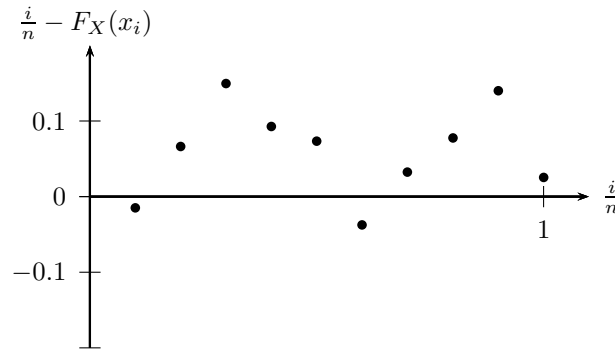


Abbildung 15.2.2: Trendbereinigter P-P-Plot der Daten von Tabelle 15.1.1 bei Vermutung von Normalverteilung mit den geschätzten Parametern $\widehat{E}(X) = \hat{\mu} = 0.389308544 = \bar{x}$ und $\widehat{V}(X) = \hat{\sigma}^2 = 0.282461202^2$

Wir werden in der Folge trendbereinigte P-P-Plots kaum verwenden. Es besteht eine gewisse Gefahr, bei diesen Plots die Abweichung der Daten von der theoretischen Verteilung zu überschätzen.

15.3 Q-Q-Plots (fakultativer Stoff)

Bei Q-Q-Plots tragen wir, wie es der Name sagt, Quantile ab: empirische und theoretische Quantile. Die empirischen Quantile sind die Daten selber. Wir tragen auf der x -Achse die empirischen Daten ab. Auf der y -Achse möchten wir die theoretischen Quantile der empirischen Verteilungsfunktion $h(x_i) = \frac{i}{n}$ abtragen, wobei sich allerdings das Problem ergibt, dass für viele Verteilungen diese für $\frac{n}{n} = 1$ nicht definiert sind (z.B. für die Normalverteilung, Exponentialverteilung. Deren Verteilungsfunktionen haben ja 1 als Asymptote und werden damit nie 1). Deshalb berechnen wir die Quantile für $h(x_i) - \frac{1}{2n}$. Wir müssten also x in $F(x) = \frac{i}{n} - \frac{1}{2n}$, bestimmen, d.h. wir berechnen die Werte

$$F^{-1}\left(\frac{i}{n} - \frac{1}{2n}\right).$$

Wir tragen dann die Punkte

$$\left(x_i, F^{-1}\left(\frac{i}{n} - \frac{1}{2n}\right)\right)$$

in einem Koordinatennetz ab. Entspricht die Verteilung der Daten der theoretischen Verteilung, sollten die Punkte ungefähr auf die $y = x$ -Achse zu liegen kommen. Mit Excel werden die Quantile der Normalverteilung durch "NORMINV($\frac{i}{n} - \frac{1}{2n}; \mu; \sigma$)" berechnet.

Am Beispiel der Daten von Tabelle 15.1.1 erhalten wir die folgenden Zahlenpaare (s. Tabelle 15.3.4):

x_i	$F^{-1}(\frac{i}{n} - \frac{1}{2n})$
0.04992132	-0.075298789
0.07626878	0.096556323
0.09694301	0.198791358
0.24706014	0.280470462
0.33711828	0.353814089
0.48861922	0.424802999
0.51154522	0.498146626
0.5559821	0.57982573
0.58895487	0.682060765
0.9406725	0.853915877

Tabelle 15.3.4: Zahlenpaare für das Zeichnen eines Q-Q-Plots für die Daten der Tabelle 15.1.1

Dies ergibt die graphische Darstellung (s. Abbildung 15.3.3)

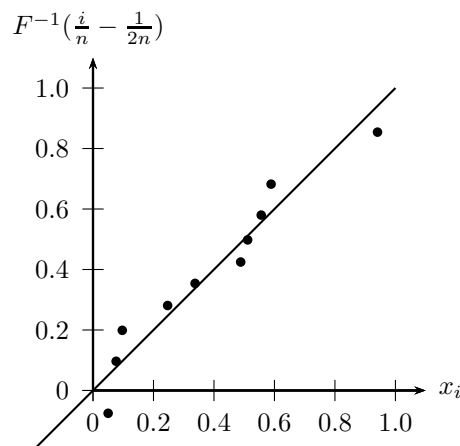


Abbildung 15.3.3: Q-Q-Plot der Daten von Tabelle 15.1.1 bei Vermutung von Normalverteilung mit geschätzten Parametern $\widehat{E}(X) = \hat{\mu} = 0.389308544 = \bar{x}$ und $\widehat{V}(X) = \hat{\sigma}^2 = 0.282461202^2$

Beachten Sie, dass im Gegensatz zu den P-P-Plots beim Q-Q-Plot die Punkte nicht im Quadrat $[0, 1] \times [0, 1]$ zu liegen brauchen, da beim Q-Q-Plot Daten abgetragen werden. Deshalb ist das obige Beispiel eher untypisch!

15.4 Trendbereinigte Q-Q-Plots (fakultativer Stoff)

Es werden die Punkte

$$(x_i, x_i - F^{-1}(\frac{i}{n} - \frac{1}{2n}))$$

abgetragen. Passt die Verteilung der Daten zur vermuteten theoretischen Verteilung, so sollten diese Punkte nicht zu stark von der x -Achse abweichen. Für die Daten von Tabelle 15.1.1 erhalten wir die folgenden Zahlenpaare (s. Tabelle 15.4.5):

Dies ergibt die graphische Darstellung (s. Abbildung 15.4.4):

x_i	$F^{-1}(\frac{i}{n} - \frac{1}{2n})$	$x_i - F^{-1}(\frac{i}{n} - \frac{1}{2n})$
0.04992132	-0.075298789	0.125220109
0.07626878	0.096556323	-0.020287543
0.09694301	0.198791358	-0.101848348
0.24706014	0.280470462	-0.033410322
0.33711828	0.353814089	-0.016695809
0.48861922	0.424802999	0.063816221
0.51154522	0.498146626	0.013398594
0.5559821	0.57982573	-0.02384363
0.58895487	0.682060765	-0.093105895
0.9406725	0.853915877	0.086756623

Tabelle 15.4.5: Zahlenpaare für das Zeichnen eines trendbereinigten Q-Q-Plots für die Daten der Tabelle 15.1.1

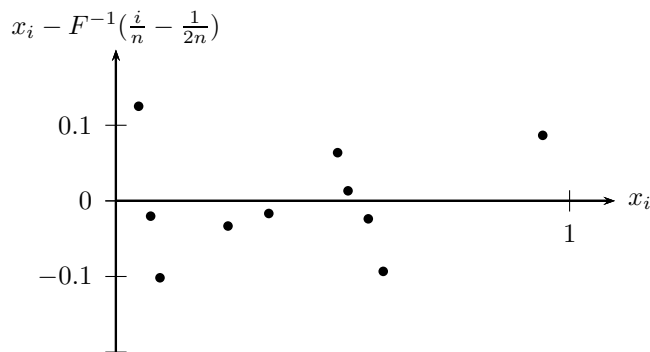


Abbildung 15.4.4: Trendbereinigter Q-Q-Plot der Daten von Tabelle 15.1.1 bei Vermutung von Normalverteilung mit geschätzten Parametern $\widehat{E(X)} = \hat{\mu} = 0.389308544 = \bar{x}$ und $\widehat{V(X)} = \hat{\sigma}^2 = 0.282461202^2$

Auch hier gilt, dass wir die trendbereinigten Q-Q-Diagramme i.A. nicht verwenden, da sie leicht zu einer Überschätzung der Abweichung der Daten von der theoretischen Verteilung führen.

15.4.1 Übungen

Die in den Übungen verwendeten Daten finden Sie unter <http://math.logik.ch> (ppplots.xls; Mit Stern versehene Aufgaben sind fakultativ).

1. Erstellen Sie von Hand einen P-P-Plot für folgende Daten, für die man eine Normalverteilung vermutet (Gewinne verschiedener Anlagen in einem Jahr auf 100 Dollar): 12.1, 13.8, 35.1, 19.3, 17.8, 17.9, 7, 13.3, 14.2, 20.9;
2. Untersuchen Sie mit Hilfe eines P-P-Plots (Excel und einer Statistik-Software), ob der Datensatz der Übung normal verteilt ist (Verkauf pro Kunde in einer Filiale pro Tag).
3. Untersuchen Sie mit Hilfe eines P-P-Plots (Excel und einer Statistik-Software), ob die Daten der Übung exponentialverteilt sind (Schraubenlängen).
4. Untersuchen Sie mit Hilfe eines P-P-Plots (Excel und einer Statistik-Software), ob die Daten der Übung uniform verteilt sind (Gewinne pro Quadratmeter in Filialen).
5. *Erstellen Sie mit Excel vier uniform verteilte Serien von Zufallszahlen (50 Zufallszahlen), so dass gilt: $X \sim U(0, 5)$ Stellen Sie den entsprechenden P-P-Plot (Q-Q-Plot) dar (Excel liefert mit dem Befehl „zufallszahl()“ uniform verteilte Zufallszahlen im Intervall $[0, 1]$).

6. *Erstellen Sie mit Excel uniform verteilte Zufallszahlen (50), so dass $X \sim U(3, 10)$. Wir nehmen an, Sie wüssten nicht, dass es sich um uniform verteilte Zufallszahlen handelt. Sie vermuten eine Normalverteilung. Erstellen Sie den P-P-Plot (Q-Q-Plot).
7. Erstellen Sie mit Excel einen Plot der empirischen Verteilungsfunktion der Zufallszahlen der Übung (Zeit bis zum Ausfallen von Maschinen, siehe Exceltabelle). Zu welcher Vermutung verleitet Sie der Plot. Überprüfen Sie Ihre Vermutung mit einem P-P-Plot.

15.4.2 Lösungen

1. Wir erhalten mit den aus den Daten durch den Mittelwert 17.14 und die Standardabweichung 7.491802928 geschätzten Parametern die folgende graphische Darstellung (s. Abbildung 15.4.5):

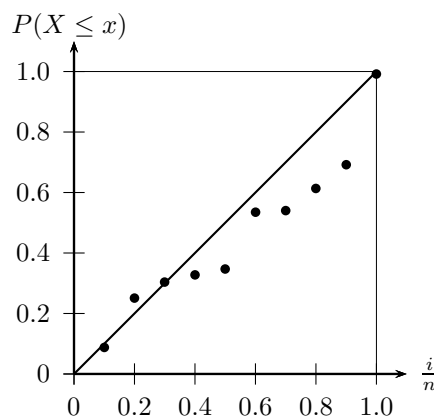


Abbildung 15.4.5: P-P-Plot der Daten von Tabelle 15.1.1 bei Vermutung von Normalverteilung mit geschätzten Parametern $\widehat{E}(X) = \hat{\mu} = 17.14 = \bar{x}$ und $\widehat{V}(X) = \hat{\sigma}^2 = 7.491802928^2$

Die Daten sind vermutlich nicht normalverteilt. Auch hier gilt, dass wir zuwenig Daten haben. Die folgenden Übungen weisen grössere Datensätze auf und müssen entsprechend mit Computer gelöst werden.

2. Wir erhalten mit den mittels des Mittelwerts 12.64000311 und der Standardabweichung 4.441030339 geschätzten Parameter(s. Abbildung 15.4.6):

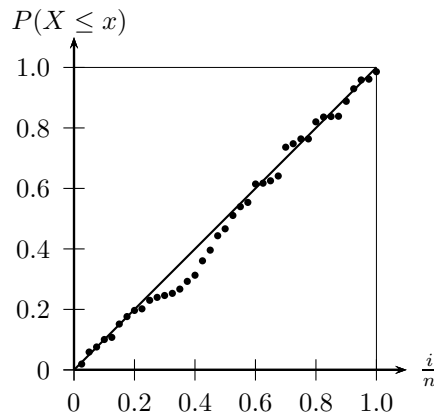


Abbildung 15.4.6: P-P-Plot der Daten von Tabelle 15.1.1 bei Vermutung von Normalverteilung mit geschätzten Parametern $\widehat{E(X)} = \hat{\mu} = 12.64000311 = \bar{x}$ und $\widehat{V(X)} = \hat{\sigma}^2 = 4.441030339^2$

Die Daten sind nach den P-P-Plots wohl normalverteilt. Es hat zwar einen Bereich mit einer Abweichung. Diese gilt jedoch nicht für den gesamten Bereich und ist nicht systematisch. Wir würden die Normalverteilungsannahme gelten lassen (die Daten entstammen einem Satz normalverteilter Zufallszahlen).

3. Wir erhalten mit dem durch den Mittelwert 1.424940549 der Daten geschätzten Parameter $\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{1.424940549} = 0.701783664$ (s. Abbildung 15.4.7):

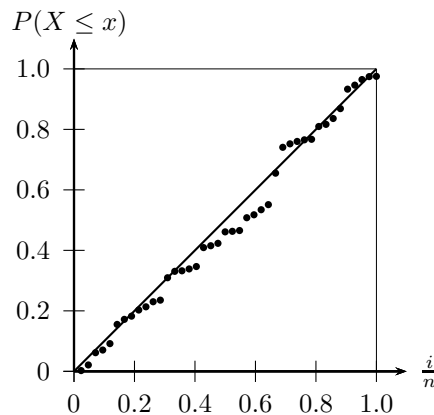


Abbildung 15.4.7: P-P-Plot bei Vermutung von Exponentialverteilung mit geschätzten Parametern $\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{1.424940549} = 0.701783664$

Die Daten sind offenbar exponentialverteilt.

4. Die stetige uniforme Verteilung ist durch zwei Parameter a und b bestimmt ($U(a, b)$). Wir schätzen diese durch das kleinste und das grösste Datum 10.09 und 17.97 und erhalten den folgenden P-P-Plot (s. Abbildung 15.4.8):

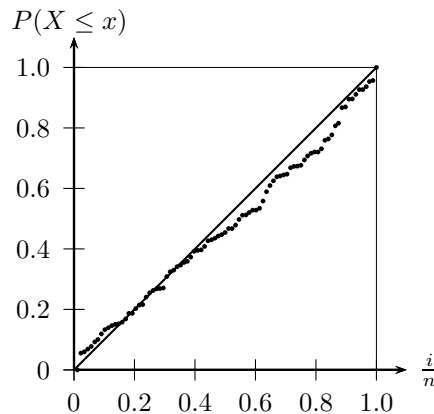


Abbildung 15.4.8: P-P-Plot bei Vermutung von uniformer Verteilung mit geschätzten Parametern $\hat{a} = \min(x) = 10.05$ und $\hat{b} = \max(x) = 17.97$

Die Daten sind offenbar uniform verteilt.

5. Bei der Übung geht es darum, ein Gefühl dafür zu bekommen, wie Zufallszahlen, die zu einer bestimmten Verteilung gehören, um die entsprechende theoretische Verteilung streuen. Die folgenden P-P-Plots sind das Resultat von Zufallszahlen. Entsprechend werden Sie nicht dasselbe Resultat erhalten (s. Abbildungen [15.4.9](#) - [15.4.12](#)).

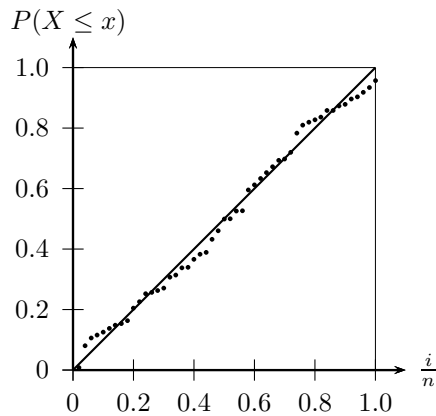


Abbildung 15.4.9: P-P-Plot für Zufallszahlen, die laut $U(0,5)$ verteilt sind (P-P-Plot mit Verteilungsfunktion von $U(0,5)$ berechnet)

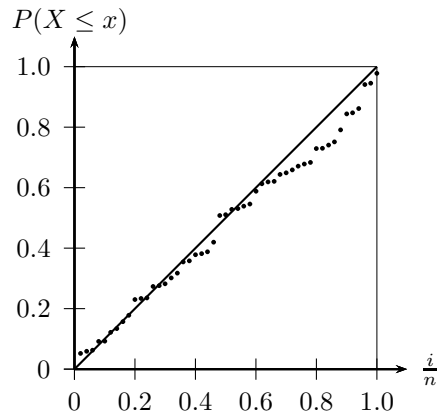


Abbildung 15.4.10: P-P-Plot für Zufallszahlen, die laut $U(0,5)$ verteilt sind (P-P-Plot mit Verteilungsfunktion von $U(0,5)$ berechnet)

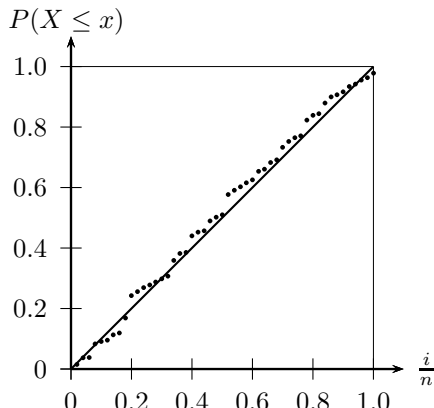


Abbildung 15.4.11: P-P-Plot für Zufallszahlen, die laut $U(0, 5)$ verteilt sind (P-P-Plot mit Verteilungsfunktion von $U(0, 5)$ berechnet)

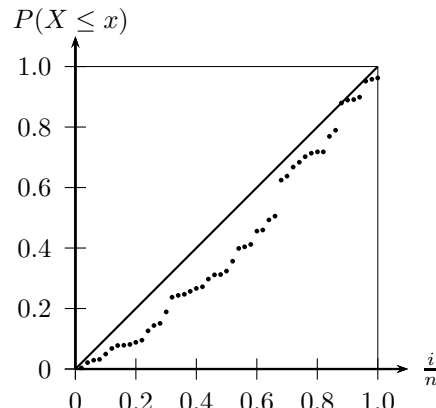


Abbildung 15.4.12: P-P-Plot für Zufallszahlen, die laut $U(0, 5)$ verteilt sind (P-P-Plot mit Verteilungsfunktion von $U(0, 5)$ berechnet)

Während die ersten drei Sätze als uniformverteilt gelten könnte, würden wir das beim vierten Satz ablehnen, obwohl er genau gleich produziert wurde. Zufälliger weise können Daten von der Verteilung, aus der sie stammen, abweichen. Je mehr Daten wir haben, desto unwahrscheinlicher ist dies jedoch (laut dem Satz 13.0.1).

6. Je nach den erhaltenen Zufallszahlen ergibt sich ein anderer P-P-Plot. Es ist durchaus möglich, dass Sie einen Plot erhalten, der mit einer Normalverteilung verträglich wäre. Typisch für uniform verteilte Daten, die mit P-P-Plots auf Normalverteilung überprüft werden, ist die S-förmige Struktur um die $y=x$ -Achse. (s. Abbildungen 15.4.13 - 15.4.16).

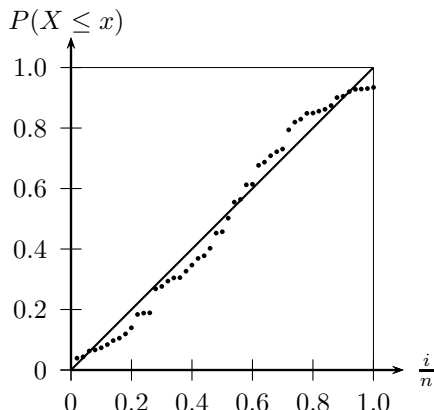


Abbildung 15.4.13: P-P-Plot für Zufallszahlen, die laut $U(3, 10)$ verteilt sind (P-P-Plot mit Verteilungsfunktion der geschätzten Normalverteilung berechnet)

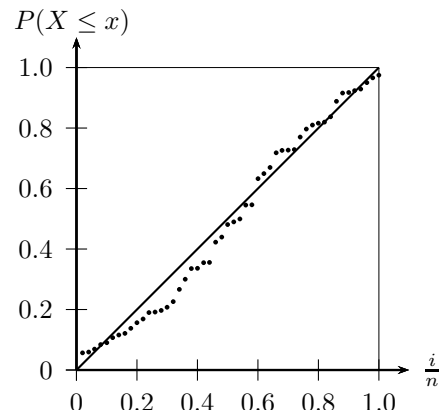


Abbildung 15.4.14: P-P-Plot für Zufallszahlen, die laut $U(3, 10)$ verteilt sind (P-P-Plot mit Verteilungsfunktion der geschätzten Normalverteilung berechnet)

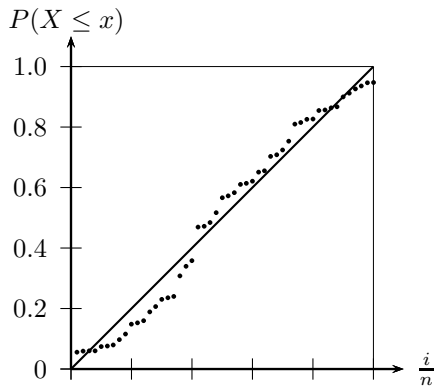


Abbildung 15.4.15: P-P-Plot für Zufallszahlen, die laut $U(3,10)$ verteilt sind (P-P-Plot mit Verteilungsfunktion der geschätzten Normalverteilung berechnet)

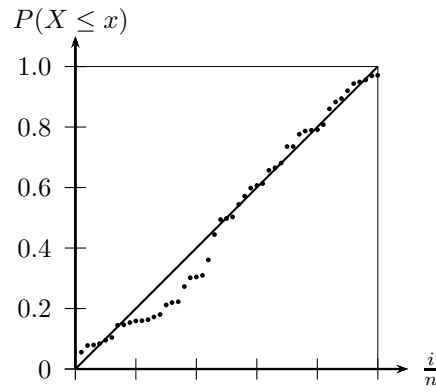


Abbildung 15.4.16: P-P-Plot für Zufallszahlen, die laut $U(3,10)$ verteilt sind (P-P-Plot mit Verteilungsfunktion der geschätzten Normalverteilung berechnet)

Man kann feststellen, dass bis auf den letzten Plot schwache, aber systematische S-förmige Abweichungen vorliegen. Bei der Beurteilung der Plots kommt es entsprechend nicht nur auf die Abweichung, sondern vor allem darauf an, ob Abweichungen systematisch sind.

7. Wir zeichnen zuerst eine empirische Verteilungsfunktion (s. Abbildungen 15.4.17). Diese führt zur Vermutung, dass die Daten gemäss einer Verteilung aus der Familie der Exponentialverteilungen verteilt sind. Wir überprüfen das mit einem P-P-Plot (s. Abbildung 15.4.18).

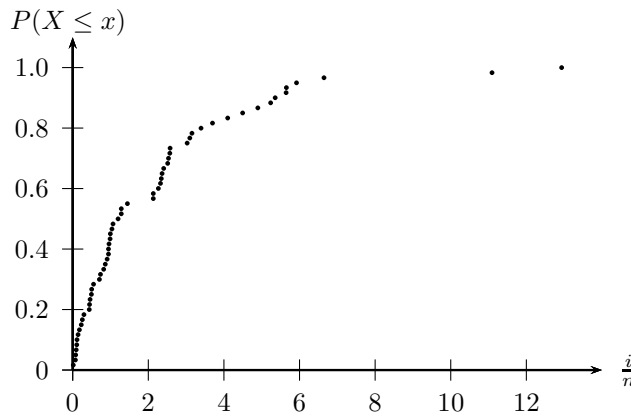


Abbildung 15.4.17: Empirische Verteilungsfunktion der Daten

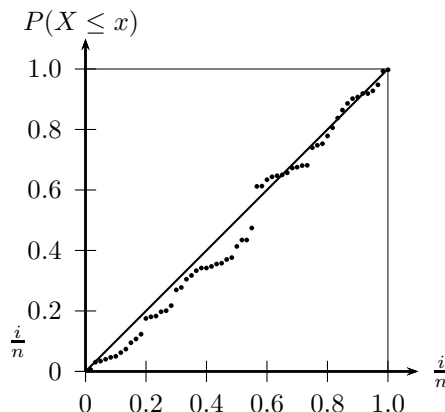


Abbildung 15.4.18: P-P-Plot mit $\text{Exp}(\hat{\lambda}) = \text{Exp}(0.443794147)$

Es scheint keine systematische Abweichung zu geben. Die Daten können als exponentialverteilt betrachtet werden.

Bemerkung 15.4.1. *Ergeben sich systematische Abweichungen (z.B. Schlangenform), so passen die Daten nicht zur entsprechenden theoretischen Verteilung.* $\diamond$

15.5 Lernziele

- P-P-Plots von Hand, mit Excel und mit einer Statistik-Software zeichnen können.

- Erläutern können, wieso man mit P-P-Plots überprüfen kann, ob Daten eine spezifische Verteilung aufweisen.
- Erläutern können, wieso mit steigender Datenzahl und angemessener Verteilungswahl die Punkte des P-P-Plots näher an der y - x -Gerade liegen müssen.
- Übungen der Art 1 und 2 lösen können.

Kapitel 16

Zufallszahlen und stetige Verteilungen (fakultativ)

Definition 16.0.1. x_i ist eine P -verteilte Zufallszahl genau dann, wenn x_i Wert einer P -verteilten Zufallsvariable X_i ist.

Excel liefert Zufallszahlen z_i uniform verteilter Zufallsvariable $Z_i \sim U(0, 1)$ (Befehl: Zufallszahl()). Es ist manchmal nützlich, anders verteilte Zufallszahlen zur Verfügung zu haben. Um Zufallszahlen x_i , die Werte von $X_i \sim U(a, b)$ sind, zu erhalten, berechnen wir $x_i = z_i(b - a) + a$. (Begründung: bei $x_i = z_i(b - a)$ wird 0 zu 0 und 1 zu $b - a$. Damit erhalten wir uniform verteilte Daten zwischen 0 und $b - a$. Nun verschieben wir diese noch um a nach links oder nach rechts und erhalten die Grenzen $0 + a = a$ und $b - a + a = b$). Für nicht uniform verteilte Verteilungen gibt es ein ebenso einfaches Verfahren, sofern:

1. die Verteilungsfunktion F von P in geschlossener Form existiert und
2. und F im Intervall, wo $0 < F(x) < 1$ gilt, strikt monoton steigend ist. Dann existiert auf $]0, 1[$ die Umkehrfunktion $F'^{-1} :]0, 1[\rightarrow \mathbb{R}$. (F' ist die Funktion, die wir aus F erhalten, indem wir den Definitionsbereich auf die Menge D' der $x \in \mathbb{R}$ einschränken, so dass $0 < F(x) < 1$, d.h. $D' = \{x : 0 < F(x) < 1\}$ und $F' : D' \rightarrow]0, 1[$).

Bei diesen Bedingungen erlaubt der folgende Satz eine Produktion von F -verteilten Daten:

Satz 16.0.2. Sei Z eine uniform verteilte Zufallsvariable mit $Z \sim U(0, 1)$ sowie der Verteilungsfunktion F_U und sei $F' : \{x : 0 < F(x) < 1\} \rightarrow]0, 1[$ strikt monoton steigend sowie $F' = F$ für $\{x : 0 < F(x) < 1\}$ und F die Verteilungsfunktion von P dann gilt: $F'(x) = F_U(z)$. Zudem gilt $x = F'^{-1} \circ F_U(z)$.

Beweis. Sei Z eine $U(0, 1)$ -verteilte Zahlen und F_U die entsprechende Verteilungsfunktion. Zudem sei $F'(x) = P(X \leq x) = z$. Dann gilt:

$F_U(z) = P(Z \leq z) = z$ (s. letzter Beweis im Unterkapitel „uniforme stetige Verteilungen“) und damit $F'(x) = F_U(z)$.

Es gilt $x = F'^{-1} \circ F'(x) = F'^{-1} \circ F_U(z)$. □

Um P -verteilte Zufallszahlen zu produzieren, können wir mit Excel $U(0, 1)$ -verteilte Zufallszahlen z_i produzieren, die Umkehrfunktion der beschränkten Verteilungsfunktion F' und $x_i = F'^{-1}(z_i)$ berechnen.

Beispiel 16.0.3. Im interessierenden Bereich $M = \{x : 0 < F(X) < 1\} =]0, \infty[$ ist die Exponentialverteilung strikt monoton steigend. Deshalb existiert dort die Umkehrfunktion. Die Umkehrfunktion von $F'(z) = 1 - e^{-\lambda z}$ (für $0 < z < 1$) ist:

$$F'^{-1}(z) = \frac{\ln(1 - z)}{-\lambda}$$

Denn: $z = 1 - e^{-\lambda x} \iff z - 1 = -e^{-\lambda x} \iff 1 - z = e^{-\lambda x} \iff$
 $\ln(1 - z) = -\lambda x \ln e = -\lambda x \iff x = \frac{\ln(1-z)}{-\lambda}$.

Für $X \sim \text{Exp}(0.5)$ erhalten wir somit: $F^{-1}(x) = \frac{\ln(1-x)}{-0.5}$ ($F' :]0, \infty[\rightarrow]0, 1[$).

Wir berechnen die exponentialverteilten Zufallszahlen x_i also durch:

$$x_i = \frac{\ln(1 - z_i)}{-0.5}$$

Am Beispiel konkreter, geordneter Zahlen (s. Tabelle 16.0.1):

Uniform verteilte Zufallszahlen (geordnet) z_i	exponential verteilte Zufallszahlen ($\lambda = 0.5$) $\frac{\ln(1-z_i)}{-0.5}$
0.085855916	0.179534158
0.232306021	0.52872818
0.421708556	1.095354617
0.511978586	1.434791988
0.623061082	1.951344252
0.645881623	2.076248045
0.74606836	2.741380365
0.948094309	5.916653692
0.950757762	6.022007067
0.957906159	6.335707699

Tabelle 16.0.1: Mit Computer produzierte $\text{Exp}(0.5)$ -verteilter Daten

Wir veranschaulichen das Vorgehen mit einer Abbildung 16.0.1. Auf der y -Achse sind die $U(0,1)$ -verteilten Daten z_i abgetragen (mit Computer produziert). Auf der x -Achse erhalten wir für $z_i \in]0, 1[$ durch die Berechnung des Wertes der Umkehrfunktion $F^{-1}(z_i) = \frac{\ln(1-z_i)}{-0.5}$ exponenti-
alverteilte Daten x_i .

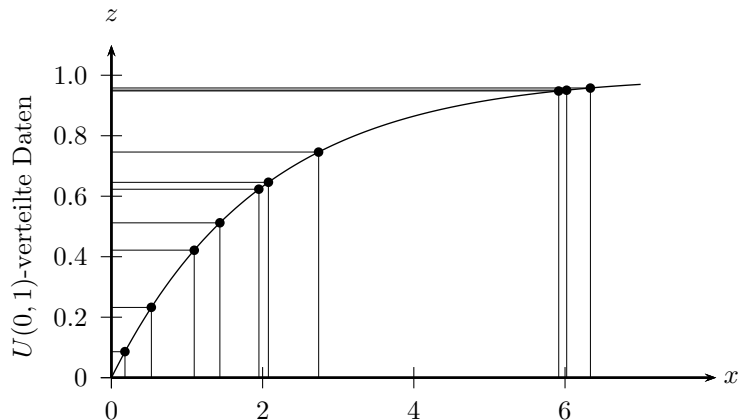


Abbildung 16.0.1: Berechnung von $\text{Exp}(0.5)$ -verteilten Daten mit Hilfe von $U(0,1)$ -verteilten Daten und der Umkehrfunktion $F^{-1}(z) = \frac{\ln(1-z)}{-0.5}$ der Verteilungsfunktion $F(x) = 1 - e^{-0.5x}$; Daten der Tabelle 16.0.1

Man kann nun überprüfen, wie gut die produzierten Daten dem Modell entsprechen, gemäß dem sie produziert wurden, wobei im Beispiel zu beachten ist, dass sehr wenige Daten erzeugt wurden (s. Abbildung 16.0.2).

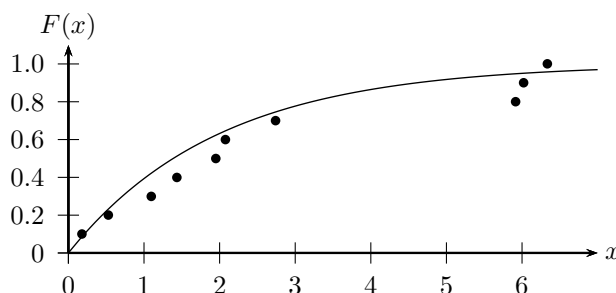


Abbildung 16.0.2: Graphische Darstellung der Verteilungsfunktion F der Exponentialverteilung mit dem Parameter $\lambda = 0.5$ und der empirischen Verteilungsfunktion der in Tabelle 16.0.1 erzeugten Daten

◇

Intuitiv einsichtiger als der gelieferte Beweis des Satzes 16.0.2 mögen als Begründung für das eingeführte Verfahren zur Produktion F -verteilter Daten folgende Überlegungen sein: Bei einer uniform verteilten Zufallsvariable im Intervall $[0, 1]$ auf der y -Achse liegen z.B. ca. 70% der Zufallszahlen unter der $z = 0.7$ Marke. Berechnen wir die Umkehrwerte dieser Zufallszahlen, liegt genau dieselbe Anzahl von Zahlen links des Wertes von $F^{-1}(0.7) := x$. Es sind dies ca. 70%. Die Funktion F ordnet - wie erfordert - nun aber x genau 0.7 zu. Dies gilt für alle z .

Die Verteilungsfunktion der *Standardnormalverteilung* ist zwar in $] -\infty, \infty[$ strikt monoton steigend, jedoch nicht als berechenbare Formel angebar. Entsprechend können wir die Umkehrfunktion nicht berechnen. Wir können jedoch näherungsweise standardnormalverteilte Zufallszahlen produzieren. Ein entsprechendes Verfahren liefern wir später. Wir können aber schon jetzt mit der entsprechenden Berechnungsfunktion von Excel arbeiten. Das Programm offeriert eine Berechnungsfunktion für die Produktion von normalverteilten Zufallsvariablen: "Norminv()". Wir wenden diese Berechnungsfunktion auf die durch Excel gelieferten Zufallszahlen aus $U(0, 1)$ an.

16.0.1 Übungen

1. Berechnen Sie mit Hilfe von Excel vier Sätze von exponentialverteilten Zufallszahlen (je 50) und erstellen Sie einen Plot der empirischen Verteilungsfunktion. Diskutieren Sie die Resultate mit Ihrem Tischnachbarn. ($\lambda = 2$)
2. Berechnen Sie mit Hilfe von Excel einen Satz (60) von F -verteilten Zufallszahlen, wobei F im Intervall, wo $F(x) \neq 0$ und $F(x) \neq 1$ durch $F_X(x) = 0.005x^2$ bestimmt ist. Zeichnen Sie einen Plot der empirischen Verteilungsfunktion und beurteilen Sie die Anpassung an das Modell $F_X(x) = 0.005x^2$.
3. Berechnen Sie einen Satz von 40 standardnormalverteilter Zufallszahlen. Erstellen Sie den Plot der Eckpunkte der empirischen Verteilungsfunktion und beurteilen Sie die Anpassung an das Modell.
4. Als Vorarbeit für das nächste Kapitel: Berechnen Sie
 - (a) 20 Sätze von exponentialverteilten Zufallszahlen (je 50, $\lambda = 1$)
 - (b) Berechnen Sie den empirischen Mittelwert $\bar{x}_i$ für jeden der 20 Sätze
 - (c) Zeichnen Sie einen Plot der empirischen Verteilungsfunktion der $\bar{x}_i$ und kommentieren Sie das Ergebnis.
 - (d) Berechnen Sie $P(\bar{X} \leq 0.7)$. $\bar{X}$ ist eine Zufallsvariable, die $\bar{x}_i$ als Funktionswerte aufweist.
 - (e) Berechnen Sie $P(0.9 \leq \bar{X} \leq 1.1)$.
 - (f) Berechnen Sie für jeden der 20 Sätze das empirische $\bar{\lambda}_i := \frac{1}{\bar{x}_i}$. Sind die $\bar{\lambda}_i$ normalverteilt?
5. Eine Spanplattenfabrik misst während einiger Zeit die Dicke der produzierten Spanplatten und kommt zu folgendem Ergebnis: (siehe Exceldatei: Aufgabe 5 Zufallszahlen).

- a) Finden Sie eine passende Verteilung für die Daten.
 - b) Platten, die dünner als 1.9 cm sind, können dem Auftraggeber nicht verkauft werden. Wie gross ist der Ausschuss bei künftiger Produktion im Durchschnitt?
 - c) Berechnen Sie $P(1.85 \text{ cm} < X \leq 2.15 \text{ cm})$
6. Eine Firma überlegt die Anschaffung von neuen Maschinen, um die Verluste durch ungenaue Verpackung zu verringern. Dabei geht es um die Abfüllung von Kaffee. Die Gläser sollen möglichst genau 500 g Kaffee enthalten. Gläser, die weniger als 495 Gramm enthalten, dürfen nicht verkauft werden. Gläser, die mehr als 500 g enthalten, verbleiben im Verkauf, führen aber zu Verlusten, da nur 500 g verrechnet werden dürfen. Auf Grund eines Experimentes wurden folgende Daten für die Maschine A und folgende für die Maschine B ermittelt (Siehe Excel-Datei Aufgabe 6 Zufallszahlen). Die Maschine A kostet 50'000, die Maschine B 70'000. Ab welcher Menge ist die Maschine B günstiger, wenn das Gramm nicht berechneten Kaffees 0.007 Fr. kostet und ein ausgemustertes Glas Kosten von 0.02 Franken verursacht.
7. Berechnen Sie die mittlere Abweichung von exponentialverteilten Zufallszahlen von ihrem Modell (= Mittelwert der Differenzen zwischen den Werten der theoretischen und der empirischen Verteilungsfunktion) ($\lambda = 0.5$) für 10, 50, 100 und 1000 Daten. Diskutieren Sie das Resultat.

16.0.2 Lösungen

1. Siehe Exceldatei: Lösung Übung 1 Zufallszahlen
2. $F(x) = 0.005x^2$ bestimmt die Funktion im folgenden Intervall: $[0, 14.142]$, da $F(x) = 1 = 0.005x^2$ nach x aufgelöst 14.142 ergibt. Zudem verläuft die Kurve durch 0.
Wenn $F(x) = 0.005x^2$, dann ist $F^{-1}(x) = \sqrt{\frac{x}{0.005}}$. Um aus uniform verteilten Zufallszahlen z_i F -verteilte zu erhalten, berechnen wir: $x_i = \sqrt{\frac{z_i}{0.005}} = 14.142\sqrt{z_i}$
Für eine Lösung mit spezifischen Zufallszahlen siehe Exceldatei: Lösung Übung 2 Zufallszahlen
3. Siehe Exceldatei: Lösung Übung 3 Zufallszahlen
4. Siehe Exceldatei: Lösung Übung 4 Zufallszahlen.
$$P(\bar{X} \leq 0.7) = \Phi\left(\frac{0.7 - 1.001215947}{0.118842402}\right) = 0.005629082$$
$$P(0.8 \leq \bar{X} \leq 1.2) = P(\bar{X} \leq 1.2) - P(\bar{X} \leq 0.8) = 0.952803829 - 0.045215084 = 0.907588745$$
5. Siehe Exceldatei: Lösung Übung 5 Zufallszahlen.
$$P(X \leq 1.9 \text{ cm}) = \Phi\left(\frac{1.9 - 2.0066}{0.118514961}\right) = 0.184202614 \text{ (ca. 18.4\%)}$$
$$P(1.85 \text{ cm} < X < 2.15 \text{ cm}) = \Phi\left(\frac{2.15 - 2.0066}{0.118514961}\right) - \Phi\left(\frac{1.85 - 2.0066}{0.118514961}\right) = 0.886855468 - 0.093192047 = 0.793663421$$
6. Ab einer Menge von 6'259'540 Gläsern. Bei den auszusondernden Gläsern genügt es, „Menge der produzierten Gläser $\cdot P(X \leq 498)$ $\cdot$ die Kosten pro Glas“ zu rechnen.
Bei den Kosten, die durch Überfüllung verursacht werden, begnügen wir uns mit einer Näherung: Wir berücksichtigen nur ganze Gramm. Dazu müssen wir $P(n - 1 \leq X \leq n)$ für alle relevanten n : 501, 502, 503, etc. berechnen.
Diese Zahlen multiplizieren wir mit der Anzahl Gläser, um den Anteil der Gläser zu erhalten, die in die entsprechende Kategorie fallen. Die Anzahl der Gläser pro Kategorie multiplizieren wir dann mit der Anzahl Gramm, die zuviel sind. Dieses Resultat multiplizieren wir mit den Kosten pro Gramm. Das Endresultat wurde durch Versuchen ermittelt (siehe Lösung Übung 6 Zufallszahlen).
7. siehe Excel-Datei: Lösung Übung 7 Zufallszahlen. Für das konkrete Beispiel erhalten wir die folgenden mittleren Abweichungen: für 10: 0.086775863; für 50: 0.041784468; für 100: 0.025988834

0.024725104; für 1000: 0.006701792

Je mehr Daten wir haben, desto kleiner ist die mittlere Abweichung vom Modell (Gesetz der grossen Zahlen). Daraus folgt dann: je mehr Daten wir haben, desto kleiner muss die mittlere Abweichung sein, damit wir die Daten als zu einem Modell passend beurteilen dürfen.

16.1 Lernziele

- Wissen, dass für beliebig verteilte Datensätze mit Hilfe von Zufallszahlen entsprechend verteiltes Datenmaterial "fabriziert" werden kann, sofern die Umkehrfunktion der Verteilungsfunktion existiert. Diese existiert immer, wenn die Verteilungsfunktion im interessierenden Intervall in geschlossener Form mit Hilfe einer Gleichung ausgedrückt werden kann und wenn sie im interessierenden Intervall strikt monoton steigend ist. Für einfache Funktionen die Umkehrfunktion berechnen und entsprechend verteilte Datensätze mit Hilfe von Zufallszahlen produzieren können.
- Uniform-, exponential- und normalverteilte Datensätze mit Hilfe von Zufallszahlen mit Hilfe von Excel herstellen können.

Kapitel 17

Der Zentrale Grenzwertsatz

Erheben wir Daten durch eine Zufallsstichprobe, so können wir - wie bereits erwähnt - die einzelnen Daten x_i als Werte von - auf die Ergebnismenge Ω definierten - Zufallsvariablen X_i betrachten. Auf Grund der Daten können wir Kennzahlen berechnen (Mittelwert, Varianz, Standardabweichung, Median, etc.). Diese Kennzahlen betrachten wir wiederum als Werte von spezifischen Zufallsvariablen, den Statistiken (s. Seite 214). Wir haben diese Zusammenhänge formal konstruiert, indem wir die Statistiken auf $\times_{i=1}^n \text{Bild}(X_i) = \times_{i=1}^n \mathfrak{X}_i$ (= das kartesische Produkt der Bilder der Zufallsvariablen X_i) definierten und die aus der Stichprobe berechneten Kennzahlen als Werte dieser Zufallsvariablen betrachteten (siehe entsprechende Ausführungen im Kapitel „Zufallsvariablen“, die Sie bitte wiederholen möchten).

17.1 Verteilung der Statistik Mittelwert

Kennen wir von Statistiken die Wahrscheinlichkeitsverteilung, ist dies für die Testtheorie, wie wir sehen werden, von besonderer Bedeutung. In diesem Kapitel wenden wir uns der Verteilung der Statistiken „Summe“ und „Mittelwert“ zu. Diese haben wir für spezifische diskrete Verteilungen im Kapitel „Zufallsvariablen“ bereits berechnet. Bevor wir für diskrete wie für stetige Verteilungen einen Satz über die Verteilung dieser Statistiken festhalten, folgt ein Beispiel zur Wiederholung der Grundgedanken. Wir berechnen die Wahrscheinlichkeitsfunktion der Statistik Mittelwert für zwei und drei Würfelspiele mit sechs möglichen Ausgängen (Gleichverteilung). Die Zufallsvariable ordnet den sechs Ausgängen die Zahlen von 1 bis 6 zu. Wir erhalten die folgenden Graphiken (s. Abbildungen 17.1.1 und 17.1.2):

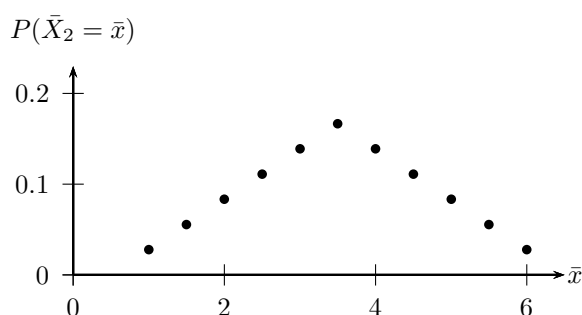


Abbildung 17.1.1: Wahrscheinlichkeitsfunktion der Statistik Mittelwert für zwei diskret uniform verteilte Zufallsvariablen mit sechs möglichen Ausgängen

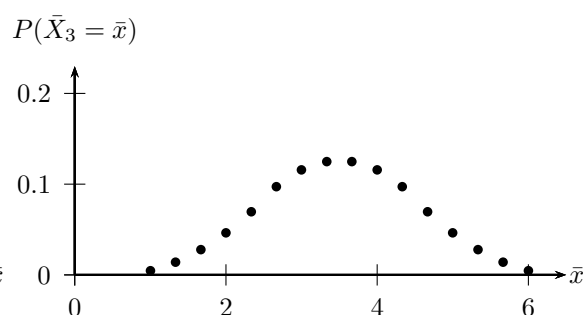


Abbildung 17.1.2: Wahrscheinlichkeitsfunktion der Statistik Mittelwert für drei diskret uniform verteilte Zufallsvariablen mit sechs möglichen Ausgängen

Wir stellen fest, dass die Werte der Wahrscheinlichkeitsfunktion des Mittelwertes dreier iden-

tisch und uniformverteilter, unabhängiger Zufallsvariablen auf einer Kurve liegen, die der Dichtfunktion einer Normalverteilung ähnelt. Dies wollen wir nun noch graphisch präziser überprüfen. Dazu standardisieren wir die Mittelwerte $\bar{x}_i$, woraus Zahlen z_i resultieren. Wir tragen dann

$$(z_i, P(\bar{X}_n \leq \bar{x}_i))$$

in das Koordinatennetz ab und legen die Verteilungsfunktion Φ der Standardnormalverteilung darüber. Wir können die Werte der Statistik Mittelwert wie folgt standardisieren:

$$z_i = \frac{\bar{x}_i - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}}$$

Dabei gilt (s. Sätze 9.5.18 und 9.5.36):

$$E(\bar{X}_n) = E(X_i) = \frac{1}{6} \sum_{k=1}^6 k = 3.5$$

$$\begin{aligned} V(\bar{X}_n) &= \frac{1}{n} V(X_i) = \frac{1}{n} \frac{1}{6} \sum_{k=1}^6 (k - E(X_i))^2 = \\ &= \frac{1}{n} \frac{1}{6} \cdot 17.5 \end{aligned}$$

Mit diesen Werten wird dann berechnet:

$$z_i = \frac{\bar{x}_i - 3.5}{\sqrt{\frac{17.5}{6n}}}$$

Für die Verteilung der Zufallsvariable Z gilt offensichtlich

$$P(\bar{X}_n \leq \bar{x}_i) = P(Z \leq z_i)$$

Die Verteilung der derart standardisierten Werte von $\bar{X}_n$ können wir mit der Standardnormalverteilung vergleichen. Wir betrachten die entsprechenden Verteilungsfunktionen für $n = 3$ (s. Abbildung 17.1.3):

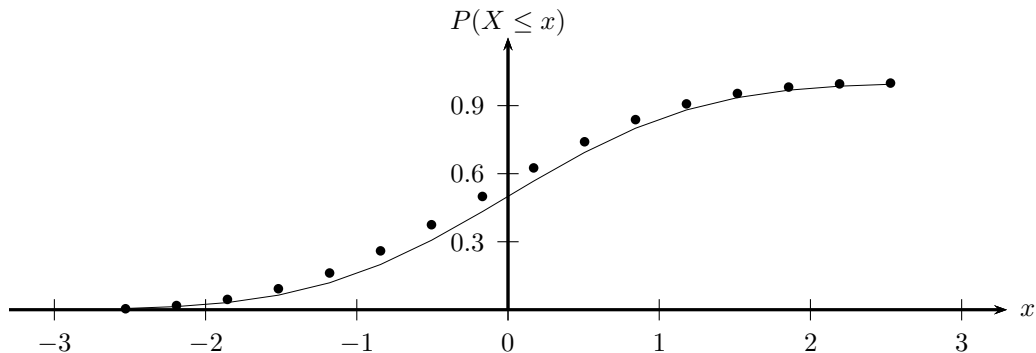


Abbildung 17.1.3: Verteilungsfunktion der Standardnormalverteilung und Verteilungsfunktion des Mittelwertes dreier uniform verteilter Zufallsvariablen mit sechs möglichen Ausgängen ($n=3$)

Man sieht, dass die zwei Verteilungsfunktionen recht gut zueinander passen. Betrachten wir den Mittelwert $\bar{X}$ von mehr identisch und uniform verteilten sowie unabhängigen Zufallsvariablen X_i , so nähert sich die Verteilungsfunktion der standardisierten Werte von $\bar{X}$ zunehmend der Verteilungsfunktion der Standardnormalverteilung. Diese Aussage gilt sogar für beliebige i.i.d verteilte Zufallsvariablen. Diese Aussage ist der wichtige *zentrale Grenzwertsatz*, den wir ohne Beweis formulieren:

Satz 17.1.1. $\bar{X}_n$ sei der Mittelwert von n Zufallsvariablen X_i . Für beliebig verteilte i.i.d. Zufallsvariablen X_i mit $0 < V(X_i) < \infty$ gilt:

$$\lim_{n \rightarrow \infty} \frac{\bar{X}_n - E(X_i)}{\sqrt{\frac{V(X_i)}{n}}} \sim N(0, 1)$$

oder in alternativer Formulierung

$$\lim_{n \rightarrow \infty} \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}} \sim N(0, 1)$$

(Für steigendes n nähert sich die Verteilung der Zufallsvariable $Z_n := \frac{\bar{X}_n - E(X_i)}{\sqrt{\frac{V(X_i)}{n}}}$ der Standardnormalverteilung).

Bemerkung 17.1.2. Die beiden Formulierungen des Zentralen Grenzwertsatzes sind äquivalent, da für gemäss Satz 9.5.18 und 9.5.36 für i.i.d. X_i gilt:

$$E(X_i) = E(\bar{X}_n)$$

und

$$\sqrt{\frac{V(X_i)}{n}} = \sqrt{V(\bar{X}_n)}$$

◇

Der zentrale Grenzwertsatz begründet u.a. die bedeutende Rolle der Normalverteilung in der Statistik: standardisierte Mittelwerte von Datensätzen, die aus Zufallsstichproben resultieren (= Werte der Statistik *Mittelwert*), sind näherungsweise standardnormalverteilt. Nicht-standardisierte Statistiken *Mittelwert* sind näherungsweise normalverteilt, wie der folgende Satz ausdrückt:

Satz 17.1.3. Für i.i.d. Zufallsvariablen X_i mit $0 < V(X_i) < \infty$ gilt:

$$\lim_{n \rightarrow \infty} \bar{X}_n \sim N(E(\bar{X}_n), V(\bar{X}_n))$$

oder äquivalent

$$\lim_{n \rightarrow \infty} \bar{X}_n \sim N(E(X_i), \frac{V(X_i)}{n})$$

Beweis. Aus

$$Z_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{V(\bar{X}_n)}}$$

folgt durch Auflösen nach $\bar{X}_n$

$$\bar{X}_n = Z_n \cdot \sqrt{V(\bar{X}_n)} + E(\bar{X}_n),$$

woraus unter Anwendung von Satz 14.2.4 folgt

$$\bar{X}_n \sim N(0 \cdot \sqrt{V(\bar{X}_n)} + E(\bar{X}_n), V(\bar{X}_n) \cdot 1) = N(E(\bar{X}_n), V(\bar{X}_n))$$

Zur Äquivalenz der beiden Formulierungen siehe die letzte *Bemerkung 17.1.2.*

□

Aus dem zentralen Grenzwertsatz folgt schliesslich, dass die standardisierte Statistik *Summe* von i.i.d. Zufallsvariable X_i ebenfalls näherungsweise normalverteilt ist:

Satz 17.1.4. Für i.i.d. X_i mit $0 < V(X_i) < \infty$ und $X_n = \sum_{i=1}^n X_i$ gilt:

$$\lim_{n \rightarrow \infty} \frac{X_n - E(X_n)}{\sqrt{V(X_n)}} \sim N(0, 1)$$

Beweis.

$$\begin{aligned} \frac{X_n - E(X_n)}{\sqrt{V(X_n)}} & \stackrel{\text{laut Voraussetzung}}{=} \frac{\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)}{\sqrt{V\left(\sum_{i=1}^n X_i\right)}} \stackrel{\text{mit } \frac{1}{n} \text{ erweitern}}{=} \\ & = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} E\left(\sum_{i=1}^n X_i\right)}{\frac{1}{n} \sqrt{V\left(\sum_{i=1}^n X_i\right)}} \stackrel{\text{Def Mittelwert}}{=} \frac{\bar{X}_n - \frac{1}{n} n E(X_i)}{\frac{1}{n} \sqrt{n V(X_i)}} = \frac{\bar{X}_n - E(X_i)}{\sqrt{\frac{V(X_i)}{n}}} \end{aligned}$$

□

17.2 Anwendungen des zentralen Grenzwertsatzes

Im Kapitel „Familie der Normalverteilungen“ haben wir die Berechnung von Werten der Binomialverteilung mit Hilfe der Normalverteilung damit gerechtfertigt, dass die Dichtefunktion der Standardnormalverteilung die Grenzfunktion einer Folge von spezifischen Treppenfunktionen ist, die wir durch die Standardisierung des entsprechenden Dichtefunktion der Binomialverteilung erhalten. Mit Hilfe des Satzes 17.1.4, dass die standardisierte Statistik *Summe* näherungsweise standardnormalverteilt ist, haben wir eine weitere Möglichkeit, um die Berechnung der Werte der Binomialverteilung mit Hilfe der Standardnormalverteilung zu begründen.

Satz 17.2.1. Sei $X \sim B(n, p)$. Dann ist $\frac{X - np}{\sqrt{np(1-p)}}$ näherungsweise standardnormalverteilt.

Beweis. Für $X \sim B(n, p)$ gilt: $X = \sum_{i=1}^n X_i$, wobei $X_i \sim B(1, p)$ (siehe Satz 10.2.3: $X \sim B(n, p)$ ist die Summe von n nach $B(1, p)$ verteilten Zufallsvariablen) mit X_i i.i.d. Damit gilt laut dem Satz 17.1.4 näherungsweise

$$\frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{n(p(1-p))}} \sim N(0, 1).$$

(der letzte Schritt gilt, da für die Binomialverteilung gilt:

$$E(X) = np; V(X) = np(1-p)).$$

□

Es handelt sich um dieselbe Formel wie die Formel, die wir bereits am Ende des Kapitels über die Normalverteilung verwendet hatten.

Auf Grund ähnlicher Überlegungen können wir die Werte $P(X \leq x)$ von *Poissonverteilungen* ebenfalls mit Hilfe der Normalverteilung berechnen, wenn λ genügend gross ist. Die Herleitung erfolgt analog zur obigen bezüglich der Binomialverteilung:

Satz 17.2.2. Sei $X \sim P(\lambda)$. Dann ist $\frac{X - \lambda}{\sqrt{\lambda}}$ näherungsweise standardnormalverteilt.

Beweis. Für $X \sim P(\lambda)$ gilt gemäss Satz 10.3.10: $X = \sum_{i=1}^n X_i$ mit i.i.d. $X_i \sim P(\frac{\lambda}{n})$. Damit gilt gemäss Satz 17.1.4 näherungsweise

$$\frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \lambda}{\sqrt{\lambda}} \sim N(0, 1).$$

Der letzte Schritt gilt, da für die Poissonverteilung gilt: $E(X) = \lambda; V(X) = \lambda$. □

Beispiel 17.2.3. Die Anzahl der Maschinenstillstände pro Tag sei poissonverteilt mit $\lambda = 15$. Wieviel beträgt die Wahrscheinlichkeit, dass mindestens 8 Maschinen an einem Tag stillstehen? Mit Hilfe der Näherungsformel erhalten wir für

$$P(X \leq 7) \approx \Phi\left(\frac{7 - 15}{\sqrt{15}}\right) = \Phi(-2.0656) = 0.019433$$

Die gesuchte Wahrscheinlichkeit ist somit: $1 - 0.019433 = 0.980566934$

Als exakten Wert erhalten wir $P(X \geq 8) = 1 - \sum_{i=0}^7 \frac{15^i e^{-15}}{i!} = 1 - 0.018002193 = 0.982$. (Die Differenz zwischen dem tatsächlichen und dem näherungsweisen Wert beträgt somit: $0.982 - 0.980566934 = 0.0014331$). ◇

Im Allgemeinen ist es möglich, bei Poissonverteilungen die Normalverteilung zu verwenden, wenn gilt: $\lambda \geq 9$.

Bemerkung 17.2.4. Wir können laut dem zentralen Grenzwertsatz beliebig verteilte i.i.d. Zufallszahlen verwenden, um standardnormalverteilte Zufallszahlen zu erhalten. Die jeweiligen Mittelwerte sind näherungsweise normalverteilt. So können wir z.B. uniform verteilte Zufallszahlen im Intervall $[0, 1]$ verwenden, wobei wir statt des empirischen Mittelwertes und der empirischen Standardabweichung für die Standardisierung die Werte des Modells verwenden (Ist $X \sim U[0, 1]$, gilt $V(X) = \frac{1}{12}$ und $E(X) = \frac{1}{2}$). Wir erhalten:

$$\frac{\frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{2}}{\sqrt{\frac{1}{12n}}} = \frac{\frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{2}}{\frac{1}{\sqrt{12n}}} = \frac{\sum_{i=1}^n x_i - \frac{n}{2}}{\sqrt{12n}} = \frac{\sum_{i=1}^n x_i - \frac{n}{2}}{\sqrt{\frac{n^2}{12n}}} = \frac{\sum_{i=1}^n x_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \sim N(0, 1) \quad \diamond$$

17.2.1 Übungen

1. Reproduzieren Sie die Abbildungen 17.1.1 und 17.1.2 mit Excel.
2. Reproduzieren Sie die Abbildungen 17.1.3 mit Excel.
3. Auf dem Übungs-Excelblatt Ueb3 in „ZentralerGrenzwertSatz.xls“ finden Sie eine Anzahl Datensätze, deren Daten als Werte von i.i.d. verteilten Zufallsvariablen betrachtet werden können. Berechnen Sie für alle Datensätze die Mittelwerte und zeichnen Sie die Verteilungsfunktion der standardisierten Mittelwerte (Schätzen sie den Erwartungswert und die Standardabweichung aus den Mittelwerten). Zeichnen Sie diese in ein Koordinatennetz und legen Sie dann die Verteilungsfunktion der Standardnormalverteilung darüber. Kommentieren Sie das Resultat.
4. Berechnen Sie die folgenden Werte der folgenden Poissonverteilungen (exakt und mit Näherung):
 $P(X \leq 6), X \sim P(13); \quad P(X \leq 4), X \sim P(5)$

17.2.2 Lösungen

1. Siehe Exceltabelle „ZentralerGrenzwertSatz.xls“ LUeb1.

Mit R: `n2=expand.grid(1:6,1:6); n2s=apply(n2,1,mean)`
`plot(x=unique(n2s),y=table(n2s)/6^2)`

`n3=expand.grid(1:6,1:6,1:6); n3s=apply(n3,1,mean)`
`plot(x=unique(n3s),y=table(n3s)/6^3)`

2. Siehe Exceltabelle „ZentralerGrenzwertSatz.xls“ LUeb2.

Mit R: `n3=expand.grid(1:6,1:6,1:6); n3s=apply(n3,1,mean)`
`plot((x=unique(n3s)-3.5)/sqrt(17.5/(6*3)),y=cumsum(table(n3s)/6^3))`
`lines(seq(-2.5,2.5,0.1),pnorm(seq(-2.5,2.5,0.1)),lty=1,col="red")`

3. Siehe Exceltabelle „ZentralerGrenzwertSatz.xls“ LUeb3. Kommentar: Die standardisierten Mittelwerte sind offensichtlich standardnormalverteilt. Dies könnte man zusätzlich mit einem P-P-Plot überprüfen, der ebenfalls eine sehr gute Übereinstimmung ergibt (s. LUeb3).

4. Da $\lambda = 13 \geq 9$, dürfen wir die Normalverteilung verwenden:

$$\frac{6-13}{\sqrt{13}} = -1.9415; \quad \Phi(-1.9415) = 0.02609876 = P(X \leq 7)$$

Der genaue Wert beträgt: 0.025886915

Da $\lambda = 5 < 9$, dürfen wir nicht die Normalverteilung verwenden:

Der genaue Wert beträgt:

$$P(X \leq 4), X \sim P(5); \quad P(X \leq 4) = 0.440493285$$

(Mit der Normalverteilung würden wir erhalten: $\frac{4-5}{\sqrt{5}} = -.44721$

$P(X \leq 4) \approx 0.327361743 = \Phi(-0.44721)$ - was doch bedeutend daneben wäre).

17.3 Lernziele

1. Den Zentralen Grenzwertsatz in Worten formulieren können.
2. Die Begründung - mittels des zentralen Grenzwertsatzes - für die Berechnung von Wahrscheinlichkeiten von binomial- und poisson-verteilten Zufallsvariablen durch die Standardnormalverteilung liefern können (fakultativ).
3. Übungen der obigen Art (mit Excel) machen können.

Kapitel 18

Weitere stetige Verteilungsfamilien

Wir werden in diesem Rahmen nicht sämtliche, mit Namen versehenen stetigen Verteilungsfamilien betrachten, sondern nur jene, die eine besondere Rolle in der Test-Theorie spielen: Die χ_n^2 -Verteilungen (Chi-Quadrat), die t_n -Verteilungen und die $F_{m,n}$ -Verteilungen. Diese Verteilungen werden von der Normalverteilung hergeleitet. Die entsprechenden Herleitungen, die auf Operationen auf den Dichtefunktionen beruhen, sind technisch anspruchsvoll und brauchen uns nicht weiter zu interessieren.

18.1 Die Familie der χ_n^2 -Verteilungen

Definition 18.1.1. Y ist χ_n^2 -verteilt genau dann, wenn

- (i) $Y = \sum_{i=1}^n X_i^2$
- (ii) $X_i \sim N(0, 1)$ und
- (ii) X_i sind unabhängig.

Wir nennen Y eine χ_n^2 -verteilte Zufallsvariable mit n Freiheitsgraden. n ist der einzige Parameter der Verteilung. Die Anzahl der Freiheitsgrade ist mit der Anzahl der unabhängigen Zufallsvariablen identisch. Je nach dieser Anzahl sieht die Verteilung etwas anders aus. Wir kürzen Freiheitsgrade künftig oft mit „df“ ab (= degrees of freedom) ab.

Die Dichtefunktion der χ_n^2 -Quadratverteilung mit n Freiheitsgraden ist:

$$f_{\chi_n^2} : \mathbb{R} \rightarrow \mathbb{R}_+ \\ f_{\chi_n^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}} & (\text{für } x > 0) \\ f_{\chi_n^2}(x) = 0 & \text{sonst} \end{cases}$$

wobei Γ die Gammafunktion ist. Die Formel der Dichtefunktionen folgt aus der Definition der χ_n^2 -Verteilungen.

18.1.1 Ein paar Bemerkungen zur Gammafunktion (fakultativ)

Die Gammafunktion ist definiert durch:

$$\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+ \\ \Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Die Gammafunktion ist die Funktion, welche die diskrete Funktion $n!$ fortführt, d.h. so verbindet, dass eine stetige Funktion ohne überflüssige Krümmungen entsteht. Es gilt nämlich: $\Gamma(n+1) = n!$. Damit ist z.B. $\Gamma(6) = \Gamma(5+1) = 5! = 120$. (s. Abbildung 18.1.1)

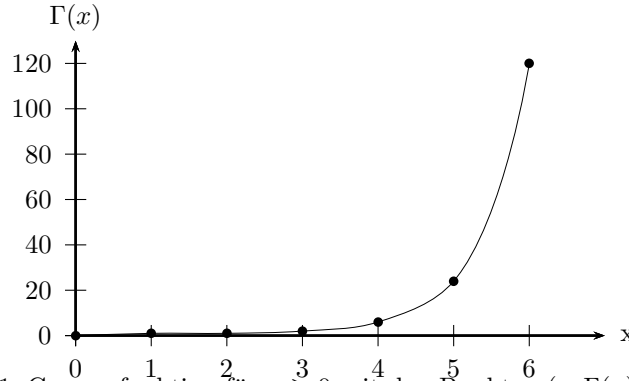


Abbildung 18.1.1: Gammafunktion für $x \geq 0$ mit den Punkten $(x, \Gamma(x))$, mit $\Gamma(x+1) = x!$

Die Werte der Gammafunktion kann man mit Excel berechnen:

Excel (Befehl: =GAMMALN()) liefert die logarithmierten Werte der Gammafunktion (natürlicher Logarithmus $\ln$). Die Werte der Gammafunktion erhalten wir somit durch: $\exp(\text{GAMMALN}()) = \Gamma(x)$.

Faktisch berechnen wir in den hier verwendeten Formeln die Gammawerte von ganz- oder halbzahlig Werten. Dabei gilt:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Da für die Gammafunktion zusätzlich gilt

$$\Gamma(x+1) = x\Gamma(x),$$

erhalten wir alle gewünschten Werte nichtganzzahliger Argumente wie folgt:

$$\Gamma\left(\frac{3}{2}\right) = \Gamma\left(1 + \frac{1}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi} = 0.886\,23$$

$$\Gamma\left(\frac{5}{2}\right) = \Gamma\left(\frac{3}{2} + 1\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2} \cdot \frac{1}{2}\sqrt{\pi} = 1.329\,3$$

$$\Gamma\left(\frac{7}{2}\right) = \Gamma\left(\frac{5}{2} + 1\right) = \frac{5}{2}\Gamma\left(\frac{5}{2}\right) = \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2}\sqrt{\pi} = 3.323\,4$$

Allgemein:

$$\Gamma\left(\frac{2n+1}{2}\right) = \left(\prod_{i=1}^n \frac{2i-1}{2}\right) \sqrt{\pi}$$

$$\text{(z.B. } \Gamma\left(\frac{19}{2}\right) = \Gamma\left(\frac{2 \cdot 9 + 1}{2}\right) = \left(\prod_{i=1}^9 \frac{2i-1}{2}\right) \sqrt{\pi} = \frac{17}{2} \cdot \frac{15}{2} \cdot \frac{13}{2} \cdot \frac{11}{2} \cdot \frac{9}{2} \cdot \frac{7}{2} \cdot \frac{5}{2} \cdot \frac{3}{2} \cdot \frac{1}{2} \sqrt{\pi} = 1192\,92.461995)$$

18.1.2 Graphische Darstellung der Dichtefunktionen

Konkrete Dichtefunktionen von χ_n^2 -Quadratverteilungen mit n Freiheitsgraden erhalten wir, indem wir in der obigen Formel n durch eine spezifische natürliche Zahl und $\Gamma\left(\frac{n}{2}\right)$ durch den entsprechenden Wert der Gamma-Funktion ersetzen.

Für *einen* Freiheitsgrad erhalten wir die Dichtefunktion:

$$f_{\chi^2_1}(x) = \frac{1}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} \cdot x^{\frac{1}{2}-1} \cdot e^{-\frac{x}{2}} = 0.398\,94 \cdot x^{-0.5} e^{-0.5x}$$

Wir erhalten den Graphen (s. Abbildung 18.1.2):

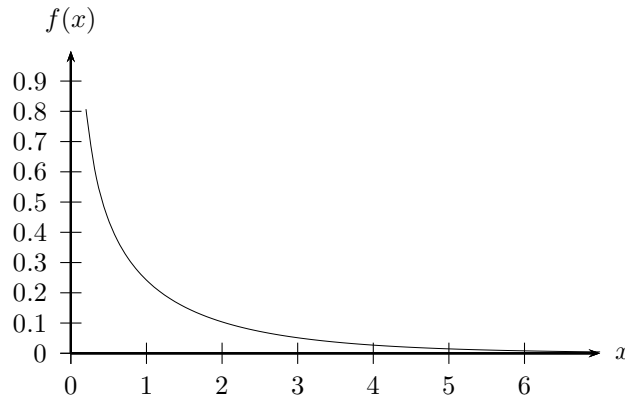


Abbildung 18.1.2: Graphische Darstellung der Dichtefunktion der χ^2_1 -Verteilung

Die Kurve ist nicht symmetrisch. Entsprechend fallen der Median und der Mittelwert nicht zusammen.

Für *zwei* Freiheitsgrade erhalten wir, da $\Gamma(\frac{2}{2}) = \Gamma(1) = 1$:

$$f_{\chi^2_2}(x) = \frac{1}{2^{\frac{2}{2}}\Gamma(\frac{2}{2})} \cdot x^{\frac{2}{2}-1} \cdot e^{-\frac{x}{2}} = \frac{1}{2} e^{-\frac{1}{2}x}$$

Wir erhalten den Graphik (s. Abbildung 18.1.3):

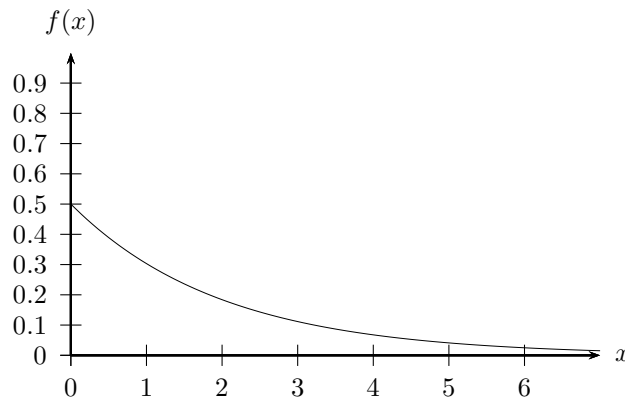


Abbildung 18.1.3: Graphische Darstellung der Dichtefunktion der χ^2_2 -Verteilung

Für n Freiheitsgrade erhalten wir:

$$\begin{aligned}
n = 3 : f_{\chi_3^2}(x) &= \frac{1}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})} \cdot x^{\frac{3}{2}-1} \cdot e^{-\frac{x}{2}} = 0.39894\sqrt{x}e^{-0.5x} \\
n = 4 : f_{\chi_4^2}(x) &= \frac{1}{2^{\frac{4}{2}}\Gamma(\frac{4}{2})} \cdot x^{\frac{4}{2}-1} \cdot e^{-\frac{x}{2}} = 0.25xe^{-0.5x} \\
n = 5 : f_{\chi_5^2}(x) &= \frac{1}{2^{\frac{5}{2}}\Gamma(\frac{5}{2})} \cdot x^{\frac{5}{2}-1} \cdot e^{-\frac{x}{2}} = 0.13298\sqrt{x^3}e^{-0.5x} \\
n = 8 : f_{\chi_8^2}(x) &= \frac{1}{2^{\frac{8}{2}}\Gamma(\frac{8}{2})} \cdot x^{\frac{8}{2}-1} \cdot e^{-\frac{x}{2}} = 0.010417x^3e^{-0.5x} \\
n = 10 : f_{\chi_{10}^2}(x) &= \frac{1}{2^{\frac{10}{2}}\Gamma(\frac{10}{2})} \cdot x^{\frac{10}{2}-1} \cdot e^{-\frac{x}{2}} = 0.0013021x^4e^{-0.5x}
\end{aligned}$$

Graphik (s. Abbildung 18.1.4):

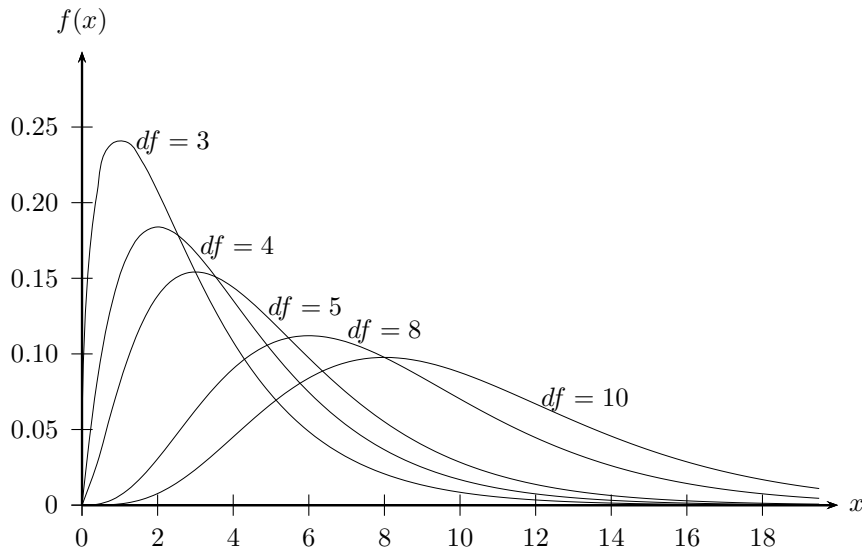


Abbildung 18.1.4: Graphische Darstellung der Dichtefunktion der χ_i^2 -Verteilung mit $i \in \{3, 4, 5, 8, 10\}$

Die Graphen haben für $n > 2$ ein Maximum in $n - 2$. Es handelt sich um rechtsschiefe (linkssteile) Verteilungen.

Für beliebige χ_n^2 -Verteilungen gilt: Der Erwartungswert ist mit n identisch und die Varianz mit $2n$. Damit liegt der Erwartungswert rechts vom Maximum der Dichtefunktion.

An den obigen Graphiken stellen wir fest: mit steigendem n werden die Kurven immer flacher. Die Kurven werden zudem symmetrischer. Mit steigendem n nähern sich das Maximum und der Erwartungswert. Es gilt ja $\lim_{n \rightarrow \infty} \frac{n-2}{n} = 1$. In der Tat unterscheidet sich die Kurve bei hohem n nur mehr geringfügig von einer Normalverteilung mit $\mu = n$ und $\sigma^2 = 2n$. Als Beispiel betrachten wir $X \sim N(50, 100)$ und X' ist χ_{50}^2 -verteilt. Wir erhalten die Formeln

$$\begin{aligned}
f_{N(50,100)}(x) &= \frac{1}{10\sqrt{2 \cdot \pi}} e^{-0.5\left(\frac{x-50}{10}\right)^2} \\
f_{\chi_{50}^2}(x) &= \frac{1}{2^{\frac{50}{2}} \cdot \Gamma\left(\frac{50}{2}\right)} \cdot x^{\frac{50}{2}-1} \cdot e^{-\frac{x}{2}} = 4.8034 \times 10^{-32} x^{24} e^{-0.5x}
\end{aligned}$$

und die Graphiken (s. Abbildung 20.2.4):

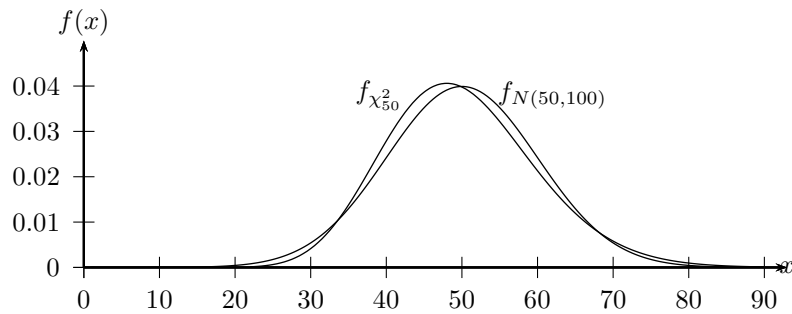


Abbildung 18.1.5: Graphische Darstellung der Dichtefunktion der χ^2_{50} -Verteilung und der Normalverteilung mit $\mu = 50$ und $\sigma = 10$

Die folgende Tabelle gibt Auskunft über die Unterschiede für $P(X \leq x)$ fürs Beispiel der Abbildung 20.2.4 (s. Tabelle 18.1.1)

x	$P(X \leq x)$	$P(X' \leq x)$	Differenz
	mit $X \sim N(50, 100)$	mit X' ist χ^2_{50} -verteilt	
10	3.16712E-05	1.59959E-10	3.16711E-05
20	0.001349898	4.69494E-05	0.001302949
30	0.022750132	0.011164779	0.011585353
40	0.158655254	0.15677259	0.001882664
50	0.5	0.526601536	-0.026601536
60	0.841344746	0.84275797	-0.001413224
70	0.977249868	0.96762589	0.009623978
80	0.998650102	0.995517343	0.003132759
90	0.999968329	0.999550754	0.000417575
100	0.999999713	0.999965451	3.42627E-05

Tabelle 18.1.1: Tabelle zu den Differenzen der Wahrscheinlichkeiten für $P(X \leq x)$ mit $X \sim N(50, 100)$ und mit X' ist χ^2_{50} -verteilt

Für die Verteilungsfunktion einer χ^2_n -verteilten Zufallsvariable brauchen wir künftig die Abkürzung

$$F_{\chi^2_n}(x) := P(X \leq x) \text{ für } \chi^2_n\text{-verteilte } X.$$

Anwendungsbeispiele für die Chi-Quadrat-Verteilungen werden wir im Rahmen der Testtheorie sehen.

18.1.3 Berechnung von Wahrscheinlichkeiten

Auch hier gibt es für die Verteilungsfunktion keine geschlossene Formel. Wir könnten die Werte der Funktion durch eine Näherungsmethode bestimmen (z.B. Trapezmethode, siehe Verfahren bei Normalverteilung). Wir werden die Werte jedoch mit Hilfe von Excel oder R bestimmen:

Excel: =CHIQU.VERT(x;Freiheitsgrade;1) die Werte der Verteilungsfunktion $F_{\chi^2_n}(x)$. Setzt man statt 1 0, werden die Werte der Dichtfunktion geliefert.

Beispiel: Wir möchten $F_{\chi^2_6}(8.3)$ berechnen. Wir geben „=chiqu.vert(8.3;6;1)“ ein und erhalten 0.783061924

Mit R: pchisq(8.3,6)

18.1.4 Übungen

Bestimmen Sie mit Hilfe von Excel und R die folgenden Werte:

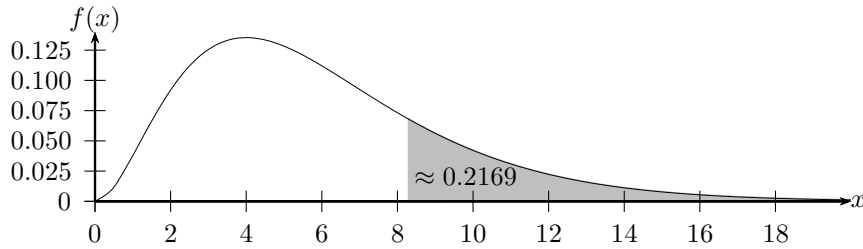


Abbildung 18.1.6: Excelausgabe für „=1-chiqu.vert(8.3;6;1)“ entspricht der grauen Fläche unter der Dichtefunktion der entsprechenden Chi-Quadrat-Verteilung.

1. $P(X \leq 1.64)$ für die χ_6^2 -verteilte Zufallsvariable X .
2. $P(X \leq 21.9)$ für die χ_{11}^2 -verteilte Zufallsvariable X .
3. $P(X > 14.9)$ für die χ_4^2 -verteilte Zufallsvariable X .
4. $P(X > 0.6)$ für die χ_5^2 -verteilte Zufallsvariable X .
5. $P(28 < X \leq 80)$ für die χ_{50}^2 -verteilte Zufallsvariable X .

18.1.5 Lösungen

1. 0.050341
2. 0.97484
3. 0.004913182
4. 0.988003243
5. 0.9905

18.2 Die Familie der t_n -Verteilungen

Definition 18.2.1. Y ist t_n -verteilt genau dann, wenn

- (i) $Y = \frac{X_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}$
- (ii) $X_i \sim N(0, 1)$ für $0 \leq i \leq n$
- (iii) die X_i sind unabhängig (für $0 \leq i \leq n$).

Wir nennen Y eine t -verteilte Zufallsvariable mit n Freiheitsgraden. n ist der Parameter der Verteilung. t_n -Verteilungen sind symmetrische Verteilungen mit dem Erwartungswert 0 (für $n \geq 2$) und der Varianz $\frac{n}{n-2}$ (sofern dieser Term positiv ist, d.h. $n > 2$). Eine t_n -verteilte Zufallsvariable ist laut Definition der Quotient von zwei Zufallsvariablen, wobei der Nenner die Wurzel aus einer chiquadratverteilten Zufallsvariable ist, die mit Kehrwert ihrer Freiheitsgrade gewichtet wurde.

Die Dichtefunktion der t_n -Verteilung ist:

$$f_{t_n} : \mathbb{R} \rightarrow \mathbb{R}_+$$

$$f_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{\pi n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

18.2.1 Graphische Darstellung der Dichtefunktionen

Für

$$n = 1 : f_{t_1}(x) = \frac{\Gamma\left(\frac{1+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\sqrt{\pi}} \left(1 + \frac{x^2}{1}\right)^{-\frac{1+1}{2}} = \frac{0.318\,31}{1+x^2}$$

$$n = 2 : f_{t_2}(x) = \frac{\Gamma\left(\frac{2+1}{2}\right)}{\Gamma\left(\frac{2}{2}\right)\sqrt{\pi \cdot 2}} \left(1 + \frac{x^2}{2}\right)^{-\frac{2+1}{2}} = \frac{0.353\,55}{\left(\sqrt{1+0.5x^2}\right)^3}$$

$$n = 10 : f_{t_{10}}(x) = \frac{\Gamma\left(\frac{10+1}{2}\right)}{\Gamma\left(\frac{10}{2}\right)\sqrt{\pi \cdot 10}} \left(1 + \frac{x^2}{10}\right)^{-\frac{10+1}{2}} = \frac{0.389\,11}{\left(\sqrt{1+0.1x^2}\right)^{11}}$$

Graphik (s. Abbildung 18.2.7):

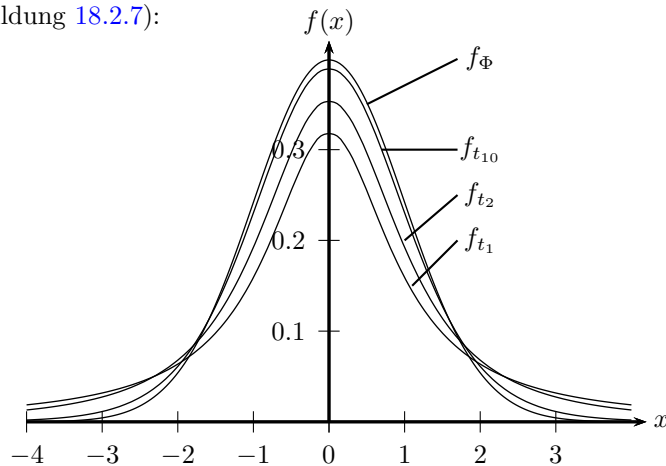


Abbildung 18.2.7: Dichtefunktion f_Φ der Standardnormalfunktion; Dichtefunktionen f_{t_i} der t -Verteilungen mit i Freiheitsgraden

Wir stellen fest: Die t_n -Verteilungen sind flacher als die Standardnormalverteilung. Sie nähern sich für steigende n aber zunehmend einer Standardnormalverteilung an (s. Tabelle 18.2.2).

x	$P(X \leq x)$		Differenzen
	für $X \sim N(0,1)$	für X ist t_{50} -verteilt	
-5	2.86652E-07	3.71661E-06	-3.42995E-06
-4	3.16712E-05	0.000104595	-7.29239E-05
-3	0.001349898	0.002100852	-0.000750953
-2	0.022750132	0.025473532	-0.0027234
-1	0.158655254	0.161062823	-0.002407569
0	0.5	0.5	0
1	0.841344746	0.838937177	0.002407569
2	0.977249868	0.974526468	0.0027234
3	0.998650102	0.997899148	0.000750953
4	0.999968329	0.999895405	7.29239E-05
5	0.999999713	0.999996283	3.42995E-06

Tabelle 18.2.2: Differenzen von $P(X \leq x)$ für $X \sim N(0,1)$ und X ist t_{50} -verteilt

Für die Verteilungsfunktion einer t_n -verteilten Zufallsvariable brauchen wir künftig die Abkürzung

$$F_{t_n}(x) := P(X \leq x) \text{ für } X \text{ ist } t_n - \text{verteilt.}$$

Anwendungsbeispiele für die t -Verteilungen werden wir im Rahmen der Testtheorie sehen.

18.2.2 Berechnung der Wahrscheinlichkeiten

Die Werte der t_n -Verteilung können wir mit Hilfe von Näherungsverfahren aus der Dichtefunktion berechnen (Trapezmethode). In der Praxis werden wir sie mit R oder mit Hilfe von Excel bestimmen.

Excel: `f.vert(x;5;1)` liefert den Wert $F_{t_n}(x)$ der Verteilungsfunktion.

Wir möchten $F_{t_{15}}(2.1)$ berechnen. Mit „`=t.vert(2.1;15;1)`“ erhalten wir 0.973472372.

R: `pt(2.1,15)`

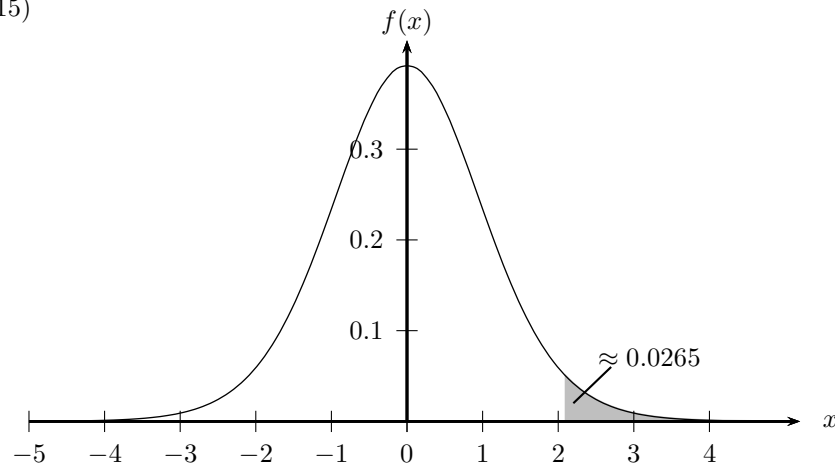


Abbildung 18.2.8: Excel-Ausgabe für $1 - t.vert(2.1; 15; 1)$ entspricht der grauen Fläche unter der Dichtefunktion der entsprechenden t -Verteilung.

18.2.3 Übungen

Geben sie die folgenden Wahrscheinlichkeiten an (mit R und mit Excel):

1. $P(X \leq 4.54)$ für eine $t_{(3)}$ -verteilte Zufallsvariable X .
2. $P(X \leq 2.06)$ für eine $t_{(25)}$ -verteilte Zufallsvariable X .
3. $P(X > 2.23)$ für eine $t_{(10)}$ -verteilte Zufallsvariable X .
4. $P(1.7 < X \leq 3.39)$ für eine $t_{(30)}$ -verteilte Zufallsvariable X .
5. $P(X \leq -3.3)$ für eine $t_{(30)}$ -verteilte Zufallsvariable X .

18.2.4 Lösungen

1. 0.99
2. 0.97502
3. 0.024921235
4. 0.048751508
5. 0.001249654

18.3 Die Familie der $F_{m,n}$ -Verteilungen

Definition 18.3.1. Y ist eine $F_{m,n}$ -verteilte Zufallsvariable genau dann, wenn

- (i) $Y = \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{i=1}^n X_i^2}$
- (ii) $X_i \sim N(0, 1)$ für $1 \leq i \leq m$ und $1 \leq i \leq n$
- (iii) die X_i sind unabhängig.

Wir nennen Y eine $F_{m,n}$ -verteilte Zufallsvariable mit m Zählerfreiheitsgraden und n Nennerfreiheitsgraden. m und n sind die Parameter der Verteilung. Damit ist eine $F_{m,n}$ -verteilte Zufallsvariable ein Quotient von zwei chiquadratverteilten Zufallsvariablen, die mit dem Kehrwert ihrer Freiheitsgrade gewichtet sind. (Die Abkürzung $F_{m,n}$ für diese Verteilung wurde zu Ehren von R.A. Fisher, einem der Begründer der Statistik und Erfinder der $F_{m,n}$ -Verteilung, gewählt).

Die Dichtefunktion der $F_{m,n}$ -Verteilung ist:

$$f_{F_{m,n}} : \mathbb{R} \rightarrow \mathbb{R}_+$$

$$f_{F_{m,n}}(x) = \begin{cases} \frac{x^{\frac{m}{2}-1} \left(\frac{m}{n}\right)^{\frac{m}{2}}}{\frac{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}{\Gamma(\frac{m+n}{2})}} (1 + \frac{m}{n}x)^{-\frac{m+n}{2}} & (\text{für } x > 0) \\ f_{F_{m,n}}(x) = 0 & \text{sonst.} \end{cases}$$

18.3.1 Graphische Darstellung der Dichtefunktionen

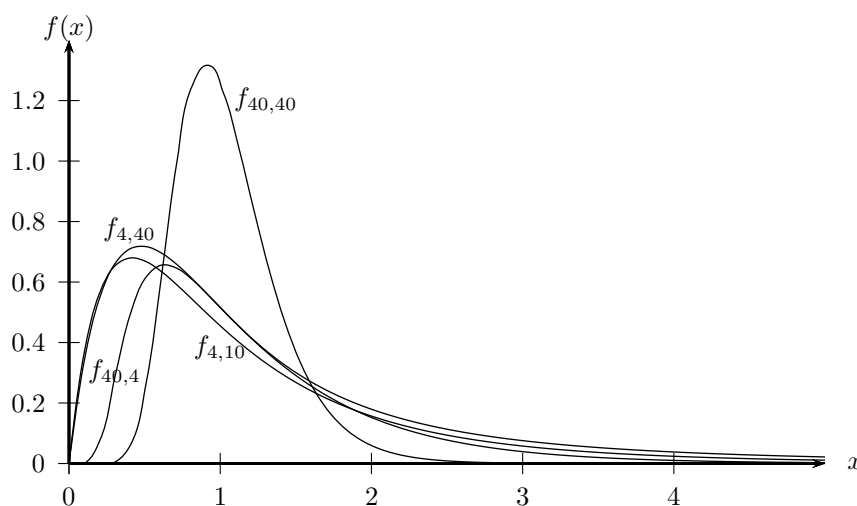
Wir zeichnen ein paar Dichtefunktionen von $F_{m,n}$ -Verteilungen (s. Abbildung 18.3.9).

$$f_{F_{4,10}}(x) = \frac{x^{\frac{4}{2}-1} \left(\frac{4}{10}\right)^{\frac{4}{2}}}{\frac{\Gamma(\frac{4}{2})\Gamma(\frac{10}{2})}{\Gamma(\frac{4+10}{2})}} (1 + \frac{4}{10}x)^{-\frac{4+10}{2}} = \frac{24x}{5(1+0.4x)^7}$$

$$f_{F_{4,40}}(x) = \frac{x^{\frac{4}{2}-1} \left(\frac{4}{40}\right)^{\frac{4}{2}}}{\frac{\Gamma(\frac{4}{2})\Gamma(\frac{40}{2})}{\Gamma(\frac{4+40}{2})}} (1 + \frac{4}{40}x)^{-\frac{4+40}{2}} = \frac{4.2x}{(1+0.1x)^{22}}$$

$$f_{F_{40,4}}(x) = \frac{x^{\frac{40}{2}-1} \left(\frac{40}{4}\right)^{\frac{40}{2}}}{\frac{\Gamma(\frac{40}{2})\Gamma(\frac{4}{2})}{\Gamma(\frac{40+4}{2})}} (1 + \frac{40}{4}x)^{-\frac{40+4}{2}} = 4.2 \times 10^{22} \frac{x^{19}}{(1+10x)^{22}}$$

$$f_{F_{40,40}}(x) = \frac{x^{\frac{40}{2}-1} \left(\frac{40}{40}\right)^{\frac{40}{2}}}{\frac{\Gamma(\frac{40}{2})\Gamma(\frac{40}{2})}{\Gamma(\frac{40+40}{2})}} (1 + \frac{40}{40}x)^{-\frac{40+40}{2}} = 1.3785 \times 10^{12} \frac{x^{19}}{(1+x)^{40}}$$

Abbildung 18.3.9: Dichtefunktionen einiger $F_{m,n}$ -Verteilungen

Der Erwartungswert der $F_{m,n}$ -Verteilung ist:

$$E_{F_{m,n}}(X) = \frac{n}{n-2}$$

($n > 2$; n = Nennerfreiheitsgrade). Die Varianz ist:

$$V_{F_{m,n}}(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$$

für $n > 4$ (m = Zählerfreiheitsgrade, n = Nennerfreiheitsgrade).

Für die Verteilungsfunktion einer $F_{m,n}$ -verteilten Zufallsvariable brauchen wir künftig die Abkürzung

$$F_{F_{m,n}}(x) := P(X \leq x) \text{ für } X \text{ ist } F_{m,n} \text{ - verteilt.}$$

Anwendungen der $F_{m,n}$ -Verteilungen werden wir im Rahmen der Testtheorie sehen.

18.3.2 Berechnung von Wahrscheinlichkeiten

Die Werte der $F_{m,n}$ -Verteilungen können mit Hilfe der Dichtefunktion und einem Näherungsverfahren bestimmte werden (z.B. Trapezmethode). In der Praxis werden sie mit Hilfe eines Rechners oder mit Excel bestimmt:

Excel: `f.vert(x;Zählerfreiheitsgrade m; Nennerfreiheitsgrade n;1)` liefert die Werte der Verteilungsfunktion $F_{F_{m,n}}$ der $F_{m,n}$ -Verteilung.

Beispiel: Wir möchten $F_{F_{5,7}}(2)$ berechnen. Wir erhalten mit „`f.vert(2;5;7;1)`“ 0.804326755
R: `pf(2,5,7)`

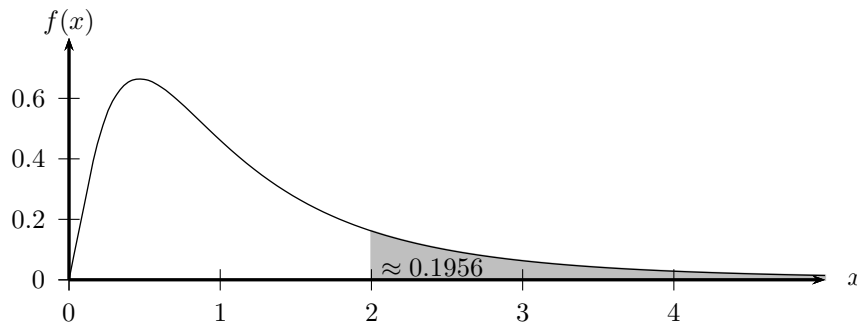


Abbildung 18.3.10: Excel-Ausgabe für „1=f.vert(2;5;7;1)“ entspricht der grauen Fläche unter der Dichtefunktion der entsprechenden F-Verteilung.

18.3.3 Übungen

Schauen Sie die folgenden Wahrscheinlichkeiten (mit Excel und R) nach

1. $P(X \leq 3.347)$ für eine $F_{10,8}$ -verteilte Zufallsvariable X .
2. $P(X \leq 3.289)$ für eine $F_{3,6}$ -verteilte Zufallsvariable X .
3. $P(X > 14.47)$ für eine $F_{9,3}$ -verteilte Zufallsvariable X .
4. $P(4.107 < X \leq 15.98)$ für eine $F_{4,4}$ -verteilte Zufallsvariable X .
5. Lösen Sie die Übungen von Weitere_stetige_Verteilungen_weitere_Uebungen_Juni_10.xls, die Sie auf der Home-Page <http://math.logik.ch> unter Statistik, W-Theorie II finden.

18.3.4 Lösungen

1. 0.949 99.
2. 0.900 01.
3. 0.025007557.
4. 0.090012472

18.4 Lernziele

- Mit Excel und einer anderen Methode die Wahrscheinlichkeiten für χ^2 -, t - und $F_{m,n}$ -verteilte Zufallsvariablen berechnen können.
- Wissen, was Freiheitsgrade ausdrücken (Anzahl unabhängiger Zufallsvariablen, die in die jeweiligen Statistiken (= Zufallsvariablen) einfließen).
- Definitionen der χ^2 -, t - und $F_{m,n}$ -Verteilungen kennen (Zusammenhang mit der Normalverteilung). Wissen, dass Dichtefunktionen der χ^2 - und $F_{m,n}$ -Verteilungen nur auf $\mathbb{R}^+$ von 0 verschieden sind und dass sie - besonders für kleine Freiheitsgrade rechtsschief sind. Wissen, dass die Dichtefunktion der t -Verteilung symmetrisch zur y -Achse ist.
- Die Form der Dichtefunktionen der behandelten Verteilungen zeichnen können (in welchen Quadranten, ungefähre Form - Schiefe, Symmetrie).

Teil IV

Schliessende Statistik - Testtheorie

Einleitung

In der schliessenden Statistik geht es um folgende Grundprobleme:

1. Ist eine bestimmte Hypothese mit den Ergebnissen eines Experimentes verträglich? Unterscheiden sich zwei Testergebnisse? (z.B. Sind zwei Rohstoffe bezüglich der Qualität des Endproduktes unterscheidbar? Führt eine bestimmte Investition in den Maschinenpark zu einer besseren Qualität? Hat ein bestimmtes Medikament die gewünschte Wirkung? Führt eine neue Mastmethode zu höheren Erträgen? etc.). In der „Wirklichkeit“ sind entsprechende Messwerte nämlich immer von vielen Einflüssen geprägt. Kein Produkt ist genau gleich wie ein anderes. Kein Medikament wirkt auf alle gleich. Keine Mastmethode führt bei Tieren genau zum gleichen Gewicht. Diese Beispiele zeigen: Messwerte streuen und es stellt sich die Frage, wie man angesichts dieser Streuung Qualität oder Wirksamkeit beurteilen kann.

Wie in der beschreibenden Statistik kann man in der Testtheorie univariate von bivariaten Tests unterscheiden. Bei den univariaten Tests beschäftigen wir uns nur mit einer Variablen (z.B. Körpergewicht). Bei den bivariaten Tests wird der Zusammenhang zweier Variablen untersucht (z.B. Rauchverhalten und Krebsrisiko). Die bivariaten Tests können auf multivariate Tests erweitert werden, mit denen wir den Einfluss mehrere Variablen auf eine Variable untersuchen (z.B. Einfluss von Einkommen, Geschlecht und Ausbildung auf Kaufverhalten bezüglich Schmuck). Wir werden in diesem Kapitel zuerst die univariate Testtheorie ausführen. Später folgen Beispiele für bivariate Tests und am Schluss ein paar kleine Ausblicke in die multivariaten Tests.

2. Schätzen von Verteilungen und von Parametern von Verteilungen (z.B. Schätzen des Erwartungswertes, der Varianz einer Verteilung). In diesem Zusammenhang werden wir ein paar Gütekriterien für Schätzer von Verteilungsparametern erwähnen.
3. Schätzen von Vertrauensintervallen für Teststatistiken: in der Testtheorie treffen wir auf Werte von Zufallsvariablen (Statistiken), welche die „tatsächlichen“ Parameter der entsprechenden Verteilungen schätzen sollen. Es ist nützlich zu wissen, in welchem Intervall sich die tatsächlichen Werte mit welcher Wahrscheinlichkeit befinden.
4. Stichprobentheorie: Schätzen von Parametern einer Grundgesamtheit mit Hilfe einer Stichprobe (Anteile, Mittelwerte: Beispiel: Meinungsumfragen, Marktforschung). Auch in diesem Zusammenhang sind wir an Vertrauensintervallen interessiert: z.B. Bei einer Marktforschungsstudie bemängeln 45% der Befragten einen bestimmten Aspekt eines Produktes. Dieser Wert ist als Wert einer Zufallsvariablen zu betrachten. Es wäre nützlich zu wissen, in welchem Intervall sich der tatsächliche Wert mit einer Wahrscheinlichkeit von 95% findet.

Wir werden in der Folge diese Grundprobleme eingehender behandeln, wobei zuerst etwas Testtheorie gemacht wird (Kapitel 19 bis und mit Kapitel 22). Anschliessend werden Vertrauensintervalle für Teststatistiken behandelt (Kapitel 23). Dann behandeln wir bivariate Tests (Teil V) und geben einen kurzen Einblick in die multiple Regression. Schliesslich behandeln wir kurz die Stichprobentheorie (Kapitel 31). Aus Zeitgründen können wir die Schätztheorie (Kapitel 30) nicht behandeln, wobei darauf hinzuweisen ist, dass wir mit Schätztheorie bereits zu tun hatten (Erwartungstreue der Statistik Mittelwert, Satz 9.5.18, S. 218; Schätzen von Parametern einer Verteilung am Beispiel der Poissonverteilung, s. S. 245; Schätzen von Verteilungsfamilien und Verteilungsfunktionen s. Kapitel 13).

Kapitel 19

Testen von Hypothesen - Grundbegriffe; Tests mit diskreten Verteilungen

Wir wollen uns das Problem des Testens von Hypothesen zuerst an einem einfachen Beispiel klarmachen:

19.1 Problemlage

Beispiel 19.1.1. *Eine Person behauptet, die Fähigkeit zu haben, Walliser Dôle von Waadtländer Gamay in einem Blindtest unterscheiden zu können. Wir überreden die Person, sich einem Test zu unterziehen. Wir nehmen z.B. folgende Anordnung vor. Wir füllen 16 Gläser mit Dôle und mit Gamay von jeweils verschiedenen Produzenten (Die entsprechenden Anteile der Gläser sollten nach Zufallsprinzip festgelegt werden, indem man z.B. mit Excel =runden(zufallszahl()*16;0) eingibt). Die Versuchsperson darf die Anteile nicht kennen. Wir setzen voraus, dass die Wahrscheinlichkeit einer zufälligen, richtigen Entscheidung weder vom Vorliegen von Gamay oder Dôle noch von der jeweils vorangehenden Degustation beeinflusst wird). Wir ordnen die Gläser nach dem Zufallsprinzip - indem wir z.B. den Gläsern der Reihe nach Zufallszahlen zuordnen und die Gläser dann der Grösse der Zufallszahlen nach auf einem Tisch anordnen. Wir fordern die Versuchsperson auf, die Gläser der Reihe nach zu degustieren und jeweils festzulegen, ob es sich um Dôle oder Gamay handelt.*

Wie gross ist die Wahrscheinlichkeit, zufälligerweise alle Gläser richtig als Dôle oder als Gamay zu klassifizieren? Die Wahrscheinlichkeit, bezüglich eines bestimmten Glases die richtige Entscheidung zu treffen, beträgt bei zufälliger Entscheidung 0.5. Somit beträgt die Wahrscheinlichkeit, alle Gläser zufälligerweise richtig zu klassifizieren $0.5^{16} = 0.000015259 = P(X = 16) = \binom{16}{16} 0.5^{16} 0.5^0$. Gelingt es der Versuchsperson, alle Gläser richtig zu klassieren, ist es zwar immer noch denkmöglich, dass sie in Tat und Wahrheit die beiden Weinsorten blind nicht unterscheiden kann und dass sie zufälligerweise die richtige Wahl getroffen hat. Die Wahrscheinlichkeit dafür ist jedoch sehr klein. Entsprechend ist das Ergebnis schlecht durch Zufall erklärbar und wir sind bereit, die Hypothese zu verwerfen, dass sie die Gläser nicht unterscheiden kann.

Wir sind aber vermutlich nicht nur bereit, die ausserordentlichen Fähigkeiten der Versuchsperson zu akzeptieren, wenn sie alle Gläser richtig klassiert, sondern auch, wenn sie die meisten Gläser richtig klassiert. So ist es z.B. auch sehr unwahrscheinlich, höchstens ein Glas zufälligerweise falsch zu raten (höchstens ein Glas falsch bedeutet: kein Glas falsch oder nur eines falsch. „Höchstens ein Glas falsch“ ist bei 16 Gläsern äquivalent mit „Mindestens 15 Gläser richtig“). Die Wahrscheinlichkeit, zufälligerweise mindestens 15 Gläser richtig zu bestimmen, ist: $P(X \geq 15) = P(X = 15) + P(X = 16) = 1 - P(X \leq 14) = 0.000259399$ mit $X \sim B(16, 0.5)$. Wir würden

bei 15 richtig bestimmten Gläsern vermutlich zum Schluss gelangen, dass das Resultat schlecht durch Zufall zu erklären ist. Wir verwerfen die Meinung, die Person habe nur geraten und wir akzeptieren deren ausserordentlichen Fähigkeiten. $\diamond$

Systematische Analyse des Beispiels

Die Ergebnismenge ist im Beispiel $\Omega = \{\text{Die Versuchsperson sagt von einem Glas Wein das richtige aus; Die Versuchsperson sagt von einem Glas Wein das falsche aus}\} = \{\text{richtig, falsch}\}$. Unter der Hypothese, dass die Versuchsperson die Weine nicht unterscheiden kann, gilt: Graph von $f_\Omega = \{(\text{falsch}, 0.5), (\text{richtig}, 0.5)\}$. Sinnvoll ist die Zufallsvariable (Graph): $X_i = \{(\text{falsch}, 0), (\text{richtig}, 1)\}$. Wenn wir derart die Summe der Werte der Zufallsvariablen bilden, ergibt sich nämlich die Anzahl der korrekten Tipps. Es gilt: Graph von $f_{X_i} = \{(0, 0.5), (1, 0.5)\}$. Wir haben 16 solcher unabhängiger Zufallsvariablen X_i auf Ω für die gilt: $X_i \sim B(1, 0.5)$. Die Anzahl der richtigen Mutmassungen ist die Summe X dieser 16 identisch verteilten, unabhängigen Bernoulli-Zufallsvariablen. Diese Summe ist eine binomialverteilte Zufallsvariable (= Statistik) $X \sim B(16, 0.5)$. Die folgende Übersicht spiegelt die Situation:

$$\Omega^{16} \xrightarrow{(X_1, \dots, X_{16})} \prod_{i=1}^{16} \mathfrak{X}_i \xrightarrow{X} \mathbb{N}_{16}$$

($\prod_{i=1}^{16} \mathfrak{X}_i$ = Kartesisches Produkt der Bilder von X_i ; $\rightarrow$ bezeichnet eine Abbildung; $\mathbb{N}_{16}$ die ersten 16 natürlichen Zahlen inklusive 0).

Wir nennen die Statistik X in diesem Zusammenhang „Teststatistik“ und den konkreten Wert, welchen die Zufallsvariable X annimmt „Testwert“.

19.2 Nullhypothese und Alternativhypothese

Wir sind im allgemeinen bereit, eine Hypothese H zu verwerfen, wenn die Wahrscheinlichkeit sehr klein ist, dass *bei der Annahme von H* das spezifische Testergebnis oder ein extremeres Ergebnis zufällig zustande kommt. Das Ergebnis ist dann ja schlecht durch Zufall zu erklären. Im Beispiel nehmen wir als Hypothese H an, die Versuchsperson könne die Weinsorten nicht unterscheiden. Unter dieser Hypothese ist die Wahrscheinlichkeit sehr klein, dass sie z.B. mindestens 15 richtige Entscheidungen trifft, d.h. 15 oder 16 richtige Entscheidungen trifft. Die Hypothese H , die wir verwerfen, wenn das Testergebnis oder ein extremeres Ergebnis bei der Annahme von H unwahrscheinlich ist, nennen wir künftig Nullhypothese (H_0). Die Negation von H_0 bezeichnen wir mit H_A (= Alternativhypothese). Im Weinbeispiel ist

H_0 : „Die Versuchsperson kann die Weingläser nicht korrekt klassifizieren“

H_A : „Die Versuchsperson kann die Weingläser korrekt klassifizieren“.

Ein Test setzt voraus, dass wir eine Zufallsvariable (= Teststatistik) mit bekannter Wahrscheinlichkeitsverteilung unter der Nullhypothese zur Verfügung haben, mit der wir die Wahrscheinlichkeit des Testergebnisses (oder eines extremeren Ergebnisses) unter der Voraussetzung der Nullhypothese beurteilen können (im Beispiel verwendeten wir die binomialverteilte Zufallsvariable X : $X \sim B(16, 0.5)$).

Bemerkung 19.2.1. Die Nullhypothese ist die Hypothese, für die wir eine Teststatistik haben, deren Verteilung wir kennen. Die Nullhypothese besagt, dass der Testwert zur bekannten Verteilung passt oder davon nicht zu stark abweichen, wobei wir die Nullhypothese immer am konkreten Fall inhaltlich ausformulieren. Wenn im obigen Beispiel die Versuchsperson nicht die Fähigkeit hat, die Weingläser korrekt zu klassifizieren, rät sie nur, und die Wahrscheinlichkeit, den Wein im Einzelversuch korrekt zu klassifizieren beträgt 0.5. Damit haben wir eine Verteilung für die Teststatistik, und die Nullhypothese ist, dass die Testperson die Weine nicht richtig klassifizieren kann. Die Alternativhypothese ist die Negation der Nullhypothese. $\diamond$

Bemerkung 19.2.2. Das Wort „Hypothese“ ist aus den griechischen Wörtern „*hypó*“ (unter, unterhalb) und *thésis* (Satz, Behauptung) zusammengesetzt. Man übersetzt es in unserem Zusammenhang am besten mit „Vermutung“, wobei solche Vermutungen nicht willkürlich sind, sondern oft wohl begründet. Alle wissenschaftlichen Gesetze sind Hypothesen, manche sind allerdings so gut bestätigt, dass wir an ihnen nicht zweifeln (z.B. Alle Metalle leiten Strom). $\diamond$

19.3 Signifikanzniveau

Traditionell legt man ein bestimmtes Niveau für die Verwerfung der Nullhypothese fest. Das Verwerfungsniveau wird *Signifikanzniveau* genannt und gewöhnlich auf $\alpha = 0.05$ oder $\alpha = 0.01$ festgelegt. Fällt der Testwert in einen Bereich, in den bei Annahme der Nullhypothese der Testwert *oder ein extremerer Wert* nur mit einer Wahrscheinlichkeit von weniger als 0.05 oder 0.01 fallen, so verwerfen wir die Nullhypothese. Gewöhnlich spricht man dann von einem *statistisch signifikanten Testwert*. Signifikanzniveaus sind jeweils vor einem Test festzulegen. Oft wird die Signifikanz von Testwerten mit einem Stern für das 0.05–Niveau, mit zwei Sternen für das 0.01–Niveau und mit drei Sternen für das 0.001–Niveau gekennzeichnet.

Legen wir im obigen Beispiel 0.05 als Signifikanzniveau fest, sind wir bereit, die Nullhypothese zu verwerfen, wenn für den Testwert x des Versuchs bei Gültigkeit der Nullhypothese gilt: $P(X \geq x) < 0.05$. Berechnen wir die entsprechende (umgekehrte), kumulative Wahrscheinlichkeitsverteilung, erhalten wir (s. Tabelle 19.3.1):

Wert der ZV x_i	$P(X = x_i)$	$P(X \geq x_i)$
16	1.52588E-05	1.52588E-05
15	0.000244141	0.000259399
14	0.001831055	0.002090454
13	0.008544922	0.010635376
12	0.027770996	0.038406372
11	0.066650391	0.105056763
10	0.122192383	0.227249146

Tabelle 19.3.1: Von oben kumulierte Werte einer binomialverteilten Zufallsvariable ($n = 16$; $p = 0.5$)

Somit sind wir bereit, die Nullhypothese zu verwerfen, wenn die Versuchsperson mindestens 12 richtige Zuordnungen vornimmt.

Die Menge der Werte, bei deren Annahme durch die Statistik die Nullhypothese verworfen wird, wird auch „Verwerfungsbereich zum Signifikanzniveau α “ genannt (kurz „ α –Verwerfungsbereich“). Im obigen Beispiel ist der 0.05-Verwerfungsbereich $\{12, 13, 14, 15, 16\}$. Statt zu sagen, dass wir die Nullhypothese verwerfen, wenn $P(X \geq x) < 0.05$, können wir ebenso gut sagen, dass wir die Nullhypothese verwerfen, wenn der Testwert x Element des Verwerfungsbereichs ist. Die Grenze des Verwerfungsbereichs nennen wir „kritische Grenze“. Im obigen Beispiel ist die kritische Grenze 12.

„signifikant“ bedeutet soviel wie „bedeutsam“ oder „wichtig“. Bei uns heisst „signifikant“ „bedeutsam, weil schlecht durch Zufall erklärbar“. Bedeutsamkeit ist damit in einem rein statistischen Sinne aufzufassen. Ob Unterschiede inhaltlich (ökonomisch) bedeutsam sind, ist ein anderes Problem. So kann eine neue Mastmethode bei Schweinen zu einem durchschnittlichen Mehrertrag von 200 g führen. Die Abweichung kann bei genügend grosser Stichprobe durchaus statistisch signifikant sein (= schlecht durch Zufall zu erklären). Sie braucht aber ökonomisch nicht bedeutsam zu sein, wenn die Einführung der neuen Methode in Hinblick auf den Ertrag zu viel kostet. Allerdings kann ein Unterschied nicht ökonomisch bedeutsam sein, wenn er statistisch nicht signifikant ist. Denn dann ist der Unterschied recht gut durch Zufall erklärbar und kann entsprechend mit hoher Wahrscheinlichkeit nicht ökonomisch relevant sein.

19.4 Zweiseitige und einseitige Tests

Im obigen Beispiel 19.1.1 können wir festlegen, dass wir die Nullhypothese verwerfen, wenn die Versuchsperson mindestens 12 richtige Entscheidungen trifft. Wir sprechen in diesem Falle von einem *rechtsseitigen Test* (s. Abbildung 19.4.1).

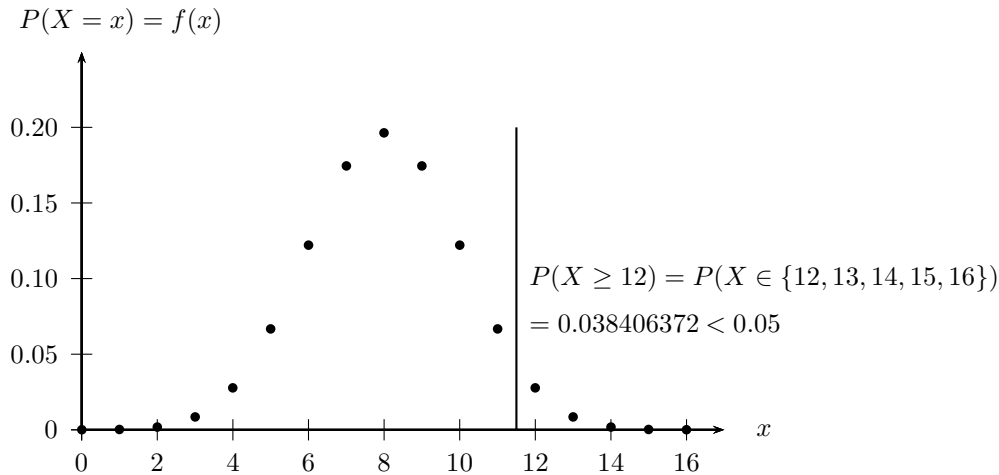


Abbildung 19.4.1: Rechtsseitiger Test: $\alpha = 0.05$. Verwerfen der Nullhypothese bei mindestens 12 richtigen Klassierungen; $P(X \geq 12) = 0.038406372 < 0.05$; 0.05-Verwerfungsbereich = $\{12, 13, 14, 15, 16\}$

Es ist auch möglich, auf Grund inhaltlicher Überlegungen den Verwerfungsbereich (der Nullhypothese) auf der linken Seite der Verteilung festzulegen.

Beispiel 19.4.1. In einer Hühnerfarm wurden neue Wärmelampen installiert. Mit den alten Wärmelampen gibt es durchschnittlich 20% Ausfälle (80% schlüpfen, 20% nicht). Es wird ein Test mit neuen Wärmelampen gemacht ($n = \text{Stichprobengrösse} = 50$). Dabei gibt es nur 8% Ausfälle. Führen die neuen Wärmelampen zu einer Reduktion der Ausfälle, die schlecht durch Zufall zu erklären sind (d.h. die neuen Wärmelampen wären mit grosser Wahrscheinlichkeit besser als die alten)? (Signifikanzniveau $\alpha = 0.05$).

H_0 : mit den neuen Wärmelampen gibt es nicht weniger Ausfälle als mit den alten.

H_A : mit den neuen Wärmelampen gibt es weniger Ausfälle als mit den alten ($\Rightarrow$ linksseitiger Test). $X \sim B(50, 0.2)$. 8% von 50 sind 4. Damit gilt es zu berechnen: $P(X \leq 4) =$

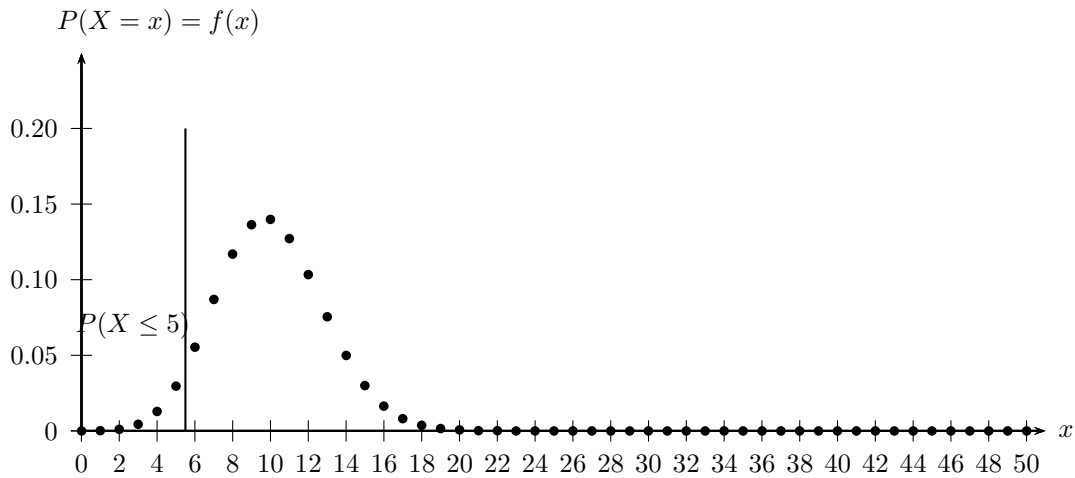
$\sum_{i=0}^4 \binom{50}{i} 0.2^i 0.8^{50-i} = 0.018496 < 0.05$. Wir können die Nullhypothese verwerfen. Die neuen Wärmelampen sind mit hoher Wahrscheinlichkeit besser (s. Abbildung 19.4.2).

Statt zu sagen, dass $P(X \leq 4) = 0.018496 < 0.05$, und damit die Nullhypothese als genügend unwahrscheinlich zu verwerfen ist, können wir auch hier sagen, dass der Testwert $4 \in \{0, 1, 2, 3, 4, 5\}$ und damit Element des Verwerfungsbereichs ist. $\diamond$

Rechtsseitige oder linksseitige Tests nennen wir „einseitige Tests“. Neben einseitigen Tests gibt es auch zweiseitige Tests: Nehmen wir an, die Person im obigen Beispiel 19.1.1 würde 16 falsche Entscheidungen treffen. Es wäre dann wohl sinnvoll anzunehmen, sie könne die Weine unterscheiden, würde diesen aber jeweils falsche Ausdrücke zuordnen. Dies würde uns zu folgenden Hypothesen führen:

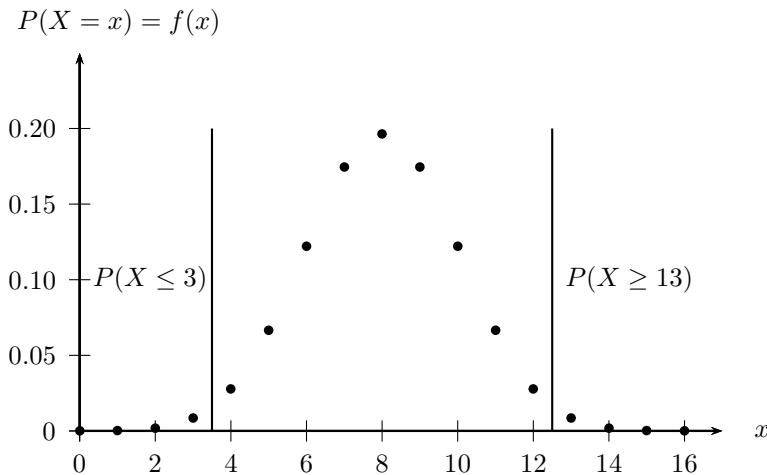
H_0 : Die Person kann die Gläser nicht unterscheiden (und klassifiziert sie nicht überwiegend falsch oder nicht überwiegend korrekt).

H_A : Die Person kann die Gläser unterscheiden (und klassifiziert sie überwiegend falsch oder dann überwiegend korrekt)



Abbildungung 19.4.2: linksseitiger Test; $P(X \leq 5) = 0.0480279 < 0.05$ für $X \sim B(50, 0.2)$; 0.05-Verwerfungsbereich = $\{0, 1, 2, 3, 4, 5\}$

Entsprechend müssten wir den Verwerfungsbereich auf beiden Seiten der Verteilung festlegen. Bei einem Signifikanzniveau von 0.05 ergibt sich auf beiden Seiten eine Wahrscheinlichkeitsgrenze von 0.025. Betrachten wir wieder unser Weindegustations-Beispiel: Bei beidseitigem Verwerfungsbereich würden wir die Nullhypothese verwerfen, wenn höchstens 3 oder mindestens 13 richtige Entscheidungen getroffen werden (bei $p = 0.5$, ist die Binomialverteilung symmetrisch. Wir können somit durch die Bestimmung einer Grenze mit Hilfe der obigen Tabelle auch die andere Grenze bestimmen). Ob beidseitig oder einseitig getestet werden soll, muss vorgängig zu einem Test auf dem Hintergrund von *inhaltlichen* Überlegungen festgelegt werden (s. Abbildung 19.4.3).



Abbildungung 19.4.3: Zweiseitiger Test: $\alpha = 0.05$. Verwerfen der Nullhypothese bei höchstens 3 oder mindestens 13 richtigen Klassierungen; $P(X \leq 3) + P(X \geq 13) = 0.010635376 + 0.010635376 = 0.02127 < 0.05$; 0.05-Verwerfungsbereich $\{0, 1, 2, 3, 13, 14, 15, 16\}$

Bemerkung 19.4.2. „Rechts-, links- und zweiseitiger Test“ bezieht sich auf die Verwerfungsbereiche und wo diese auf der x -Achse bezüglich der Dichte- oder Wahrscheinlichkeitsfunktion der Verteilung der Teststatistik liegen. Ein Test ist rechtsseitig, wenn der Verwerfungsbereich rechts das Signifikanzniveau α abschneidet. Ein Test ist linksseitig, wenn der Verwerfungsbereich links

bei der Dichte- oder der Wahrscheinlichkeitsfunktion α abschneidet. Ein Test ist zweiseitig, wenn der Verwerfungsbereich bei der Dichte- oder Wahrscheinlichkeitsfunktion sowohl rechts als auch links je $\frac{\alpha}{2}$ abschneidet (Der Sprachgebrauch ist allerdings in der Literatur nicht einheitlich. Das Statistikprogramm SPSS braucht z.B. rechts- und linksseitig nicht immer auf diese Art). $\diamond$

Bemerkung 19.4.3. Ob ein Test rechts- oder linksseitig ist, liegt oft an der Formulierung der Hypothesen. So führt die Formulierung „Es hat mit der neuen Brutmaschine mehr ausgeschlüpfte Küken“ auf einen rechtsseitigen Test und die Formulierung „Es hat mit der neuen Brutmaschinen weniger nicht ausgebrütete Eier“ auf einen linksseitigen Test. $\diamond$

19.5 p-Wert

Trifft in einem *rechtsseitigen* Test der Testwert x ein, so wird $P(X \geq x)$ „ p -Wert“ des Testwertes genannt. Das Testergebnis ist, wie gesehen, signifikant, wenn der p -Wert kleiner als das Signifikanzniveau α ist. Im *linksseitigen Fall* ist der p -Wert des Testwertes x die Wahrscheinlichkeit $P(X \leq x)$. Auch hier ist das Testergebnis signifikant, wenn der p -Wert kleiner als das Signifikanzniveau α ist.

Im *beidseitigen Test* sollte das Testergebnis ebenfalls signifikant sein, wenn die Wahrscheinlichkeit, den Testwert x oder ein extremeres Ergebnis zu haben, kleiner als α ist. Dabei kann der Testwert links oder rechts der Verteilung in einen unter der Nullhypothese unwahrscheinlichen Bereich fallen. Damit verwerfen wir die Nullhypothese genau dann, wenn

$$P(X \leq x) < \frac{\alpha}{2} \text{ oder } P(X \geq x) < \frac{\alpha}{2}$$

- oder anders formuliert - genau dann, wenn

$$\min\{P(X \leq x), P(X \geq x)\} < \frac{\alpha}{2}.$$

Um das folgende Verwerfungsprinzip möglichst einheitlich formulieren zu können, multiplizieren wir beide Seiten mit 2 und erhalten äquivalent

$$2 \cdot \min\{P(X \leq x), P(X \geq x)\} < \alpha.$$

Definieren wir den p -Wert im zweiseitigen Fall durch

$$2 \cdot \min\{P(X \leq x), P(X \geq x)\}$$

können wir für alle drei Fälle festhalten, dass wir die Nullhypothese verwerfen, wenn der p -Wert kleiner als α ist. Zusammenfassend halten wir fest:

Definition 19.5.1. Sei x der Testwert.

Bei einem rechtseitigen Test ist der p -Wert: $P(X \geq x)$

Bei einem linksseitigen Test ist der p -Wert: $P(X \leq x)$

Bei einem zweiseitigen Test ist der p -Wert: $2 \cdot \min\{P(X \leq x), P(X \geq x)\}$

Wir betrachten ein Beispiel zum zweiseitigen Fall: wir legen vorgängig zum Weinbeispiel fest, dass wir zweiseitig testen wollen. Die Versuchsperson entscheidet 10 mal richtig.

$$\begin{aligned} P(X \leq 10) &= 0.894943237 \\ P(X \geq 10) &= 1 - F(X \leq 9) = 1 - 0.77275 = 0.22725 \end{aligned}$$

Damit ist der p -Wert

$$2 \cdot \min\{0.894943237, 0.22725\} = 2 \cdot 0.22725 = 0.4545.$$

Man könnte im zweiseitigen Fall statt $2 \cdot \min\{P(X \leq x), P(X \geq x)\}$ mit α zu vergleichen, wie die obige Argumentation zeigt, auch das Signifikanzniveau halbieren und dieses direkt mit $P(X \leq x)$ und mit $P(X \geq x)$ vergleichen - in der Praxis werden wir das oft auch tun. Wir könnten dann allerdings das folgende Verwerfungsprinzip nicht so einfach formulieren.

Definition 19.5.2. Verwerfungsprinzip: *Wenn*

- (1) x eintrifft und
 - (2) p -Wert (von x) $< \alpha$ unter der Voraussetzung der Geltung von H_0
- dann verwerfen wir H_0 (α ist das Signifikanzniveau; im Allgemeinen $\alpha = 0.05$ oder $\alpha = 0.01$)

Alternativ könnte man das Verwerfungsprinzip auch mit Hilfe des Verwerfungsbereichs definieren: Wenn (1) x eintrifft und

- (2) $x \in \alpha$ -Verwerfungsbereich V_α , dann verwerfen wir H_0 (Der α -Verwerfungsbereich wird im rechtsseitigen Fall bestimmt als $V_\alpha = \{y \mid P(X \geq y) < \alpha\}$; im linksseitigen Fall als $V_\alpha = \{y \mid P(X \leq y) < \alpha\}$; im zweiseitigen Fall als $V_\alpha = \{y \mid P(X \leq y) < \frac{\alpha}{2}\} \cup \{y \mid P(X \geq y) < \frac{\alpha}{2}\}$)

Wir werden künftig im Allgemeinen die erste Variante verwenden. Diese hat den Vorteil, dass durch die Angabe des p -Wertes klar wird, ob ein Testergebnis knapp oder deutlich signifikant wird.

19.6 Fehler erster und zweiter Art

Wir nehmen an, dass wir einen Testwert erhalten, für dessen p -Wert gilt: $p\text{-Wert} < \alpha$. Es ist zwar recht unwahrscheinlich, aber trotzdem möglich, dass die Nullhypothese wahr ist. In diesem Falle verwerfen wir die Nullhypothese fälschlicherweise. Ein Verwerfen der richtigen Nullhypothese wird „*Fehler erster Art*“ genannt (Wenn man einen Test mit einem Signifikanzniveau von 0.05 und einer richtigen Nullhypothese 100 mal wiederholt, so werden wir mit hoher Wahrscheinlichkeit einige signifikante Resultate erhalten, nämlich ca. 5 auf 100. Die Anzahl der signifikanten Resultate ist dabei als Wert einer Zufallsvariable zu betrachten). Handkehrum ist es möglich, dass wir einen nicht-signifikanten p -Wert erhalten, obwohl die Nullhypothese nicht zutrifft. Diesen Fehler nennt man „*Fehler zweiter Art*“. Die Fehlerwahrscheinlichkeiten für die beiden Fehlertypen hängen vom Signifikanzniveau α ab. Wird das Signifikanzniveau bei diskret verteilten Teststatistiken genügend stark gesenkt, so sinkt die Fehlerwahrscheinlichkeit erster Art und die Fehlerwahrscheinlichkeit zweiter Art wird erhöht.

Wir betrachten dazu zuerst das Beispiel einer binomialverteilten Statistik (Weinbeispiel). Nehmen wir an, die Person könne die Weine korrekt klassifizieren, wobei die Wahrscheinlichkeit einer korrekten Klassifikation 0.9 betrage. Die tatsächliche Verteilung kennen wir i.A. nicht. Wir nehmen hier einfach eine an, um Überlegungen zum Fehler 1. und 2. Art anzustellen. Tragen wir die Wahrscheinlichkeitsfunktionen der Teststatistik unter der Nullhypothese und der Hypothese ab, dass die Person mit $p = 0.9$ korrekt klassifiziert, in einer Graphik ab, so erhalten wir (s. Abbildung 19.6.4)

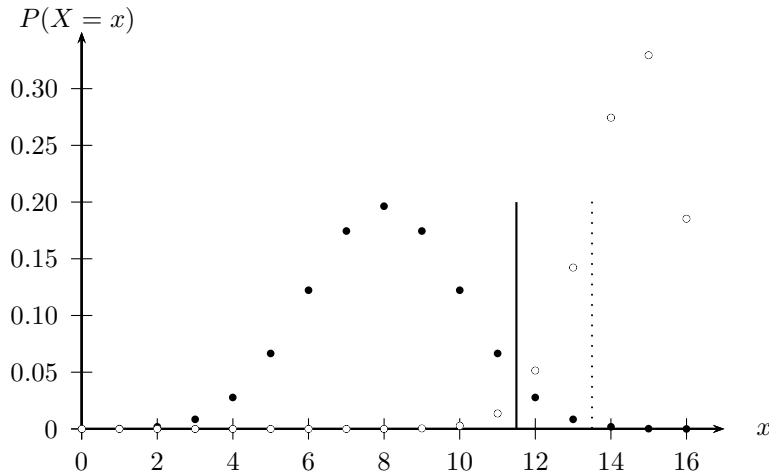


Abbildung 19.6.4: Fehler 1. und 2. Art für einem rechtsseitigen Test bei $\alpha = 0.05$ und $\alpha = 0.01$. Unter der Nullhypothese ist die Statistik $B(16, 0.5)$ verteilt (fette Punkte). Die tatsächliche Wahrscheinlichkeitsfunktion sei durch $B(16, 0.9)$ gegeben (nicht ausgefüllte Punkte).

Die Verkleinerung des Signifikanzniveaus (von 0.05 auf 0.01) bewirkt eine Verminderung der Fehlerwahrscheinlichkeit erster Art und eine Vergrößerung der Fehlerwahrscheinlichkeit zweiter Art. Bei $\alpha = 0.05$ (durchgezogene Linie) beträgt die Fehlerwahrscheinlichkeit erster Art $P(X \geq 12) = 0.038406372$ für $X \sim B(16; 0.5)$ und die Fehlerwahrscheinlichkeit zweiter Art $P(X \leq 11) = 0.017003998$ für $X \sim B(16; 0.9)$. Bei $\alpha = 0.01$ (punktierter Linie) beträgt die Fehlerwahrscheinlichkeit erster Art $P(X \geq 14) = 0.002090454$ für $X \sim B(16; 0.5)$ und die Fehlerwahrscheinlichkeit zweiter Art $P(X \leq 13) = 0.21075066$ für $X \sim B(16; 0.9)$.

Bei stetig verteilten Teststatistiken beträgt die Fehlerwahrscheinlichkeit erster Art genau α . In der folgenden Graphik betrachten wir eine standardnormalverteilte Teststatistik (dargestellt durch die Dichtefunktion der Zufallsvariable unter der Nullhypothese). Daneben betrachten wir eine $X \sim N(3, 1)$ verteilte Zufallsvariable (dargestellt durch die Dichtefunktion der tatsächlichen Verteilung der untersuchten Grösse; die tatsächliche Verteilung kennen wir auch hier gewöhnlich nicht. Wir nehmen sie einfach an; s. Abbildungen 19.6.5 und 19.6.6). Man sieht, dass die Reduktion der Fehlerwahrscheinlichkeit 1. Art nur durch eine Erhöhung der Fehlerwahrscheinlichkeit 2. Art erkauft werden kann - und umgekehrt.

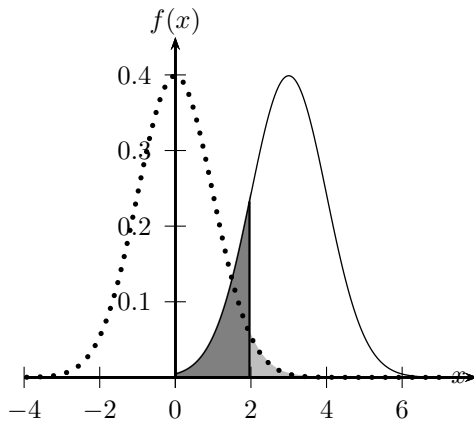


Abbildung 19.6.5: gepunktete Dichtefunktion = Dichtefunktion unter der Verteilung der Nullhypothese; durchgezogene Dichtefunktion = Dichtefunktion der „wirklichen“ Verteilung; rechtsseitiger Test mit Verwerfungsbereich = $]1.96, \infty[$; hellere graue Fläche = Fehlerwahrscheinlichkeit 1. Art (= 0.025); dunklere graue Fläche = Fehlerwahrscheinlichkeit 2. Art (= 0.14916995)

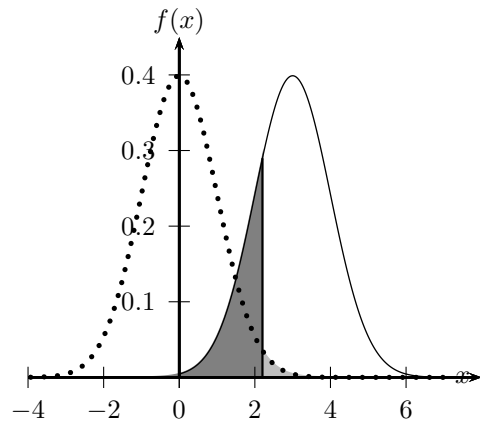


Abbildung 19.6.6: gepunktete Dichtefunktion = Dichtefunktion unter der Verteilung der Nullhypothese; durchgezogene Dichtefunktion = Dichtefunktion der „wirklichen“ Verteilung; rechtsseitiger Test mit Verwerfungsbereich = $]2.2, \infty[$; hellere graue Fläche = Fehlerwahrscheinlichkeit 1. Art (= 0.014); dunklere graue Fläche = Fehlerwahrscheinlichkeit 2. Art (= 0.211855399)

Auf Grund von aussermathematischen Überlegungen zu den beiden Fehlertypen kann man die Festlegung des Signifikanzniveaus unter Umständen begründen, indem man die Kosten beim Auftreten der beiden Fehler analysiert. Wir betrachten als Beispiel die Verwendung eines Medikamentes mit starken, negativen Nebenwirkungen. Führt uns ein Test fälschlicherweise dazu, die Wirksamkeit des Medikamentes zu postulieren, so werden Patienten den Nebenwirkungen ausgesetzt, ohne dass das Medikament wirkt. Dies sollte man vermeiden. Entsprechend muss der Fehlerwahrscheinlichkeit erster Art reduziert werden und das Signifikanzniveau ist klein anzusetzen (kleines Alpha). Hat ein Medikament demgegenüber keine bekannten Nebenwirkungen, ist es weniger wichtig, die Verabreichung des Medikamentes an eine möglichst klar bestätigte Wirksamkeit zu binden. Das Signifikanzniveau kann höher liegen (Alpha kann grösser sein).

Beispiel 19.6.1. *Weiteres Beispiel für Fehler erster und zweiter Art: In einer Population sei der Anteil von Personen, die eine Schokolade gut finden 0.4. Die Firma verändert die Rezeptur, um zu sehen, ob der Anteil an Personen, welche die Schokolade gut finden, gesteigert werden kann. Es wird eine Zufallsstichprobe (Stichprobengrösse $n = 100$) gezogen, um die Anteilsentwicklung zu überprüfen. Ein Fehler erster Art liegt vor, wenn sich der Anteil an Personen, welche die Schokolade mit neuer Rezeptur gut finden, in der Population nicht ändert, die Firma aber zufälliger Weise einen Stichprobe gezogen hat, welche z.B. einen Anteil von 0.5 solcher Personen aufweist. Bei einem Signifikanzniveau von 0.05 wird die Nullhypothese, dass sich der Anteil nicht erhöht hat, nämlich in diesem Fall verworfen (es gilt $P(X \geq 50) = 0.027099198 < 0.05$ mit $X \sim B(100, 0.4)$). Ein Fehler zweiter Art liegt vor, wenn sich der Anteil in der Population z.B. auf 0.5 erhöht hat, die Firma aber zufälliger Weise eine Stichprobe zieht, in der der Anteil 0.4 ist. Die Nullhypothese wird nicht verworfen, obwohl man sie hätte verwerfen sollen.* $\diamond$

Beispiel 19.6.2. *Noch ein ökonomisches Beispiel für die Problematik der Festlegung des Signifikanzniveaus: Wir nehmen an, eine Warenlieferung werde mit Hilfe eines Testes auf die Erfüllung der minimalen, vertraglich abgemachten Qualitätsstandards untersucht. Laut diesen dürfen höchstens 5% der Produkte einer Warenlieferung defekt sein. Bei einer Warenlieferung von 200 Produkten dürfen entsprechend nur 10 Produkte defekt sein. Da eine Untersuchung aller Produkte zu kostspielig ist, ziehen wir eine Zufallsstichprobe von 30 Produkten, um diese auf ihren Zustand*

zu untersuchen. Da die untersuchte Gesamtheit relativ klein ist und wir defekte Objekte nicht zurücklegen möchten, verwenden wir als Teststatistik eine hypergeometrisch verteilte Zufallsvariable. Wir nehmen an, in der Stichprobe befänden sich 5 defekte Produkte. Erfüllt die Lieferung die Qualitätsanforderungen?

H_0 : Die Lieferung weist nicht mehr defekte Objekte auf als erlaubt.

H_A : Die Lieferung weist mehr defekte Objekte auf als erlaubt (rechtsseitiger Test).

Wie ist das Signifikanzniveau festzulegen? Bei einem kleinen α wird die Wahrscheinlichkeit, eine Warenlieferung zu akzeptieren, die den Qualitätsstandards nicht genügt, vergrößert (= Wahrscheinlichkeit, Fehler zweiter Art zu begehen, wird vergrößert). Dies ist nicht im Interesse des Beliefernden. Andererseits wird dadurch die Wahrscheinlichkeit, eine Warenlieferung zurückzuweisen, die den Ansprüchen genügt, verkleinert (= Wahrscheinlichkeit, Fehler erster Art zu begehen, wird verkleinert). Dies liegt im Interesse des Lieferanten. Die Festlegung des Signifikanzniveaus liegt damit im Schnittpunkt widerstrebender Interessen. Es liegt an den Vertragspartnern, den möglichen Schaden von Fehleinschätzungen überzeugend ins Feld zu führen, um ein für sie möglichst günstiges Signifikanzniveau auszuhandeln.

Um ein weiteres Beispiel für die Durchführung eines Tests zu haben, führen wir das eben eingeführte Beispiel weiter. Wir nehmen an, die Vertragspartner hätten sich auf ein Signifikanzniveau von 0.01 geeinigt. Wir berechnen $P(X \geq 5)$ mit $X \sim H(200, 10, 30)$. $P(X \geq 5) = \sum_{i=5}^{10} \frac{\binom{190}{30-i} \binom{10}{i}}{\binom{200}{30}} = 0.0080426 < 0.01$. Die Nullhypothese kann verworfen werden. Es ist unwahrscheinlich, dass unter der Nullhypothese mindestens 5 defekte Objekte in der Stichprobe auftauchen. Die Lieferung entspricht den Qualitätsansprüchen nicht. $\diamond$

Bemerkung 19.6.3. Auch im Beispiel 19.6.2 kann man die Situation mit einer Übersicht der folgenden Art darstellen, da man eine hypergeometrisch verteilte Zufallsvariable als Summe von bernoulli-verteilten Zufallsvariablen betrachten kann (s. Seite 252).

$$\Omega^{30} \xrightarrow{(X_1, \dots, X_{30})} \prod_{i=1}^{30} \mathfrak{X}_i \xrightarrow{X} \mathbb{N}_{10}$$

$\diamond$

19.7 Macht eines Testes

Als Macht, Güte oder Trennschärfe eines Testes wird die Wahrscheinlichkeit bezeichnet, die falsche Nullhypothese zu verwerfen. Die Macht des Testes ist also die Wahrscheinlichkeit, eine korrekte Alternativhypothese zu akzeptieren. Es handelt sich um die Gegenwahrscheinlichkeit zur Fehlerwahrscheinlichkeit zweiter Art. Beträgt die Fehlerwahrscheinlichkeit zweiter Art z.B. 0.2, beträgt die Macht des Testes 0.8.

Es folgt ein Beispiel für einen Test, der auf einer poissonverteilten Teststatistik beruht:

Beispiel 19.7.1. Einer Firma wird vorgeworfen, die Asbest-Grenzwerte in ihren Hallen nicht einzuhalten. Man einigt sich auf folgenden Test: Es werden 10 mal jeweils ein Liter Luft untersucht und die Anzahl Asbestfasern gezählt. Der offizielle Grenzwert liegt bei höchstens 3 Fasern pro Liter. Die Nullhypothese ist: „Die tatsächliche Anzahl liegt nicht über dem Grenzwert“. Die Alternativhypothese ist: „Die Anzahl liegt über dem Grenzwert“ ($\implies$ rechtsseitiger Test). Wir erhalten für die Proben folgende Stichprobenresultate: 1; 3; 5; 7; 2; 6; 4; 3; 2; 1 (Anzahl Fasern pro Liter). Unter der Nullhypothese können wir von 10 poisson-verteilten Zufallsvariablen X_i ausgehen, mit $E(X_i) = 3$. Die Summe dieser 10 unabhängigen und identisch verteilten Zufallsvariablen ist dann $X \sim P(30)$. (siehe Sätze zur Summe von Poissonverteilungen im Kapitel „Poissonverteilung“; Ergebnismenge Ω = mögliche Anzahl von Fasern, abzählbar unendlich viele Elemente!; $\mathfrak{X}_i = \mathbb{N}$). Wir gelangen also wieder zur Übersicht:

$$\Omega^{10} \xrightarrow{(X_1, \dots, X_{10})} \prod_{i=1}^{10} \mathfrak{X}_i \xrightarrow{X} \mathbb{N}$$

Die Teststatistik X nimmt im vorliegenden Fall den Testwert $1 + 3 + 5 + 7 + 2 + 6 + 4 + 3 + 2 + 1 = 34$ an. Durchschnittlich liegt das Resultat somit über dem Grenzwert. Es stellt sich die Frage, ob der Testwert die Nullhypothese widerlegt. Der p -Wert ist $P(X \geq 34)$. Wir berechnen $P(X \leq 33) = 0.74445$ für $X \sim P(30)$. Der p -Wert ist damit $1 - 0.74445 = 0.25555 = P(X > 33) = P(X \geq 34) \geq 0.05$ und er liegt über dem Signifikanzniveau. Die Nullhypothese kann nicht verworfen werden (Da $P(X \leq 38) = 0.93516$ und $P(X \leq 39) = 0.953746962$ für $X \sim P(30)$, müsste somit die Nullhypothese erst bei 40 und mehr Fasern verworfen werden ($\alpha = 0.05$)) $\diamond$

Die eingeführten Grundbegriffe (einseitiger und zweiseitiger Test, Verwerfungsniveau = Signifikanzniveau, Verwerfungsbereich, Nullhypothese, Alternativhypothese, p -Wert, Teststatistik, Testwert, Stichprobe, Fehler erster und zweiter Art) werden auch für die später behandelten Tests, die eine stetige Verteilung voraussetzen, verwendet. Es ist wichtig, diese Begriffe verstanden zu haben, da sie in der Folge immer wieder auftauchen. Entsprechend sollten sie wiederholt werden, sobald an weiteren Beispielen die Vorgehensweise des Testens deutlicher geworden ist.

19.7.1 Übungen

1. Eine Unternehmung fertigt mit einer Maschine täglich elektronische Bauteile. Darunter fallen durchschnittlich 3% defekte an. Es wird eine neue Maschine getestet. Aus der Produktion von 500 Objekten der neuen Maschine wird eine Zufallsstichprobe von 50 Produkten gezogen. Es befinden sich 2 defekte darunter. Ist die neue Maschine besser? ($\alpha = 0.05$)
2. Eine Unternehmung fertigt mit einer Maschine täglich elektronische Bauteile. Darunter fallen durchschnittlich 4% defekte an. Es wird eine neue Maschine getestet. Aus der Produktion der neuen Maschine wird eine Zufallsstichprobe von 75 Produkten gezogen. Es befindet sich 1 defektes darunter. Ist die neue Maschine besser? ($\alpha = 0.05$)
3. Eine Firma wird von der Gewerkschaft beschuldigt, weniger als durchschnittlich für die Sicherheit ihrer Mitarbeiter vorzukehren, da die Unfallquote im Jahre 2000 10 auf 1000 betrug, während sie in vergleichbaren Firmen nur 5 auf 1000 betrug. Kann die vorliegende Unfallhäufigkeit auch durch Zufall erklärt werden? (Signifikanzniveau 0.05)
4. Bei einer Lieferung von Rohstoffteilen dürfen laut Liefervertrag durchschnittlich nur 0.5% mangelhaft sein. Dabei wird der folgende Test verwendet, um die Einhaltung der Lieferbedingungen zu überprüfen: Es werden 20 Teile zufällig gezogen (ohne Zurücklegen). Wir nehmen an, bei einer solchen Stichprobe und einer Lieferung von 200 Teilen seien 2 Bauteile fehlerhaft gewesen. Entspricht die Lieferung den Lieferbedingungen? ($\alpha = 0.01$)
5. Bei einer Lieferung von Zwischenprodukten dürften laut Liefervertrag durchschnittlich nur 1% mangelhaft sein. Dabei wird der folgende Test verwendet, um die Einhaltung der Lieferbedingungen zu überprüfen: Es werden 20 Teile zufällig gezogen (ohne Zurücklegen). Wir nehmen an, bei einer solchen Stichprobe und einer Lieferung von 500 Teilen seien 2 Bauteile fehlerhaft gewesen.
 - a) Entspricht die Lieferung den Lieferbedingungen? ($\alpha = 0.05$)
 - b) Welche Interessen kommen bei der Festlegung des Signifikanzniveaus ins Spiel? (Ist der Lieferant oder der Belieferte an einem möglichst hohen Signifikanzniveau interessiert? Welche Gewinne und Verluste fahren die beiden Beteiligten beim Vorliegen eines Fehlers erster oder zweiter Ordnung ein?).
6. In einer Zeitung steht, die Unfälle mit Todesfolge hätten in der Schweiz massiv zugenommen, da sie von 1000 auf 1100 gestiegen sind. Testen Sie!

7. Eine Person gibt vor, hellseherische Kräfte zu haben. Insbesondere will sie voraussagen können, ob sich bei einem Münzwurf Kopf oder Zahl ergibt. Denken Sie sich einen statistischen Test aus, um diese Behauptung zu überprüfen (Legen sie dabei H_0 und α fest, sowie ob einseitig oder zweiseitig getestet werden soll).
8. Eine Firma beschwert sich bei einem Lieferanten über unterschiedliche Qualitäten von Lieferungen. Dabei könnten von Auge deutlich zwei Qualitäten unterschieden werden. Der Lieferant bestreitet die Behauptung. Man einigt sich auf folgenden Test: 10 Personen untersuchen je 10 Stichproben, die aus den (angeblich oder wirklich) unterschiedlichen Qualitäten stammen. Als Signifikanzniveau wird 0.01 festgelegt. Man testet beidseitig. Legen Sie die Nullhypothese und die Alternativhypothese fest und berechnen x_i , so dass der p-Wert von $x_i \leq 0.01$.
9. Im Boulevard-Gratis-Blatt „20 minuten“ steht am 11. November 2003 auf der ersten Seite der Titel „Jugendkriminalität stieg um 4 Prozent“. Unter diesem Titel werden folgende Zahlen geliefert: Im Jahr 2002 sind 13'000 Minderjährige wegen Straftaten verurteilt worden. Das zeigt eine vom Bundesamt für Statistik am 10. November 2003 veröffentlichte Erhebung. Seit der Statistik 1999 haben die Verurteilungen damit um rund 500 Fälle oder vier Prozent zugenommen. Ist der Unterschied der beiden Jahre signifikant? ($\alpha = 0.05$).
10. (Beispiel aus dem Tages-Anzeiger, 29. Juli 04, S. 26). Bei der Untersuchung von Invaliden-Renten-Aspiranten wird unter anderem auch Statistik eingesetzt. Der Neuropsychologe Thomas Merten untersuchte einen 22-Jährigen, der vorgab, nach einem Schädel-Hirn-Trauma nicht mehr rechnen zu können. Merten liess den Patienten 40 einfache Rechenaufgaben machen, z.B. $3 + 3$ und bot ihm jeweils zwei Lösungen an: die richtige (im Beispiel 6) und eine falsche, die nahe bei der richtigen lag (im Beispiel etwa 7). Der Patient entschied sich 11 mal für die richtige Lösung, 29 mal für die falsche. Merten schliesst aus dem Ergebnis, dass der Patient simuliert, denn der Begutachtete wählte mehr Falsche als eine Person, die wirklich nicht rechnen kann und damit zufällig zwischen wahr und falsch unterscheidet. Führen Sie den Test durch. ($\alpha = 0.05$)
11. Kassasturz SF, 6. November 2012. Ein Pendler, der behauptet, ein mit seinen geheimen Methoden von magnetischen Wellen befreites Handy unter anderen Handis mit Hilfe von Pendeln identifizieren zu können, macht in der Sendung folgenden Test: 10 mal hat er die Möglichkeit, Handis, die unter 10 weissen Plastikschrüsseln liegen, zu bependeln. Er weiss, dass es in jeder Testserie genau ein gemäss seiner Methode von magnetischen Wellen befreites Handy hat. Er hat 3 Treffer. Wie hoch ist die Wahrscheinlichkeit, mindestens 3 Treffer zu haben, unter der Nullhypothese, dass er das strahlenfreie Handy nicht erpendeln kann? Das Schweizer Fernsehen gibt an, dem Pendler eine faire Chance zu geben, seine Fähigkeit unter Beweis zu stellen: wenn er 8 richtige Treffer habe, dann würde ihm die Preissumme von 10 000 Franken zugestanden. Wie hoch ist die Wahrscheinlichkeit, unter der Nullhypothese mindestens 8 richtige Treffer zu haben? Wie beurteilen Sie die Testanlage des SF?
12. Denken Sie sich je eine betriebswirtschaftlich relevante Problemlage aus, die Sie mit Hilfe eines Binomial- und eines Poisson-Testes lösen können. Erfinden Sie eine solche Problemlage auch für einen hypergeometrischen Test.

19.7.2 Lösungen

1. H_0 : Die neue Maschine produziert nicht weniger Ausschuss.
 H_A : Die neue Maschine produziert weniger Ausschuss ($\implies$ linksseitiger Test). Als Teststatistik verwenden wir eine hypergeometrisch verteilte Zufallsvariable X mit $X \sim H(500, 50, 15)$ (denn die untersuchte Gesamtheit ist klein und wir möchten defekte Objekte nicht zurücklegen. 3% von 500 sind 15); $P(X \leq 2) = \sum_{i=0}^2 \frac{\binom{485}{50-i} \binom{15}{i}}{\binom{500}{50}} = 0.81832 > 0.05$. Die neue Maschine ist

nicht besser (Dies ist auch nicht verwunderlich: 3% von 50 sind $0.03 \cdot 50 = 1.5$. Es hat somit sogar mehr defekte in der neuen Produktion als im Durchschnitt bei der alten Maschine).

2. H_0 : Die neue Maschine produziert nicht weniger Ausschuss.

H_A : Die neue Maschine produziert weniger Ausschuss ($\implies$ linksseitiger Test). Als Teststatistik verwenden wir eine binomialverteilte Zufallsvariable, da die Grösse der untersuchten Gesamtheit (Neuproduktion) nicht bekannt ist. Ist diese klein, können wir gezogene Stücke zurücklegen, wobei wir defekte jeweils kennzeichnen können, damit wir sie nachher entfernen können. $X \sim B(75, 0.04)$. Wir berechnen $P(X \leq 1) = \sum_{i=0}^1 \binom{75}{i} 0.04^i (1 - 0.04)^{75-i} = 0.19309 > 0.05$. Die neue Maschine ist nicht besser.

3. H_0 : Die Unfallquote der Firma ist nicht grösser als die durchschnittliche Unfallquote. H_A : Die Unfallquote der Firma ist grösser als die durchschnittliche Unfallquote ($\implies$ rechtsseitiger Test).

Wir verwenden als Teststatistik eine poisson-verteilte Zufallsvariable. Wir berechnen somit $P(X \geq 10)$ für $X \sim P(5)$. $P(X \geq 10) = 1 - P(X < 10) = 1 - 0.968171943 = 0.031828 < 0.05$. Somit ist die Abweichung von der durchschnittlichen Unfallquote als signifikant zu bezeichnen. Die Firma sollte etwas für den Unfallschutz unternehmen.

4. H_0 : In der Lieferung hat es nicht mehr defekte Objekte als erlaubt.

H_A : In der Lieferung hat es mehr defekte Objekte als erlaubt ($\implies$ rechtsseitiger Test)
 $\alpha = 0.01$

Verzichten wir auf Zurücklegen, wird die Hypergeometrische Verteilung angewendet. Da 0.5% von 200 1 ist, kann es in keiner Stichprobe 2 haben. Somit ist $P(X \geq 2) = 0$. Die Nullhypothese muss verworfen werden. Verwenden wir mit Zurücklegen einen Binomialtest erhalten wir folgendes Resultat:

Wir berechnen: $P(X \geq 2) = 1 - P(X < 2)$ für $X \sim B(20, 0.005)$ und erhalten $1 - 0.995526106 = 0.004473894 < 0.01$.

Die Nullhypothese ist widerlegt.

5. H_0 : Die Lieferung enthält nicht mehr defekte Zwischenprodukte als erlaubt.

H_A : Die Lieferung enthält mehr defekte Zwischenprodukte als erlaubt (rechtsseitiger Test) 1% auf 500 macht 5. Wir verwenden einen Hypergeometrischen Test (kleine untersuchte Gesamtheit!!!). Wir berechnen $P(X \geq 2)$ für $X \sim H(N, n, M) = H(500, 20, 5)$ (N für Anzahl untersuchte Gesamtheit; M für Anzahl Defekte in der Gesamtheit; n = Stichprobengrösse; die Reihenfolge der Angabe der Parameter kann von Buch zu Buch verschieden sein. Bei Excel wird z.B. eine andere Reihenfolge gewählt).

Werte x_i der ZV X	$P(X = x_i)$
$X \sim H(500, 20, 5)$	
2	0.013634848
3	0.000513446
4	9.11126E-06
5	6.07417E-08
$P(X \geq 2) = \sum_{i=2}^5 P(X_i = x_i) =$	0.014157466

Tabelle 19.7.2: Berechnung des p -Wertes bei der Hypergeometrischen Verteilung; im Beispiel $X \sim H(500, 20, 5)$

Da $0.014157466 < 0.05$ ist die Abweichung signifikant. Die Lieferung entspricht den Lieferbedingungen nicht.

6. H_o : Die Unfallzahlen sind im zweiten Jahr nicht höher. H_A : Die Unfallzahlen im zweiten Jahr sind höher ($\implies$ rechtsseitiger Test). Da sowohl 1000 als auch 1100 als Werte einer Zufallsvariablen zu betrachten sind, wäre es hier ungünstig, so wie im Beispiel 3 vorzugehen. Wir können jedoch die Situation so deuten, dass unter der Nullhypothese ein Unfall mit gleicher Wahrscheinlichkeit im ersten Jahr erfolgt oder im zweiten Jahr. Laut Nullhypothese ist dann die Wahrscheinlichkeit, dass ein Unfall im ersten Jahr erfolgt = 0.5. Wir verwenden entsprechend eine Binomialverteilung: Wir berechnen $P(X \geq 1100)$ für $X \sim B(2100, 0.5)$. Mit der Näherungsformel für die Normalverteilung erhalten wir: $1 - \Phi\left(\frac{1100 - (2100 \cdot 0.5)}{\sqrt{2100 \cdot 0.5 \cdot 0.5}}\right) = 1 - \Phi(2.1822) = 1 - 0.985452664 = 0.014547 < 0.05$.
Somit ist das Resultat auf einem Signifikanzniveau von 0.05 signifikant, auf einem Niveau von 0.01 nicht signifikant. Mit der Binomialverteilung erhält man: 0.015360354 (=1-BINOMVERT(1099;2100;0.5;1))
7. H_o : "Die Testperson kann nicht mehrheitlich richtige Voraussagen machen".
 H_A : Die Testperson kann mehrheitlich richtige Voraussagen machen ($\implies$ rechtsseitiger Test). (Man könnte ähnlich wie oben auch für einen zweiseitigen Test argumentieren: wenn die Person systematisch falsch liegt, scheint sie hellseherische Fähigkeiten zu haben. Man könnte die Person in diesem Falle durchaus bei entsprechenden Spielen einsetzen, um oft zu gewinnen. Man müsste jeweils auf das Gegenteil dessen setzen, das sie voraussagt. In diesem Fall würden wir die Alternativhypothese wie folgt formulieren: Die Testperson kann mehrheitlich richtige oder mehrheitlich falsche Voraussagen machen). Als Signifikanzniveau legen wir 0.05 fest. Wir könnten z.B. die folgende Versuchsanordnung festlegen. Wir werfen 20 Münzen. Die entsprechenden Ergebnisse (= Summen der richtig Vorausgesagten) sind binomialverteilt ($B(20, 0.5)$). Wir testen rechtseitig. Mit Hilfe von Excel erhalten wir (s. Tabelle 19.7.3):

x_i	$P(X = x_i)$	$P(X \geq x_i)$
20	9.53674E-07	9.53674E-07
19	1.90735E-05	2.00272E-05
18	0.000181198	0.000201225
17	0.001087189	0.001288414
16	0.004620552	0.005908966
15	0.014785767	0.020694733
14	0.036964417	0.057659149
13	0.073928833	0.131587982

Tabelle 19.7.3: Berechnung des 0.05-Verwerfungsbereich bei einer binomialverteilten Statistik; $X \sim B(20, 0.5)$; Verwerfungsbereich = $\{15, 16, 17, 18, 19, 20\}$

Beim festgelegten Signifikanzniveau wird die Nullhypothese verworfen, wenn die Testperson mindestens 15 richtige Voraussagen macht (da $P(X \geq 15) = 0.020694733 < 0.05$).

8. H_o : Die zwei Qualitäten können nicht auseinandergehalten werden.
 H_A : Die zwei Qualitäten können auseinandergehalten werden, d.h. überwiegend richtig oder überwiegend falsch klassifiziert werden ($\implies$ zweiseitiger Test).
Man wählt eine spezifische Anzahl m von Stichproben aus der ersten Qualität, und die restlichen $100 - m$ aus der anderen Qualität. Wir ordnen den Stichproben Zufallszahlen zu und ordnen sie der Grösse nach. Je zehn werden dann den Versuchspersonen zugeteilt. Wir können das ganze als einen Binomial-Test mit $X \sim B(100, 0.5)$ ansehen. Mit Excel erhalten wir: mindestens 64 müssen richtig klassiert werden oder höchstens $100 - 64 = 36$ (denn =BINOM.VERT(36;100;0.5;1)=0.00331856; $2 \cdot 0.00331856 = 0.0066371 < 0.01$, während

$=\text{BINOM.VERT}(37;100;0.5;1) = 0.006016488$ und $2 \cdot 0.006016488 = 0.012033 \geq 0.01$). (s. Tabelle 19.7.4)

x_i	$P(X = x_i)$	$P(X \geq x_i)$
66	0.000458105	0.000894965
65	0.000863856	0.001758821
64	0.001559739	0.00331856
63	0.002697928	0.006016488
62	0.00447288	0.010489368
61	0.007110732	0.0176001
60	0.010843867	0.028443967
59	0.015869073	0.04431304
		$P(X \leq x_i)$
34		0.000894965
35		0.001758821
36		0.00331856
37		0.006016488
38		0.010489368
39		0.0176001
40		0.028443967
41		0.04431304

Tabelle 19.7.4: Berechnung des 0.01-Verwerfungsbereich bei einer binomialverteilten Statistik; $X \sim B(100, 0.5)$; $P(X \leq 36) + P(X \geq 64) = 2 \cdot 0.00331856 = 0.006637121 < 0.01$; $P(X \leq 37) + P(X \geq 63) = 2 \cdot 0.006016488 = 0.012032976 > 0.01$; Verwerfungsbereich = $\{0, \dots, 36, 64, \dots, 100\}$

Wir könnten auch die Normalverteilung verwenden: $\frac{x-(100 \cdot 0.5)}{\sqrt{100 \cdot 0.5^2}} = 0.2x - 10 \sim N(0, 1)$.

$\Phi(0.2x - 10) = 0.995$, d.h. $\Phi^{-1}(0.995) = 0.2x - 10$. Wir bestimmen mit Hilfe von Excel $\Phi^{-1}(0.995)$ (mit norminv): 2.575834515. Wir erhalten somit: $0.2x - 10 = 2.575834515$, $\Rightarrow x = 62.879$. Wir müssten somit mindestens 63 richtige oder höchstens $100 - 63 = 37$ richtige Klassierungen haben - es liegt also eine kleine Abweichung vom exakten Resultat vor.

9. H_0 : Es hat nicht mehr Straftaten als vorher.

H_A : Es hat mehr Straftaten als vorher (rechtsseitiger Test).

Verteilung der Teststatistik: $X \sim B(12500 + 13000; 0.5)$. Den p-Wert können wir hier mit der Näherungsformel berechnen. p-Wert: $P(X \geq 13000) \approx 1 - \Phi\left(\frac{13000 - np}{\sqrt{np(1-p)}}\right) =$

$$1 - \Phi\left(\frac{13000 - 12750}{79.84359711}\right) = 1 - \Phi(3.131121455)$$

$= 0.000870768 < 0.05$. Die Entwicklung lässt sich schlecht durch Zufall erklären.

Exakt: $P(X \geq 13000) = 1 - \text{binom.vert}(12999, 12500 + 13000, 0.5, 1) = 0.00088918$

10. Bei jeder der 40 Aufgaben kann der Begutachtete zwischen falsch und wahr unterscheiden. H_0 : Der Begutachtete kann nicht zwischen wahr und falsch unterscheiden. H_A : Der Begutachtete kann zwischen wahr und falsch unterscheiden, wobei er oft das falsche wählen wird, weil er seine mathematische Unfähigkeit beweisen will (linksseitiger Test; er wählt weniger korrekte als jemand der zwischen wahr und falsch nicht unterscheiden kann). Wir berechnen $P(X \leq 11)$ für $X \sim B(40, 0.5)$. Wir erhalten: $P(X \leq 11) = \sum_{i=0}^{11} \binom{40}{i} 0.5^i (1 - 0.5)^{40-i} = 3.2133 \times 10^{-3} < 0.05$. Wir können die Nullhypothese verwerfen. Der Begutachtete kann zwischen falsch und wahr unterscheiden.

11. Die Teststatistik ist binomialverteilt mit $n = 10$ und $p = 0.1$. Damit gilt $P(X \geq 3) = 1 - P(X \leq 2) = 0.070190826$. Auf dem 0.05-Signifikanzniveau ist das Ergebnis fast signifikant. $P(X \geq 8) = 1 - P(X \leq 7) = 3.736 \cdot 10^{-7}$. Es handelt sich um das implizite Signifikanzniveau, das dem Test des Schweizerfernsehens zu Grunde liegt. Es ist wohl unvernünftig tief angesetzt.
12. siehe als Beispiele früherer Jahrgänge auf der Home-Page <http://math.logik.ch> „Beispiele_diskrete_Tests.pdf“.

19.8 Lernziele

- Grundbegriffe der Testtheorie kennen und korrekt verwenden können (Nullhypothese, Alternativhypothese, Signifikanzniveau, rechts- links- und beidseitiger Test, p-Wert, Fehler 1. und 2. Art, Teststatistik, Testwert, Verwerfungsbereich).
- Verwerfungsprinzip anwenden können.
- Tests mit diskreten Verteilungen durchführen können (insbesondere Binomial- und Poissonverteilung, sowie hypergeometrische Verteilung).
- Übungen von der Art des Übungsblocks lösen können.

Kapitel 20

Einstichprobentests

Nachdem wir die Grundgedanken des statistischen Testens mit Hilfe diskret verteilter Teststatistiken eingeführt haben, werden wir nun für verschiedene Problemlagen Tests etwas systematischer einführen. Wir werden dabei stetig wie diskret verteilte Teststatistiken verwenden.

In Einstichprobentests vergleichen wir gewöhnlich Kennwerte einer Stichprobe (z.B. Mittelwert, Median, Varianz, Anteile) mit vorgegebenen Parametern, um zu untersuchen, ob die Abweichung der Stichprobenkennwerte von den Parametern signifikant ist oder nicht - wir haben also *eine* Stichprobe, deren Kennwerte wir mit vorgegebenen Parametern vergleichen. Die Parameter können aus unterschiedlichen Quellen vorliegen:

- Sie können aus einer grossen Datenmenge berechnet sein. Im Test geht es darum zu überprüfen, ob die Stichprobenkennwerte mit den Kennwerten der betrachteten grösseren Datenmenge verträglich sind. So können wir z.B. überprüfen, ob ein Stichprobenmittel von Wallisern bezüglich Einkommen mit dem mittleren schweizerischen Einkommen verträglich ist oder mehr als zufällig davon abweicht.
- Die vorgegebenen Parameter können gesetzlich definiert sein. In Umwelt- und Gesundheitsgesetzen wird etwa ein Grenzwert angegeben, der nicht überschritten werden darf. Im statistischen Test geht es darum zu überprüfen, ob ein Stichprobenmittel mit dem untersuchten Grenzwert verträglich ist oder mehr als zufällig davon abweicht.
- Die Parameter können vertraglich festgelegt sein (Qualitätsstandards). So könnte z.B. die Milch eines Bauern einen Fettanteil von 3.8% aufweisen sollen. Es wird getestet, ob eine festgestellte Abweichung des Stichprobenmittels „Fettanteil“ von 20 Lieferungen gut oder schlecht durch Zufall zu erklären ist.
- Die Parameter können aus einer Theorie folgen. Im Test geht es darum zu prüfen, ob eine Stichprobe die Theorie bestätigt oder nicht. Die Theorie wird nicht bestätigt, wenn die Parameter der Stichprobe mehr als zufällig von den theoretischen Parametern abweichen.

Die vorgegebenen Parameter charakterisieren die Verteilung der Zufallsvariablen unter der Nullhypothese. Dadurch wird die Verteilung einer Statistik bestimmt, welche die Stichprobenkennzahlen als Werte annimmt. Ein konkreter Stichprobenkennwert weicht signifikant vom vorgegebenen Parameter ab, wenn sein Eintreffen unter der Nullhypothese genügend unwahrscheinlich ist.

Die folgenden Tests setzen alle

1. die Unabhängigkeit der Zufallsvariablen voraus. Dies gewährleisten wir in der Praxis durch inhaltliche Überlegungen (gibt es eine gegenseitige Beeinflussung der Ergebnisse bei den wiederholten Zufallsexperimenten, welche die Werte der Zufallsvariablen produzieren, oder nicht).

2. identisch verteilte Zufallsvariablen voraus. Die identische Verteiltheit wird durch inhaltliche Überlegungen gestützt: sind die Zufallsexperimente, welche die Daten produzieren, genügend einheitlich, so dass wir sie als Wiederholungen desselben Zufallsexperimentes betrachten können?

20.1 z-Test

Ziel: Getestet wird die Abweichung des Mittelwertes $\bar{x}$ (= Wert der Zufallsvariable $\bar{X}$) einer Stichprobe $(X_1, \dots, X_n)$ von einem vorgegebenen Parameter $\bar{x}_o$ (= Erwartungswert der X_i unter der Nullhypothese). Gegeben ist ein zusätzlicher Hilfsparameter s_o^2 (= Varianz der X_i unter der Nullhypothese).

Voraussetzungen: die Zufallsvariablen X_i , als deren Werte die Daten x_i der Stichprobe betrachtet werden, sind unabhängig und identisch verteilt. Der Parameter s_o^2 ist bekannt (bei kleinen Stichproben - $n < 30$ - müssen die Daten normalverteilt sein) (für stetig, metrisch skalierte Variablen).

Unter der Nullhypothese sind die Daten der Stichprobe als Werte von Zufallsvariablen zu betrachten, welche eine Verteilung mit den Parametern $E(X_i) = \bar{x}_o$ und $V(X_i) = s_o^2$ aufweisen. Damit können wir die Zufallsvariable $\bar{X}$, als deren Werte wir konkrete Stichprobenmittelwerte $\bar{x}$ betrachten, standardisieren:

$$\frac{\bar{X} - \bar{x}_o}{\frac{s_o}{\sqrt{n}}}.$$

Diese Zufallsvariable ist laut dem Zentralen Grenzwertsatz für genügend grosse n näherungsweise standardnormalverteilt. Dies führt somit auf die folgende Teststatistik Z :

$Z = \frac{\bar{X} - \bar{x}_o}{\frac{s_o}{\sqrt{n}}} \sim N(0, 1)$ (gilt asymptotisch)
$\bar{X}$: Statistik mit dem Wert $\bar{x}$ (Mittelwert der n Daten)
$\bar{x}_o$: Parameter der Zufallsvariablen X_i unter der Nullhypothese (Erwartungswert)
s_o : Parameter der Zufallsvariablen X_i unter der Nullhypothese (Standardabweichung)
X_i : Stichprobenzufallsvariablen
n : Stichprobengrösse (= Anzahl Stichprobenvariablen)

Einen Test, der eine standardnormalverteilte Teststatistik verwendet, nennen wir oft „z-Test“.

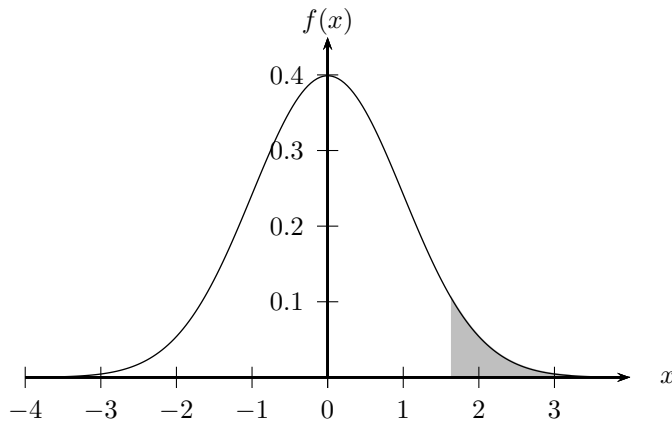


Abbildung 20.1.1: Standardnormalverteilte Teststatistik Z und rechtsseitiger Test; unter der Nullhypothese beträgt $P(Z \geq 1.645) \approx 0.05 = 5\%$ d.h. $1.645 \approx \Phi^{-1}(0.95)$. Der 0.05–Verwerfungsbereich ist also $]1.645, \infty[$

Man überlege sich die intuitive Bedeutung der Formel

$$Z = \frac{\bar{X} - \bar{x}_o}{\frac{s_o}{\sqrt{n}}} :$$

Der konkrete Testwert z weicht mehr von 0 ab, wenn die Differenz zwischen dem Wert von $\bar{X}$ und $\bar{x}_o$ grösser ist. Zudem wird der Testwert kleiner, wenn die Varianz grösser ist, d.h. je mehr die Daten streuen, desto weniger ist die Differenz von $\bar{x}$ und $\bar{x}_o$ bedeutsam. Der Testwert steigt, wenn die Zahl der Daten grösser wird, d.h. je grösser die Stichprobe, desto eher wird die Differenz von $\bar{x}$ und $\bar{x}_o$ statistisch bedeutsam. Man sieht: bei einer genügend grossen Stichprobe werden beliebig kleine Unterschiede signifikant. Diese sind vermutlich ökonomisch unbedeutend. Man darf sich nicht auf die Signifikanz alleine verlassen und muss die inhaltliche Bedeutsamkeit im Auge behalten. So kann ein mittlerer Gewichtsunterschied von 100 g bei der Schweinemast statistisch signifikant sein. Die Umstellung auf die neue Mastmethode kann aber teurer sein als der durch die höheren mittleren Masterträge zu erwartende höhere Gewinn. Am besten ist es, wenn man sich vorgängig Überlegungen über die ökonomisch interessante Unterschiede macht. Es wird dann die Stichprobengrösse berechnet, welche den minimalen ökonomisch relevanten Unterschied signifikant werden lässt.

Beispiel 20.1.1. Wir wollen testen, ob sich nach der Einführung einer neuen Mastmethode das durchschnittliche Gewicht $\bar{x}$ von n Schweinen signifikant erhöht hat, wobei die Standardabweichung s_o und der Mittelwert $\bar{x}_o$ der Ergebnisse der traditionellen Mastmethode bekannt ist. Wir nehmen an, die Einzelergebnisse x_i , die zum Mittelwert $\bar{x}$ führen, seien Realisierungen von n identisch verteilten und unabhängigen Zufallsvariablen X_i (Unabhängigkeit ist bei Schweinemast recht problematisch - ansteckende Krankheiten, die sich aufs Gewicht auswirken! Wir ziehen aus den Schweinen, die mit der neuen Mastmethode gezüchtet wurden, eine Zufallsstichprobe, um das Problem etwas zu entschärfen. Identische Verteiltheit wird durch die Anwendung einer einzigen Mastmethode gewährleistet, wobei Mastmethode weit gefasst ist und z.B. Auslauf, Boxgrösse, Luft, etc. umfasst). Unter der Nullhypothese ist der Mittelwert der Stichprobe $\bar{x}$ als Wert der Zufallsvariable $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ zu betrachten, mit $E(X_i) = \bar{x}_o$ und $V(X_i) = s_o$. In einem konkreten Fall möge gelten:

$\bar{x} = 20.5$ kg (= Mittelwert der Stichprobe aus den mit der neuen Mastmethode gezüchteten Schweine = Wert der Zufallsvariable $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$);

$\bar{x}_o = 19.5$ kg (Mittelwert der Schweine bei der alten Mastmethode)

$s_o = 2$ (Standardabweichung der Schweine bei der alten Mastmethode)

$n = 50$ (Stichprobengrösse)

H_0 : Der Mittelwert der neuen Mastmethode ist nicht grösser als der Mittelwert der alten Mastmethode.

H_A : Der Mittelwert der neuen Mastmethode ist grösser als der Mittelwert der alten Mastmethode ($\Rightarrow$ einseitiger, rechtsseitiger Test).

Signifikanzniveau: 0.01.

Wir berechnen den Testwert (= Wert der Statistik Z = Wert der standardisierten Teststatistik „Mittelwert“ $\bar{X}$):

$$\frac{20.5 - 19.5}{\frac{2}{\sqrt{50}}} = 3.5355.$$

Der p -Wert ist

$$P(Z \geq 3.5355)$$

mit $Z \sim N(0, 1)$. Da $\Phi(3.5355) = 0.999796456$, ist der p -Wert $P(Z \geq 3.5355) = 1 - 0.999796456 = 2.0354 \times 10^{-4} < 0.01$. Damit ist der Unterschied hochsignifikant (d.h. schlecht durch Zufall erklärbar!). $\diamond$

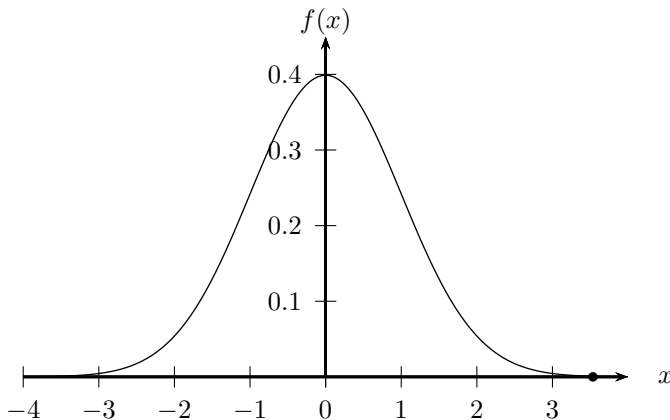


Abbildung 20.1.2: Testwert des Testes im Beispiel und Dichtefunktion der Standardnormalverteilung. Zwischen dem Testwert und Unendlich liegt zwischen der Dichtefunktion und der x -Achse eine Fläche von 0.000203454 (= p -Wert)

Bei kleinen Stichproben müssen wir prüfen, ob die Daten normalverteilt sind, da nur der Mittelwert normalverteilter Daten kleiner Stichproben normalverteilt ist. Die Verteilung der Mittelwerte anderer Verteilungen ist jedoch nur für genügend grosse n näherungsweise normal.

Beispiel 20.1.2. Die durchschnittliche Lebensdauer von Transmissionsriemen eines bestimmten Typs betrage 7 Monate (Standardabweichung: 1). Bei einem Test wird eine neue Generation von Riemen untersucht, die zwar teurer sind, für welche die Werbung jedoch eine längere Lebensdauer verspricht. Es werden 30 zufällig gewählte Riemen der neuen Generation getestet, die durchschnittlich 7.2 Monate halten. Ist der Unterschied signifikant?

H_0 : Die durchschnittliche Lebensdauer der neuen Riemen ist nicht länger.

H_A : Die Lebensdauer der neuen Riemen ist länger ($\Rightarrow$ rechtsseitiger Test).

Wir nehmen an, die Zufallsvariablen X_i , welche konkrete Lebensdauern x_i von Riemen annehmen, seien unter der Nullhypothese unabhängig und identisch verteilt - mit den Parametern $E(X_i) = \bar{x}_0 = 7$ und $V(X_i) = \sigma_0^2 = 1$. Entsprechend können wir einen z -Test durchführen (Signifikanzniveau 0.01):

Der Testwert ist somit

$$\frac{7.2 - 7}{\frac{1}{\sqrt{30}}} = 1.0954.$$

Der p -Wert ist damit - bei $Z \sim N(0, 1)$:

$$P(Z \geq 1.0954) = 1 - 0.863339161 = 0.13666 > 0.01,$$

da

$$\Phi\left(\frac{7.2 - 7}{\frac{1}{\sqrt{30}}}\right) = \Phi(1.0954) = 0.863339161.$$

Der Unterschied der durchschnittlichen Lebensdauern ist nicht signifikant. $\diamond$

Bemerkung 20.1.3. Beachten Sie, dass die Teststatistik wie bei Statistiken üblich auf das kartesische Produkt der Bilder der Stichprobenzufallsvariablen X_i definiert ist. Der Wert der Teststatistik (= Testwert) wird nämlich aus dem Datenvektor berechnet, wobei wir das i -te Datum (= Gewicht des i -ten Schweines oder im zweiten Beispiel die Lebensdauer des i -ten Riemens) als Wert der i -ten Stichprobenvariable X_i betrachten und der konkret vorliegende Datensatz (= Datenvektor) ein Element des kartesischen Produktes der Bilder der Stichprobenzufallsvariablen ist. Dies gilt für alle in der Folge eingeführten Teststatistiken. $\diamond$

Bemerkung 20.1.4. Statt der Statistik Z könnte man auch die Statistik $\bar{X}$ mit

$$\bar{X} \sim N(\bar{x}_0, \frac{s_0^2}{n})$$

verwenden. Der p -Wert des rechtsseitigen Testes ist $P(\bar{X} \geq \bar{x})$, wobei $\bar{x}$ der Mittelwert der Stichprobe ist. So erhält man fürs obige Riemenbeispiel ebenfalls $P(\bar{X} \geq 7.2) = 1 - 0.863329244$ mit $\bar{X} \sim N(7, \frac{1}{30})$, wie Sie selber nachrechnen mögen. Der Grund für die Standardisierung besteht darin, dass man früher Tabellen für die Normalverteilung hatte - und keine Computer oder Taschenrechner - und damit gezwungen war, mit nur einer Normalverteilung, und zwar mit der Standardnormalverteilung, zu arbeiten. Die Statistik Mittelwert musste deshalb standardisiert werden. $\diamond$

Bemerkung 20.1.5. Wie gross muss die Stichprobe sein, damit bei einem Einstichproben-Z-Test eine tatsächliche positive Differenz von 2 kg signifikant werden kann. Eine tatsächliche Differenz von 2 bedeutet, dass die Differenz des Erwartungswertes der Statistik Stichprobenmittel und des gegebenen Nullhypothesen-Mittelwertes $\bar{x}_0$ 2 beträgt (Standardabweichung, gemäss denen die Stichprobenzufallsvariablen verteilt sind, sei 10; Alpha = 0.05; rechtsseitiger Test). $\Phi^{-1}(0.95) = 1.644853627$

$$\begin{aligned} \frac{E(\bar{X}) - \bar{x}_0}{\frac{s_0}{\sqrt{n}}} &= \frac{2}{\frac{10}{\sqrt{n}}} > 1.644853627 \\ \frac{2\sqrt{n}}{10} &> 1.644853627 \\ n &> \left(\frac{1.644853627 \cdot 10}{2}\right)^2 \\ \left(\frac{1.644853627 \cdot 10}{2}\right)^2 &= 67.639 \\ n &= 68 \end{aligned}$$

Allerdings wäre die sogenannte Macht des Testes mit dieser Fallzahl eher gering. Als Macht eines Testes wird die Wahrscheinlichkeit betrachtet, dass die wahre Alternativhypothese akzeptiert wird. Dies ist die Gegenwahrscheinlichkeit zur Fehlerwahrscheinlichkeit 2. Art. Im Beispiel erhält man als Macht des Testes bei $n = 68$, der tatsächlichen Differenz 2 und der Standardabweichung 10 die Fehlerwahrscheinlichkeit 2. Art

$$P(\bar{X} < 1.644853627) = 0.3848141 \text{ für } \bar{X} \sim N\left(2, \frac{10^2}{68}\right)$$

und damit die Macht des Testes

$$1 - P(\bar{X} < 1.644853627) = 0.6151859$$

Man würde die wahre Alternativhypothese, dass der Unterschied der Mittel 2 beträgt, nur mit einer Wahrscheinlichkeit von 0.6151859 entdecken. Oft strebt man eine Mächtigkeit von 0.85 an. Rechnet man $1 - P(X < 1.644853627)$ für $X \sim N\left(2, \frac{10^2}{n}\right)$ und steigende n durch (mit Excel erstellt man eine arithmetische Folge $68+i$ mit $i \in \mathbb{N}$ z.B. in der Spalte A. Sei 68 in der Zelle A2. Dann schreibt man $=1-NORM.VERT(1.644853627;2;10/WURZEL(A2);1)$ in die Zelle B2 und zieht die Zelle B2 hinunter bis der Wert der Formel grösser als 0.85 ist. Mit R geht's schneller: $n=68$; `while (1-pnorm(qnorm(0.95,0,1),2,10/sqrt(n)) < 0.85){ n=n+1; n}`, erhält man als nötiges n die Zahl 852. Diese recht hohe Zahl erklärt sich durch den Umstand, dass eine Standardabweichung der einzelnen Zufallsvariablen von 10 bei einem Abstand von 2 eine sehr grosse Streuung darstellt. Bei der hohen Fallzahl von 852 würden allerdings bereits kleine Differenzen signifikant:

$$P(X > d) = 0.95 \text{ für } \bar{X} \sim N\left(0, \frac{10^2}{852}\right)$$

ergibt

$$d = 0.5635176$$

(berechnet man mit R mittels `qnorm(0.95,0,10/sqrt(852))`). Erhält man also z.B. eine signifikante Differenz von 1, eine Investition würde sich aber nur für eine Differenz von 2 lohnen, müsste man mindestens kontrollieren, ob 2 im Vertrauensintervall von 1 enthalten ist (Vertrauensintervalle behandeln wir später). $\diamond$

20.2 t-Test

Ziel: Getestet wird die Abweichung des Mittelwertes einer Stichprobe von einem gegebenen Parameter $\bar{x}_0$ (Erwartungswert der Stichprobenvariablen X_i unter der Nullhypothese).

Voraussetzungen: die Zufallsvariablen X_i , als deren Werte die Daten x_i der Stichprobe betrachtet werden, sind unabhängig und identisch normalverteilt. Die Varianz der Zufallsvariablen unter der Nullhypothese ist unbekannt und wird aus den Daten der Stichprobe geschätzt (stetig metrisch skalierte Variable).

Da wir die Varianz der Zufallsvariablen unter der Nullhypothese nicht kennen, müssen wir diese aus den Daten schätzen. Dabei ist die empirische Varianz als Wert einer Zufallsvariable

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

aufzufassen. Es gilt:

$$\frac{\bar{X} - \bar{x}_0}{\frac{S}{\sqrt{n}}} = \frac{\bar{X} - \bar{x}_0}{\sqrt{\frac{S^2}{n}}} = \frac{\frac{1}{s}(\bar{X} - \bar{x}_0)}{\frac{1}{s}\sqrt{\frac{S^2}{n}}} = \frac{\frac{1}{s}(\bar{X} - \bar{x}_0)}{\frac{1}{\sqrt{n}}\sqrt{\frac{S^2}{s^2}}} = \frac{\frac{1}{s}(\bar{X} - \bar{x}_0)}{\frac{1}{\sqrt{n}}\sqrt{\frac{(n-1)S^2}{(n-1)s^2}}} = \frac{\frac{(\bar{X} - \bar{x}_0)}{\frac{s}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{(n-1)s^2}}}$$

Zudem gilt:

$$\frac{(\bar{X} - \bar{x}_0)}{\frac{s}{\sqrt{n}}} \sim N(0, 1),$$

wenn die Daten normalverteilt sind und für diesen Fall - wie auf der Seite 393 bei der Herleitung der Verteilung der Teststatistik des Einstichproben-Varianztestes gezeigt werden wird:

$$\frac{(n-1)S^2}{s^2} \text{ ist } \chi_{n-1} - \text{verteilt.}$$

Laut Definition der t -Verteilung gilt somit:

$$\frac{\frac{(\bar{X} - \bar{x}_0)}{\frac{s}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{(n-1)s^2}}} \text{ ist } t_{n-1} - \text{verteilt.}$$

für normalverteilte Daten. Wir erhalten somit die folgende t -verteilte Teststatistik T :

$T = \frac{\bar{X} - \bar{x}_0}{\frac{s}{\sqrt{n}}}$ ist $t_{(n-1)}$ -verteilt (bei normalverteilten Daten).
$\bar{x}_0$: Parameter der Zufallsvariablen X_i unter der Nullhypothese (Erwartungswert; Mittelwert)
X_i : Stichprobenzufallsvariablen
$\bar{X}$: Zufallsvariable mit $\bar{x}$ der Stichprobe als Wert
S : Zufallsvariable mit Standardabweichung s der Stichprobe als Wert.
n : Anzahl der Daten

Wir sehen, dass gegenüber dem z -Test nur s_0 durch S ersetzt wird, dass sich jedoch die Verteilung ändert.

Beispiel 20.2.1. In einem Test wurden die folgenden Bremswege (in Meter) für ein Profil A ermittelt (Zufallsstichprobe aus Reifen mit dem Profil A):

44.5; 55.1; 52.5; 50.2; 45.3; 46.1; 52.1; 50.5; 50.6; 49.2 ($\Rightarrow s^2 = 11.55433333, \bar{x} = 49.61; n = 10$).

Diese seien die Werte von i.i.d Zufallsvariablen X_i . Für das Profil B ergab sich bisher ein Mittelwert von 55. Ist der Unterschied signifikant? ($\alpha = 0.01$)

H_0 : Der Bremswege des Profils A ist nicht kürzer.

H_A : Der mittlere Bremsweg von Profil A ist kürzer ($\Rightarrow$ linksseitiger Test).

Wir überprüfen die Daten mit einem P-P-Plot auf Normalverteilung (s. Abbildung 20.2.3).

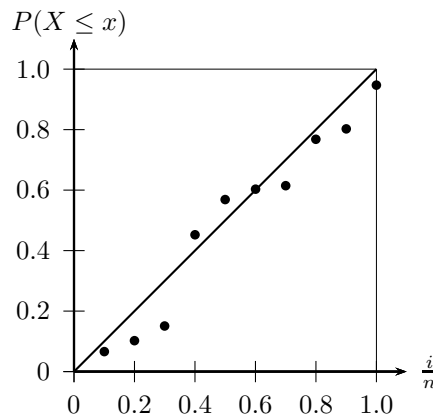


Abbildung 20.2.3: P-P-Plot zur Überprüfung der Voraussetzungen

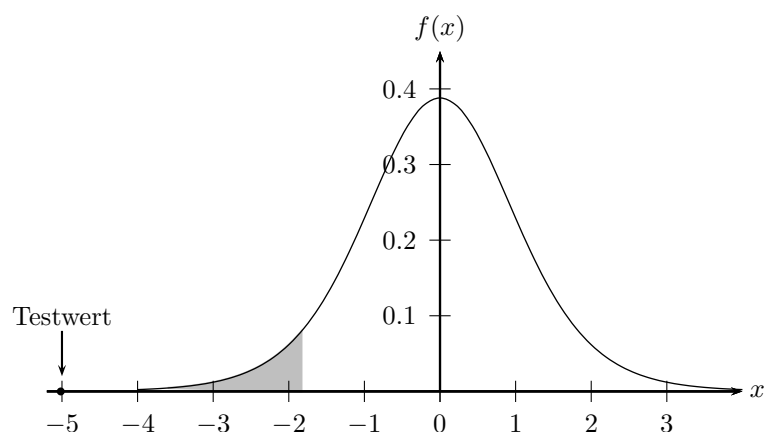
Wir entscheiden, dass die Daten (knapp) der Normalverteilungsannahme entsprechen (es hat etwas wenige Daten!).

Testwert: $\frac{49.61 - 55}{\sqrt{\frac{1}{10} \cdot 11.55433333}} = -5.0144$.

Wir berechnen den p -Wert: $P(T \leq -5.0144)$:

mit Excel: $F_{t(9)}(-5.0144) = 0.000362298 < 0.01$ (mit Excel berechnet mittels $=t.vert(-5.0144; 9; 1)$).

Somit ist das Ergebnis hochsignifikant. $\diamond$

Abbildung 20.2.4: Linksseitiger Test mit t_9 -verteilter Teststatistik

20.3 Wilcoxon-Rang-Test

Ziel: Getestet wird, ob Daten x_i einer Stichprobe zum Median unter der Nullhypothese passen.

Voraussetzungen: die Zufallsvariablen X_i , als deren Werte die Daten x_i der Stichprobe betrachtet werden, sind unabhängig und identisch verteilt. Die Daten sind in etwa symmetrisch verteilt (metrisch skalierte Variable).

Ist die Varianz nicht bekannt und gilt die Normalverteilungsannahme nicht, so kann man den Wilcoxon-Vorzeichen-Rang-Test verwenden. Wir berechnen die Differenzen der Daten zu einem vorgegebenen Median - unter der Voraussetzung der Symmetrie fällt dieser mit dem arithmetischen Mittelwert zusammen. Null-Differenzen werden weggelassen. Der Betrag dieser Differenzen wird rangiert. Dann wird die Summe der Ränge der Differenzen gebildet, welche eine positive (oder negative) Differenz zum Parameter aufweisen. Unter der Nullhypothese sollten die Rangsummen positiver und negativer Differenzen ungefähr gleich gross sein. Sind diese zwei Summen verschieden, so dass der Unterschied schlecht durch Zufall erklärbar ist, verwerfen wir die Nullhypothese.

Wir betrachten das folgende Datenbeispiel, um die Grundideen zu verdeutlichen (s. Tabelle 20.3.1 und Abbildung 20.3.5):

Daten	Differenzen zum Median der alten Grundgesamtheit	Betrag der Differenzen geordnet	Ränge	Ränge negativer Differenzen	Ränge positiver Differenzen
4.3	-0.1	0.1	1	1	0
4.6	0.2	0.2	2	0	2
4.7	0.3	0.3	3	0	3
4	-0.4	0.4	4	4	0
4.9	0.5	0.5	5	0	5
5	0.6	0.6	6	0	6
5.4	1	1	7	0	7
3.1	-1.3	1.3	8	8	0
6.2	1.8	1.8	9	0	9
6.5	2.1	2.1	10	0	10
10.5	6.1	6.1	11	0	11
Summen			66	13	53

Tabelle 20.3.1: Beispieldaten zu den Grundideen des Wilcoxon-Testes. Vergleich mit $med_0 = 4.4$

Die graphische Darstellung dieser Daten mit Hilfe eines eindimensionalen Streudiagramms zeigt, dass mehr Daten - und oft mit grösseren Rängen - auf der rechten Seite des Parameters

unter der Nullhypothese liegen:

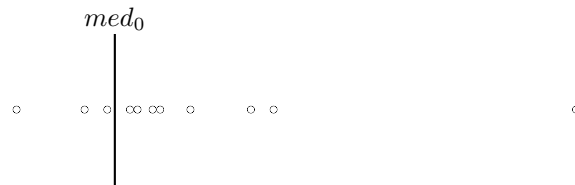


Abbildung 20.3.5: Graphische Darstellung der Daten der Tabelle 20.3.1. Die Rangsumme der Daten mit negativen Differenzen beträgt 13, die Rangsumme der Daten mit positiven Differenzen zum Median 4.4 der Grundgesamtheit beträgt 53

Es stellt sich die Frage, ob die Rangsummenunterscheide gut oder schlecht durch Zufall erklärbar sind. Wir bezeichnen mit R^- die Summe der Ränge negativer Differenzen und mit R^+ die Summe der Ränge positiver Differenzen. R^i ist als Wert einer Zufallsvariable zu betrachten, wobei gilt: $R^- + R^+ = \frac{n(n+1)}{2}$ (n = Anzahl der von 0 verschiedenen Differenzen, $\frac{n(n+1)}{2}$ = Summe aller Ränge).

Die kritischen Werte (= Grenzen des Verwerfungsbereichs; die Grenzen gehören zum Verwerfungsbereich) dieser Verteilung für $n \leq 25$ (n = Anzahl der von 0 verschiedenen Differenzen) sind für die jeweils kleinere Rangsumme in Tabellen angegeben (die Werte werden mit Hilfe von Kombinatorik ausgerechnet):

n	Kritische Werte (einseitig) bei		
	$\alpha = 0.025$	0.01	0.005
	Kritische Werte (zweiseitig) bei		
	$\alpha = 0.05$	0.02	0.01
6	0		
7	2	0	
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

Tabelle 20.3.2: Kritische Werte (= Grenzen des Verwerfungsbereichs) beim Wilcoxon-Test

Für $n > 25$ kann die Standardnormalverteilung verwendet werden:

$Z = \frac{R^i - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0, 1)$ (asymptotisch)	$n > 25$ (für $n \leq 25$ Tabelle 20.3.2!!)
n : Anzahl der von 0 verschiedenen Differenzen	
R^i : Rangsummen der negativen (oder positiven) Differenzen	

In der Tabelle 20.3.1 von Seite 20.3.1 ist $n = 11 < 25$. Mit arbeitet also mit der Tabelle 20.3.2 von Seite 20.3.2. Die kritischen Grenzen sind gemäss Tabelle je nach Signifikanzniveau und je nachdem ob einseitig oder zweiseitig getestet wird 11, 7 oder 5. Da $13 > 11$, ist das Ergebnis für kein Signifikanzniveau signifikant.

Bemerkung 20.3.1. Z ist die standardisierte Zufallsvariable R^i , wobei $V(R^i) = \sqrt{\frac{n(n+1)(2n+1)}{24}}$ und $E(R^i) = \frac{n(n+1)}{4}$. Es gilt nämlich für die Zufallsvariablen

$$W_i := \begin{cases} 1 & \text{wenn } X_i > \text{med}_0 \\ 0 & \text{sonst} \end{cases},$$

dass bei symmetrischer Verteilung unter der Nullhypothese gilt: $P(W_i = 1) = 0.5$ - es handelt sich um bernoulli-verteilte Zufallsvariablen. Für die Statistik R^+ und die Ränge r_i des i -ten Datums des geordneten Datensatz gilt:

$$R^+ = \sum_{i=1}^n r_i W_i$$

Damit gilt (die Ränge sind für gegebenes n gegeben und damit konstant)

$$\begin{aligned} E(R^+) &= E\left(\sum_{i=1}^n r_i W_i\right) \\ &= \sum_{i=1}^n r_i E(W_i) \\ &= \frac{n(n+1)}{2} \frac{1}{2}. \end{aligned}$$

Für die Varianz erhält man mit $\sum_{i=1}^n r_i^2 = \frac{n(n+1)(2n+1)}{6}$

$$\begin{aligned} V(R^+) &= V\left(\sum_{i=1}^n r_i W_i\right) = \sum_{i=1}^n V(r_i W_i) \\ &= \sum_{i=1}^n r_i^2 V(W_i) = \sum_{i=1}^n r_i^2 \frac{1}{2} \frac{1}{2} \\ &= \frac{n(n+1)(2n+1)}{24}. \end{aligned}$$

◇

Bemerkung 20.3.2. Manche Statistikprogramme berechnen Z verteilt nach t_{n-1} . Dadurch wird der p -Wert ein wenig grösser. Es handelt sich um eine etwas vorsichtigere Variante. ◇

Bemerkung 20.3.3. Bei Bindungen, d.h. beim Vorliegen gleicher Ausprägungen, empfiehlt es sich, eine Bindungskorrektur vorzunehmen. Die Varianz wird zu

$$\frac{n(n+1)(2n+1)}{24} - \frac{1}{48} \sum_{j=1}^r (b_j^3 - b_j),$$

wobei r die Anzahl der Gruppen verbundener Daten ist und b_j die Anzahl der Daten in der j -ten Gruppe verbundener Daten. Kommen z.B. 3 Gruppen mit verbundenen Daten vor, die erste mit 3 Daten, die zweite mit 2 Daten und die dritte mit 4 Daten, würde $\frac{1}{48} \sum_{j=1}^r (b_j^3 - b_j)$ zu $\frac{1}{48} ((3^3 - 3) + (2^3 - 2) + (4^3 - 4)) = 1.875$. ◇

Bemerkung 20.3.4. Die Symmetriebedingung kann mit einem Boxplot überprüft werden. ◇

Bemerkung 20.3.5. Die Symmetriebedingung muss erfüllt sein, da sonst ein Test die Abweichung des Medians der Stichprobe vom Median unter der Nullhypothese anzeigen kann, ohne dass eine solche vorliegt. Dies kann das folgende Beispiel klarmachen: Der Median unter der Nullhypothese sei 0.91. Es wurden folgende Daten erhoben (s. Tabelle 20.3.3):

Daten	Differenzen zum Median	Betrag Differenzen	Ränge	Ränge negativer Abweichungen	Ränge positiver Abweichungen
0.9	-0.01	0.01	1.5	1	0
0.92	0.01	0.01	1.5	0	2
0.899	-0.011	0.011	3	3	0
0.93	0.02	0.02	4	0	4
0.886	-0.024	0.024	5	5	0
0.935	0.025	0.025	6	0	6
0.936	0.026	0.026	7	0	7
0.88	-0.03	0.03	8	8	0
0.87	-0.04	0.04	9	9	0
0.865	-0.045	0.045	10	10	0
0.8	-0.11	0.11	11	11	0
5	4.09	4.09	12	0	12
8	7.09	7.09	13	0	13
10	9.09	9.09	14	0	14
11	10.09	10.09	15	0	15
12	11.09	11.09	16	0	16
20	19.09	19.09	17	0	17
21	20.09	20.09	18	0	18
27	26.09	26.09	19	0	19
28	27.09	27.09	20	0	20
30	29.09	29.09	21	0	21
Summen			231	47	184

Tabelle 20.3.3: Beispiel für die Notwendigkeit der Symmetriebedingung

Der Median der Stichprobe ist 0.936 und unterscheidet sich nur unwesentlich von Median der Nullhypothese 0.91. Der Test ist signifikant (zweiseitiger Test, $\alpha = 0.05$, kritischer Wert gemäss Tabelle 20.3.2: 59 - Wert der grösser als 47 ist). Der folgende Boxplot zeigt die extreme Abweichung der Daten von der Symmetriebedingung (s. Abbildung 20.3.6:

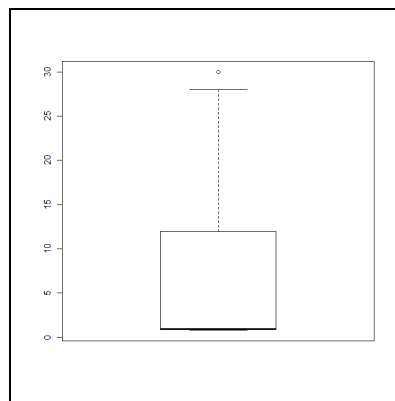


Abbildung 20.3.6: Boxplot des Beispiels der Tabelle 20.3.3

◇

Beispiel 20.3.6. (s. Beispiel 20.2.1) In einem Test wurden die folgenden Bremswege (in Meter) für ein Profil A ermittelt:

44.5; 55.1; 52.5; 50.2; 45.3; 46.1; 52.1; 50.5; 50.6; 49.2 ($\Rightarrow s^2 = 11.55433333, \bar{x} = 49.61; n = 10$).

Für das Profil B ergab sich bisher ein Median von 55. Ist der Bremsweg signifikant kleiner? (links-seitiger Test, $\alpha = 0.025$). Wir überprüfen die Symmetriebedingung mit einem Boxplot: (s. Abbildung 20.3.7)

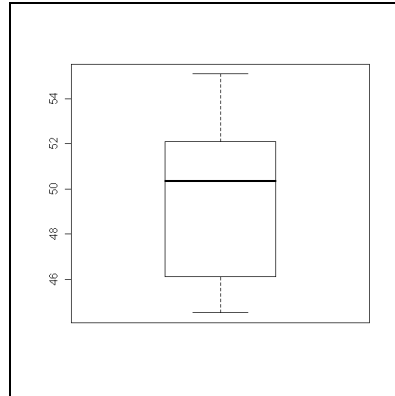


Abbildung 20.3.7: Boxplot der Reifendaten zwecks Überprüfung der Symmetrie

Die Daten sind nicht wirklich symmetrisch verteilt. Wir werden den Test trotzdem fortführen, sicherheitshalber anschließend aber einen weiteren Test auf diese Daten ausführen.

Es ergeben sich die folgenden Differenzen (s. Tabelle 20.3.4):

Bremsweg	Differenz zu 55
44.5	-10.5
55.1	0.1
52.5	-2.5
50.2	-4.8
45.3	-9.7
46.1	-8.9
52.1	-2.9
50.5	-4.5
50.6	-4.4
49.2	-5.8

Tabelle 20.3.4: Daten mit Differenzen zum Median 55

Dem geordneten Betrag der Differenzen werden Ränge zugeordnet - bei gleichen Beträgen wird wie bei der Spearman-Korrelation das jeweilige arithmetische Mittel zugeordnet (s. Tabelle 20.3.5):

Differenzen zu 55	Betrag der Differenzen	Rangierung des Betrags	Ränge negativer Differenzen	Ränge positiver Differenzen
0.1	0.1	1	0	1
-2.5	2.5	2	2	0
-2.9	2.9	3	3	0
-4.4	4.4	4	4	0
-4.5	4.5	5	5	0
-4.8	4.8	6	6	0
-5.8	5.8	7	7	0
-8.9	8.9	8	8	0
-9.7	9.7	9	9	0
-10.5	10.5	10	10	0
	Summen	55	54	1

Tabelle 20.3.5: Rangierte Daten von Tabelle 20.3.4

Die Rangsumme der Daten mit negativer Differenz ist offensichtlich viel grösser als die andere Rangsumme. Die Verteilung der Rangsummen ist symmetrisch, so dass in den Tabellen nur die kleineren Werte angegeben werden. Wir untersuchen also, ob 1 kleiner ist als der kritische Wert in der Tabelle 20.3.2. Für 10 Daten ist der kritische Wert für den einseitigen Test mit $\alpha = 0.025$ mit 8 identisch. Da $1 < 8$ sind die Bremswege der Stichprobe signifikant kleiner. $\diamond$

Beispiel 20.3.7. Nach einem Kurs in Kundenberatung möchte eine Versicherungsgesellschaft überprüfen, ob sich die Umsätze zufällig gewählter Vertreter im Vergleichsmonat August vergrößert haben. Im Jahr 2008 beliefen sich die durchschnittlichen Umsätze auf 28 000 im Monat. Im Jahr 2009 ergaben sich die folgenden Zahlen (s. Excel-Tabelle 14EinstichprobenTests.xls, BeiWilcox2). Es gilt $n = 103$. Wir könnten überprüfen, ob die Daten normalverteilt sind und den t-Test anwenden. Übungshalber wenden wir den Wilcoxon-Test an. Es ergibt sich: Rangsumme: 5356
Ränge negativer Differenzen: 1895; Ränge positiver Differenzen: 3461. Rechtsseitiger Test, $\alpha = 0.05$. Da die Rangsummen in der Tabelle nicht vorkommen, verwenden wir die Normalverteilungsnäherung:

$$z = \frac{3461 - \frac{103(103+1)}{4}}{\sqrt{\frac{103(103+1)((2 \cdot 103 + 1))}{24}}} = 2.576$$

$1 - \Phi(2.576) = 1 - 0.995002468 = 0.004997532 < 0.05$. Die Umsätze sind damit signifikant gestiegen. $\diamond$

20.4 Vorzeichentest

Ziel: Getestet wird, ob die Daten x_i einer Stichprobe zum vorgegebenen Parameter med_0 (Median der Verteilung unter der Nullhypothese) passen.

Voraussetzungen: die Zufallsvariablen X_i , als deren Werte die Daten x_i der Stichprobe betrachtet werden, sind unabhängig und identisch verteilt. (ordinal oder metrisch skalierte Variable)

Die Voraussetzungen des t-Testes sind recht stark (Normalverteilung der Daten, metrische Skalierung der Daten). Auch beim z-Test müssen wir bei kleinen Datensätzen die Normalverteilung der Daten voraussetzen. Zudem muss die Varianz der Zufallsvariablen unter der Nullhypothese bekannt sein und die Daten müssen metrisch skaliert sein. Beim Wilcoxon-Test wird immer noch die metrische Skalierung der Daten vorausgesetzt. Mit weniger Voraussetzungen kommt der Vorzeichentest aus. Es handelt sich um einen Binomialtest: unter der Nullhypothese fallen ungefähr

gleich viele Daten links und rechts des Medians, d.h. die Wahrscheinlichkeit rechts (oder links) vom Median zu fallen beträgt 0.5:

$X \sim B(n, 0.5)$
$n = \text{Anzahl der Daten, die nicht auf } med_0 \text{ fallen.}$
$med_0 = \text{Median der Zufallsvariablen unter der Nullhypothese}$

Bemerkung 20.4.1. *Statt der Anzahl der Daten zu zählen, die rechts (oder links) fallen, können wir auch die Anzahl der positiven oder negativen Vorzeichen von $x_i - med_0$ zählen. Daher der Name „Vorzeichentest“.* $\diamond$

Beispiel 20.4.2. *Im obigen Reifenbeispiel 20.2.1 sei $med_0 = 55$.*

H_0 : *Die Daten liegen nicht mehrheitlich unter einem Median von 55.*

H_A : *Die Daten liegen mehrheitlich unter einem Median von 55. ($\alpha = 0.05$)*

Wir berechnen nun alle Differenzen $x_i - 55$ und zählen die Anzahl der negativen Vorzeichen dieser Differenzen (d.h. wir zählen die Anzahl m der Bremswege x_i , für die gilt: $x_i < 55$). Es sind dies 9 negative Vorzeichen auf 10 Daten. Wir berechnen also: $P(X \geq 9)$ mit $X \sim B(10, 0.5)$. $P(X \geq 9) = 0.010742188 < 0.05$ (rechtsseitiger Test). Das Ergebnis ist signifikant und die Nullhypothese kann verworfen werden. $\diamond$

20.5 Mächtigkeitsvergleich der vier Tests

Der Vorzeichentest verwendet nicht alle Informationen der metrisch skalierten Daten: er berücksichtigt nicht, ob die Daten nahe oder weit vom „alten“ Median entfernt sind. Damit läuft man Gefahr, Opfer des Fehlers zweiter Art zu werden. Die Macht des Testes verringert sich. Wir können somit feststellen: Der Vorzeichentest macht weniger Voraussetzungen, er ist aber i.A. weniger mächtig als der t - oder der z -Test bei Daten, die die Voraussetzungen dieser Tests erfüllen. Im Allgemeinen gilt: der z -Test ist mächtiger als t -Test; der t -Test ist mächtiger als Wilcoxon-Test, Wilcoxon-Test ist mächtiger als Vorzeichentest, sofern die jeweiligen Voraussetzungen erfüllt sind. Sind die Voraussetzungen nicht erfüllt, müssen die Feststellungen allerdings nicht gelten. Je flacher eine Verteilung ist und desto mehr Gewicht Daten haben, die weiter weg vom Median liegen, desto schlechter schneidet der t -Test ab und desto besser der Wilcoxon-Test. Dieser schneidet gegenüber Vorzeichentest im Allgemeinen besser ab, ausser die Verteilung ist extrem flach, d.h. Daten, die weit weg vom Median liegen, haben ein sehr starkes Gewicht (s. für allerdings ziemlich mathematische Ausführungen zum Thema ([Hettmansperger, 1984](#))).

20.6 Test auf abweichende Varianz

Ziel: Getestet wird die Abweichung der Varianz s^2 einer Stichprobe von einem bekannten Parameter s_0^2 (Varianz der Zufallsvariablen X_i unter der Nullhypothese).

Voraussetzungen: Die Zufallsvariablen X_i , als deren Werte wir die Daten x_i auffassen, sind identisch normalverteilt und unabhängig. Die Varianz der Zufallsvariablen unter der Nullhypothese ist bekannt (stetig, metrisch skalierte Variablen) (i.i.d)

Beispiel 20.6.1. *Eine Abfüllanlage führt durch zu grosse Streuung zu hohen Verlusten. Die Firma überlegt sich die Anschaffung einer neuen Maschine, die eine signifikant kleinere Streuung aufweisen sollte als die alte. Als Streumass wird die Varianz verwendet. Die neue Maschine wird getestet, indem aus abgefüllten Gläsern 100 zufällig ausgewählt werden.*

H_0 : *Die Varianz der neuen Abfüllanlage ist nicht kleiner.*

H_A : *Die Varianz der neuen Abfüllanlage ist kleiner ($\implies$ linksseitiger Test).* $\diamond$

Um das Beispiel zu lösen, brauchen wir einen Ausdruck, der uns die Verteilung der Zufallsvariable Varianz angibt. Wir definieren die *Zufallsvariable*

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

für X_i i.i.d. Diese Zufallsvariable weist konkrete Varianzen als Werte auf. Wenn die

$$\frac{X_i - \bar{X}}{s_o}$$

standardnormalverteilt und unabhängig sind (und das gilt genau dann, wenn die Daten x_i Werte normalverteilter und unabhängiger Zufallsvariablen X_i sind), dann gilt für den Ausdruck laut Definition der χ_n^2 -Verteilung:

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_o} \right)^2 \text{ ist } \chi_{n-1}^2 \text{ - verteilt.} \quad (20.6.1)$$

- n ist die Anzahl der unabhängigen Zufallsvariablen X_i . Da $\bar{X}$ aus den X_i berechnet wird, gilt für die Zufallszahlen $D_i = X_i - \bar{X}$: n unabhängige X_i ergibt $n-1$ unabhängige D_i . Sind nämlich $n-1$ Daten und der Mittelwert gegeben, dann ist auch die n -te Zahl bestimmt. (Beispiel: Der Mittelwert von fünf Zahlen (1, 3, 10, 4, x) sei 6. Nun kann x berechnet werden: $\frac{1+3+10+4+x}{5} = 6 \Rightarrow x = 12$).

Durch Umformen erhält man aus dem linken Ausdruck von (20.6.1):

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_o} \right)^2 = \frac{1}{s_o^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{s_o^2} \frac{n-1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{(n-1)S^2}{s_o^2}$$

Im ersten Schritt klammern wir $\frac{1}{s_o^2}$ aus.

Im zweiten Schritt erweitern wir mit $n-1$.

Im dritten Schritt wenden wir die Definition von S^2 an.

Wir erhalten damit die folgende chi-quadrat verteilte Teststatistik (= Zufallsvariable) X :

$\frac{(n-1)S^2}{s_o^2}$ ist χ_{n-1}^2 -verteilt.
S^2 : Zufallsvariable mit dem Wert s^2 der Stichprobe
s_o^2 : Varianz der Zufallsvariablen unter der Nullhypothese
n : Anzahl Daten

Betrachten wir die Formel

$$\frac{(n-1)S^2}{s_o^2},$$

so erkennen wir: Nimmt S^2 den Wert s_o^2 an, so nimmt die Teststatistik

$$\frac{(n-1)S^2}{s_o^2}$$

den Wert $n-1$ an. Dies ist der Erwartungswert der Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden. Nimmt S^2 einen Wert grösser als s_o^2 an, erhalten wir einen Testwert, der rechts vom Erwartungswert liegt. Nimmt S^2 einen Wert kleiner als s_o^2 an, erhalten wir einen Testwert, der links vom Erwartungswert liegt. Wir erhalten damit einen linkseitigen Test, wenn wir testen, ob die Stichprobenvarianz signifikant tiefer ist als die Varianz der Zufallsvariablen unter der Nullhypothese und wir erhalten einen rechtsseitigen Test, wenn wir testen, ob die Stichprobenvarianz signifikant höher liegt als die Varianz der Zufallsvariablen unter der Nullhypothese.

In Fortführung des Beispiels 20.6.1 nehmen wir nun an, wir hätten für die alte Abfüllanlage eine Standardabweichung von 15 g, für die neue eine solche von 10 g. Es stellt sich die Frage, ob die Abweichung signifikant ist ($\alpha = 0.01$). Wir müssen zuerst mit Hilfe eines P-P-Plots untersuchen, ob die Daten normalverteilt sind. Wenn die Normalverteilungsannahme bezüglich der Daten akzeptiert werden kann, können wir den folgenden Testwert berechnen: $\frac{10^2}{15^2}(99) = 44$; Wir berechnen den p-Wert, wobei $X \sim \chi_{99}^2$ -verteilt ist: p-Werte = $P(X \leq 44) = F_{\chi_{99}^2}(44) = 1 - 0.999999679 = 3.21 \times 10^{-7} \leq 0.01$. Die Varianz der neuen Abfüllanlage unterscheidet sich (hoch) signifikant von der alten.

20.7 Anteilstest bei zwei Ausprägungen

Ziel: Es geht darum, zu untersuchen, ob der Anteil einer Stichprobe (Variable mit nur zwei Ausprägungen) zum vorgegebenen Parameter p (Anteil unter der Nullhypothese) passt.

Voraussetzung: i.i.d Zufallsvariablen X_i

Wir können bei Einstichproben-Anteilstest (mit nur zwei Ausprägungen) eine binomialverteilte Teststatistik verwenden:

$X \sim B(n, p)$
$n = \text{Anzahl der Daten.}$
$p = \text{Anteil unter der Nullhypothese}$

Beispiel 20.7.1. Wir nehmen an, der Anteil der Frauen bei den Konsumenten eines Produktes betrage 0.3. Die Branche möchte diesen Anteil erhöhen und lanciert eine Werbekampagne. Nach der Kampagne wird eine Stichprobe gezogen, um zu überprüfen, ob die Kampagne wirksam war. Wir nehmen an, in der Zufallsstichprobe ($n = 50$; die Stichprobe umfasst nur Konsumenten des Produktes) habe sich ein Anteil Frauen von 0.4 ergeben. Ist der Unterschied signifikant?

H_0 : Die Anteile unterscheiden sich nicht

H_A : Der Anteil der Frauen ist gestiegen ($\implies$ rechtsseitiger Test; $\alpha = 0.05$).

Unter der Nullhypothese beträgt die Wahrscheinlichkeit, eine Frau zu ziehen, 0.3. Die Situation kann offensichtlich mit Hilfe einer binomialverteilten Zufallsvariable modelliert werden, wenn die Daten aus einer genügend grossen Gesamtheit stammen, so dass wir auf Zurücklegen verzichten können: Wie hoch ist die Wahrscheinlichkeit, aus einer Gesamtheit mit einem Anteil von 0.3 Frauen, eine Stichprobe zu ziehen, in der sich mindestens ein Anteil von 0.4 Frauen ergibt. Mit $0.4 \cdot 50 = 20$, ist somit $P(X \geq 20)$ gesucht für $X \sim B(50, 0.3)$.

Wir erhalten: $P(X \geq 20) = \sum_{i=20}^{50} \binom{50}{i} 0.3^i (1 - 0.3)^{50-i} = 0.084803 > 0.05$.

Der Unterschied der Anteile ist nicht signifikant. ◇

20.8 Anteilstest bei mehr als zwei Ausprägungen

Ziel: Es werden die absoluten Häufigkeiten einer statistischen Variable (mit m Ausprägungen) einer Stichprobe mit dem Parametervektor $(k_1^e, \dots, k_m^e)$ (absoluten Häufigkeitsverteilung unter der Nullhypothese) verglichen.

Voraussetzungen: die Daten sind Werte unabhängiger, identisch verteilter Zufallsvariablen. Die Häufigkeiten $(k_1^e, \dots, k_m^e)$ sind in maximal 20% der Zellen kleiner als 5.

$(k_1^e, \dots, k_m^e)$ ist gewöhnlich aus einer absoluten Häufigkeitsverteilung $(k_1^o, \dots, k_m^o)$ oder relativen Häufigkeitsverteilung $(h_1^o, \dots, h_m^o)$ zu berechnen, und zwar mittels

$$(k_1^e, \dots, k_m^e) = n \cdot (h_1^o, \dots, h_m^o) = \frac{n}{N} \cdot (k_1^o, \dots, k_m^o),$$

wobei $N = \sum_{i=1}^m k_i^o$ die Populationsgrösse und n die Stichprobengrösse ist.

Wir entwickeln die Grundideen ausführlicher an einem

Beispiel 20.8.1. Wir nehmen an, in der Schweiz würden sich die Konsumenten wie folgt verhalten (s. Tabelle 20.8.6):

Anteile	Marke
0.2	A
0.3	B
0.35	C
0.15	D

Tabelle 20.8.6: Relative Häufigkeitsverteilung von Marken

Nach einer Werbekampagne für die Marke D wird eine Zufallsstichprobe ($n = 150$) gezogen, um die Wirkung der Werbekampagne zu überprüfen. Es ergeben sich folgende Werte (s. Tabelle 20.8.7):

	absolute Häufigkeiten	relative Häufigkeiten	Marke
	30	0.2	A
	50	0.333	B
	30	0.2	C
	40	0.266	D
Summen	150	1	

Tabelle 20.8.7: Absolute und relative Häufigkeitsverteilung von Marken nach einer Werbekampagne

Hat sich im Verteilungsgefüge ein Unterschied ergeben?

◇

Zuerst müssen wir eine Masszahl für die Abweichungen zweier Häufigkeitsverteilungen entwickeln. Es gibt dabei verschiedene Möglichkeiten. Für die Entwicklung einer Teststatistik ist die folgende nützlich: die relativen Häufigkeiten der Population werden in absolute Häufigkeiten umgerechnet, so dass deren Summe mit der Stichprobengrösse identisch ist. Diese Häufigkeiten sind die für die Stichprobe zu erwartenden Häufigkeiten, sofern die Stichprobe die gleiche Häufigkeitsverteilung aufweist wie die Population. Wir nennen deshalb diese aus der Grosspopulation berechneten Häufigkeiten „erwartete Häufigkeiten“ und verwenden dafür die Abkürzung

$$k_i^e := h_i^o \cdot n$$

n = Stichprobengrösse

h_i^o = Anteil der Klasse i an Population

Fürs obige Beispiel erhalten wir auf Grund der Häufigkeitsverteilung in der Population folgende zu erwartende absolute Häufigkeiten :

$$\begin{aligned} k_i^e & \\ 0.2 \cdot 150 &= 30 \\ 0.3 \cdot 150 &= 45 \\ 0.35 \cdot 150 &= 52.5 \\ 0.15 \cdot 150 &= 22.5 \end{aligned}$$

Als Kennzahl der Abweichung der beiden Häufigkeitsverteilungen können wir dann die quadrierten Abweichungen $K_i - k_i^e$ an k_i^e messen und diese Ausdrücke addieren. Wir erhalten:

$$\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e},$$

wobei K_i eine Zufallsvariable ist, welche die absoluten Häufigkeiten k_i der Stichprobe als Werte annimmt (m = Anzahl der Klassen; k_i^e = absolute Häufigkeiten unter der Nullhypothese = erwartete Häufigkeiten).

Man kann beweisen, dass diese Statistik näherungsweise χ^2 -verteilt ist mit $m - 1$ Freiheitsgraden, sofern die erwarteten Häufigkeiten in mindestens 20% der Klassen grösser gleich 5 sind (K_i unabhängig). Wir erhalten also die chi-quadratverteilte Teststatistik X :

$X^2 = \sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e}$ ist asymptotisch χ_{m-1}^2 -verteilt.
für $k_i^e \geq 5$ für mindestens 20% der $1 \leq i \leq m$
m : Anzahl der Ausprägungen der statistischen Variable
K_i : Zufallsvariablen, welche absoluten Häufigkeiten k_i der Stichprobe annehmen
k_i^e : erwartete absolute Häufigkeiten

Begründung: Wir können jede der m erwarteten Häufigkeiten k_i^e als den Erwartungswert einer poisson-verteilten Zufallsvariablen K_i betrachten und die Stichprobenhäufigkeiten k_i als deren Werte. Wie wir auf Grund des zentralen Grenzwertsatzes wissen, gilt näherungsweise

$$\frac{X - \lambda}{\sqrt{\lambda}} \sim N(0, 1).$$

Eingesetzt erhalten wir:

$$\frac{K_i - k_i^e}{\sqrt{k_i^e}} \sim N(0, 1).$$

Wir quadrieren:

$$\left(\frac{K_i - k_i^e}{\sqrt{k_i^e}} \right)^2 = \frac{(K_i - k_i^e)^2}{k_i^e} \text{ ist } \chi_1^2 - \text{verteilt.}$$

- laut Definition der χ_1^2 -Verteilung. Die Summe dieser Zufallsvariablen ist dann näherungsweise

$$\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e} \text{ ist } \chi_{m-1}^2 - \text{verteilt.}$$

Einen Freiheitsgrad verlieren wir, weil die Stichprobengrösse gegeben ist und damit k_i , wenn die $k_{j \neq i}$ bestimmt sind. Damit liegen $m - 1$ unabhängige Zufallsvariablen vor, die in die Summe $\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e}$ eingehen.

In Fortführung von Beispiel 20.8.1 erhalten wir:

H_0 : Die Häufigkeitsverteilungen unterscheiden sich nicht.

H_A : Die Häufigkeitsverteilungen unterscheiden sich (rechtsseitiger Test - siehe Bemerkung unten; $\alpha = 0.05$) :

Testwert der χ_3^2 -verteilten Teststatistik:

$$\sum_{i=1}^4 \frac{(k_i - k_i^e)^2}{k_i^e} = \frac{(30 - 30)^2}{30} + \frac{(50 - 45)^2}{45} + \frac{(30 - 52.5)^2}{52.5} + \frac{(40 - 22.5)^2}{22.5} = 23.810$$

$P(X \geq 23.810) = 2.73749E - 05 \leq 0.05$. Die Abweichung ist hoch signifikant. Die Nullhypothese H_0 kann verworfen werden: die Häufigkeitsverteilung der Stichprobe passt nicht zur Häufigkeitsverteilung der betrachteten Population.

Beispiel 20.8.2. Die Variable „Nettoeinkommen des meistverdienenden Elternteils der Studenten und Studentinnen der x -ten Klasse einer Schule“ weise 5 Ausprägungen auf (siehe Tabelle unten). Wir nehmen an, es würde sich folgende Klassifizierung ergeben - Häufigkeiten k_i pro Ausprägung (s. Tabelle 20.8.8):

Wir wollen überprüfen, ob gewisse soziale Schichten an dieser Schule über- oder untervertreten sind. Wir vergleichen die Häufigkeitsverteilung mit den Daten des Bundesamtes für Statistik (s. Tabelle 20.8.9):

Einkommensklassen	k_i
0-3000	0
3001-4000	7
4001-5000	6
5001-6000	7
6001-20001+	14
Summen	34

Tabelle 20.8.8: Klassifizierte Nettoeinkommen des meistverdienenden Elternteils der Studierenden der x-ten Klasse einer Schule

Lohnhöhenklasse in Fr.	Total in %	Kumul in %	Frauen in %	Kumul in %	Männer in %	Kumul in %
0-1000	4.5	4.5	9	9	1.6	1.6
1001-2000	6.2	10.60	13.3	22.3	1.6	3.2
2001-3000	10.4	21.00	20.4	42.8	3.9	7.1
3001-4000	19.1	40.10	25.4	68.1	15.2	22.3
4001-5000	23.1	63.20	16.1	84.2	27.6	49.9
5001-6000	14.7	77.90	8.2	92.3	18.9	68.8
6001-7000	8.4	86.30	3.8	96.2	11.3	80.1
7001-8000	4.8	91.10	1.7	97.8	6.8	86.9
8001-9000	2.8	93.90	0.9	98.7	4.1	91
9001-10000	1.9	95.80	0.5	99.2	2.7	93.7
10001-11000	1.2	97.00	0.3	99.5	1.8	95.5
11001-12000	0.8	97.70	0.2	99.6	1.1	96.6
12001-13000	0.5	98.30	0.1	99.7	0.8	97.4
13001-14000	0.3	98.60	0.1	99.8	0.5	97.9
14001-15000	0.3	98.90	0	99.8	0.4	98.3
15001-16000	0.2	99.00	0	99.9	0.3	98.6
16001-17000	0.2	99.20	0	99.9	0.2	98.8
17001-18000	0.1	99.30	*	99.9	0.2	99
18001-19000	0.1	99.40	*	99.9	0.1	99.1
19001-20000	0.1	99.50	*	99.9	0.1	99.2
20001+	0.5	100.00	0.1	100	0.8	100
Summen	100		100		100	

Tabelle 20.8.9: Monatlicher Nettolohn, relative Häufigkeit in %; Bundesamt für Statistik, Lohnstrukturerhebung 2000

Wir fassen die Ausprägungen so zusammen, dass sie mit der Klassifikation der Tabelle 20.8.8 übereinstimmt und rechnen die Zahlen auf absolute Zahlen bezüglich der Stichprobengrösse 34 um (dies ergibt die Spalte mit den „erwarteten Häufigkeiten“ k_i^e der Tabelle 20.8.10). Da die Summe der Prozentzahlen wegen Rundungen nicht genau 100% ergibt, korrigieren wir die entsprechende Spalte. Man erhält:

	k_i	h_i^e	h_i^e korrigiert	$k_i^e = (h_i^e \text{ korrigiert}) \cdot 34$	$\frac{(k_i - k_i^e)^2}{k_i^e}$
0-3000	0.000000	21.100000	21.057884	7.159681	7.159680639
3001-4000	7.000000	19.100000	19.061876	6.481038	0.041555325
4001-5000	6.000000	23.100000	23.053892	7.838323	0.431142299
5001-6000	7.000000	14.700000	14.670659	4.988024	0.811553364
6001-21000+	14.000000	22.200000	22.155689	7.532934	5.552012033
Summen	34	100.2	100	34	13.99594366

Tabelle 20.8.10: Vergleich Einkommensverteilung der Eltern eine Klasse mit der CH-Einkommensverteilung

Die Summe der quadrierten, gewichteten Abweichungen der absoluten faktischen und erwarteten Häufigkeiten ist also

$$\sum_{i=1}^M \frac{(K_i - k_i^e)^2}{k_i^e} = 13.99594366.$$

und man erhält den p -Wert (es hat 5 Klassen, $df = 5 - 1 = 4$):

$$1 - F_{\chi^2_{5-1}}(13.99594366) = P(X \geq 13.99594366) = 0.007308013 \text{ für } \chi^2_{5-1} - \text{verteiltes } X.$$

◇

Bemerkung 20.8.3. Im Spezialfall von zwei Anteilen erhalten wir eine Alternative (Näherung) für den früher durchgeführten exakten Anteilsvergleich (Binomialtest). Allerdings ist diese Alternative bedeutend weniger mächtig und empfiehlt sich daher kaum (siehe Beispiel in den Übungen). ◇

Bemerkung 20.8.4. Einen analogen Chi-Quadrat-Test werden wir später brauchen, um zu überprüfen, ob Daten zu einer theoretischen Wahrscheinlichkeitsverteilung passen oder nicht (Chi-Quadrat-Anpassungstest). Die Parameter entstammen einer theoretischen Wahrscheinlichkeitsverteilung und nicht der Untersuchung einer grossen Gesamtheit. ◇

Bei Chi-Quadrat-Anteils-Tests werden negative wie positive Abweichungen gleich behandelt, da wir das Quadrat der Abweichungen betrachten. Entsprechend führen wir immer einen rechtsseitigen Test durch. Ein linksseitiger Test würde testen, ob eine Überanpassung der Häufigkeitsverteilung der Stichprobe an die Häufigkeitsverteilung in der betrachteten Gesamtheit vorliegt - bei Zufallsstichproben ist es ja unwahrscheinlich, dass die Stichprobe genau dieselbe oder eine sehr ähnliche Häufigkeitsverteilung aufweist wie die Gesamtheit. Da das Vorzeichen der Abweichung im Test keine Rolle spielt, liegt es am Leser oder Studienverfasser zu überprüfen, in welche Richtung die Abweichung erfolgt ist. Dazu wird man oft eine Liste mit den *Residuen* erstellen. Es handelt sich um die Zahlen $k_i - k_i^e$ (= Abweichungen der Stichprobenhäufigkeiten von den erwarteten Häufigkeiten). Statt die absoluten Residuen zu betrachten, ist es allerdings besser standardisierte Residuen zu analysieren. Dies aus folgendem Grund: da wir die Häufigkeiten der Zellen als Werte von poisson-verteilten Zufallsvariablen betrachten können (s. Begründung auf S. 397), so nimmt die Varianz mit den erwarteten Häufigkeiten zu - bei der Poissonverteilung ist die Varianz gleich dem Erwartungswert! Entsprechend ist es schwierig, unmittelbar Abstände zwischen faktischen und erwarteten Häufigkeiten zu vergleichen. Es gibt verschiedene Standardisierungen. Die folgenden werden „Pearson-Residuen“ genannt (e für error):

$$e_i := \frac{k_i - k_i^e}{\sqrt{k_i^e}}$$

Für die Pearson-Residuen gilt offensichtlich, dass $\sum_i e_i^2 = X^2$. Unter der Nullhypothese sind die e_i näherungsweise normalverteilt mit dem Erwartungswert 0. Allerdings ist die asymptotische

Varianz etwas kleiner als 1 und zwar im Mittel um $\frac{M-1}{M}$ (M Anzahl der Zellen). Entsprechend liefert die Betrachtung der Pearson-Residuen einen konservativen Hinweis auf Abweichungen der faktischen von den erwarteten Häufigkeiten (die Abweichung wird zu wenig schnell als signifikant eingestuft). Die folgenden *Standardisierten Pearson-Residuen* hingegen sind asymptotisch standardnormalverteilt. Unter H_0 (Unabhängigkeit) ist

$$\frac{K_i - k_i^e}{\sqrt{k_i^e(1-p_i)}}$$

standardnormalverteilt ($p_i := \frac{k_i^e}{n}$ sind die relativen erwarteten Häufigkeiten der i -ten Zelle, n = Anzahl der Daten). Sind diese Residuen grösser als das $(1 - \frac{\alpha}{2})$ - oder das $\frac{\alpha}{2}$ -Quantil der Standardnormalverteilung, so liegt eine signifikante Abweichung vor. Im Allgemeinen kann man als kritische Grenzen bei einem 0.05-Signifikanzniveau 2 und -2 wählen. Werden vorgängig nicht entsprechenden Hypothesen aufgestellt, so sind signifikante Abweichungen allerdings nur als Hinweise für eventuelle spätere Untersuchungen zu betrachten. Zudem ist zu beachten, dass wir derart mehrere Abweichungen betrachten und damit die Fehlerwahrscheinlichkeit erster Art steigt. Für das obige Einkommensbeispiel erhält man:

	k_i	k_i^e	$k_i - k_i^e$	$\frac{k_i - k_i^e}{\sqrt{k_i^e}}$	$\frac{k_i - k_i^e}{\sqrt{k_i^e(1-p_i)}}$
k1	0.00	7.16	-7.16	-2.68	-3.01
k2	7.00	6.48	0.52	0.20	0.23
k3	6.00	7.84	-1.84	-0.66	-0.75
k4	7.00	4.99	2.01	0.90	0.98
k5	14.00	7.53	6.47	2.36	2.67

Bei den standardisierten Pearson-Residuen des Beispiels sieht man, dass es in zwei Kategorien signifikante Abweichungen gibt: in der tiefsten Einkommensklasse und in der grössten. Die tiefsten Einkommen sind in der Klasse untervertreten, die grössten Einkommen übervertreten. Zu beachten ist, dass der globale Chi-Quadrat-Test die Tabelle als solche testet. Es ist möglich, dass eine einzelne Abweichung signifikant ist, während der Globaltest nichts anzeigt.

Beispiel 20.8.5. Wir analysieren die Residuen des Beispiels 20.8.1. Wir erhalten (s. Tabelle 20.8.11):

	k_i	k_i^e	$k_i - k_i^e$	$\frac{k_i - k_i^e}{\sqrt{k_i^e}}$	$\frac{k_i - k_i^e}{\sqrt{k_i^e(1-p_i)}}$
A	30.00	30.00	0.00	0.00	0.00
B	50.00	45.00	5.00	0.75	0.89
C	30.00	52.50	-22.50	-3.11	-3.85
D	40.00	22.50	17.50	3.69	4.00

Tabelle 20.8.11: Berechnung der Residuen des Beispiels 20.8.1

Das grösste absolute Residuum $k_i - k_i^e$ liegt für das Produkt C vor. Man erhält für die Zeile B das folgende „Pearson-Residuum“:

$$e_1 := \frac{50 - 45}{\sqrt{45}} = 0.745\,36$$

Das entsprechende Standardisierte Pearson-Residuum ist

$$\frac{50 - 45}{\sqrt{45(1 - \frac{45}{150})}} = 0.890\,87$$

Bei den standardisierten Pearson-Residuen des Beispiels sieht man, dass es zwei Abweichungen der beobachteten von den erwarteten Häufigkeiten gibt, die signifikant sind (die letzten zwei). Die Werbekampagne scheint also für D Erfolg gehabt zu haben (zu Lasten des zweitletzten Produktes). $\diamond$

20.9 Tests auf Kennzahlen bei bekannter, nicht-normaler Verteilung oder ohne bekannte Verteilung (fakultativ)

Kennt man die Verteilung einer Teststatistik nicht, so kann man oft mit Computerexperimenten arbeiten. Es gibt verschiedene Varianten davon.

20.9.1 Computerexperiment bei Kenntnis der Verteilung der Zufallsvariablen unter der Nullhypothese

Ziel: Getestet wird die Abweichung des Kennwertes einer Stichprobe (= Wert der Teststatistik für die vorliegende Stichprobe) von einem vorgegebenen Parameter (Parameter der i.i.d. Stichprobenvariablen unter der Nullhypothese). Die Verteilung der Teststatistik ist unbekannt.

Voraussetzungen: Die Verteilung der Zufallsvariablen X_i , als deren Werte wir die Daten x_i auffassen, ist bekannt (Verteilung unter der Nullhypothese) und im Rahmen einer vorgegebenen Verteilungsfamilie durch die Parameter bestimmt (stetig metrisch skaliert).

Die Verteilung der Daten unter der Nullhypothese kann z.B. aus der Literatur bekannt sein (eher unwahrscheinlicher Fall) oder wir haben Daten unter der Nullhypothese, mit denen wir die Stichprobendaten vergleichen wollen. Im letzten Fall kann man die Verteilung unter der Nullhypothese mittels bereits eingeführter Verfahren schätzen: Wahl der Verteilungsfamilie mit Hilfe der empirischen Verteilungsfunktion und anschließendes Schätzen der Parameter.

Um eine Näherung an die Verteilung der Teststatistik unter der Nullhypothese zu berechnen, erstellen wir mit Excel - oder effizienter mit R - B Datensätze mit der bekannten Verteilung der Zufallsvariablen unter der Nullhypothese. Für jeden Datensatz berechnen wir die entsprechende Kennzahl (Wert der Statistik), wodurch wir eine empirische Verteilung der Statistik unter der Nullhypothese erhalten. Den Verwerfungsbereich erhalten wir, indem wir die entsprechenden Quantile des Vektors der künstlich geschaffenen Kennzahlen verwenden. Verfahren, welche die Berechnung von Statistiken und deren Verteilungen durch (Computer-)Zufallsexperimente ersetzen, nennt man nach dem bekannten Spieler-Zentrum in Monaco „Monte Carlo-Verfahren“.

Beispiel 20.9.1. *Eine Firma möchte die Varianz der Lebensdauer eines Produktes verringern. Auf Grund langjähriger und vieler Daten weiss man, dass die Lebensdauern der Produkte, die dem gegenwärtigen Produktionsprozess entstammen, exponentialverteilt sind ($\lambda = 1 \implies s^2 = 1$). Der neue Produktionsprozess führt zu einer Varianz von 0.7, wobei 50 Produkte getestet wurden. Ist der Unterschied der Varianzen signifikant?*

H_0 : Die Varianzen unterscheiden sich nicht.

H_A : Die Varianz der Stichprobe ist kleiner als die Varianz der alten Daten ($\implies$ linksseitiger Test). ($\alpha = 0.05$).

Kennen wir die Verteilung der Varianz nicht, so können wir z.B. 1000 exponentialverteilte Datensätze erstellen (siehe Excel-Datei: „Einstichprobentests.xls“, „BeiMC“, wo 1020 Sätze verwendet wurden) und für jeden Datensatz die Varianz berechnen. Im Beispiel finden wir als Näherung an $P(X \leq 0.7)$ (= die Wahrscheinlichkeit, dass die Statistik Varianz kleiner gleich 0.7 ist) die relative Häufigkeit der Varianzen, die kleiner gleich 0.7 sind, nämlich 0.245. Somit ist die Abweichung nicht signifikant. Signifikant unterschiedlich auf dem 0.05-Niveau wäre eine Varianz von ca. 0.48. Siehe R-Skript für ein effizientes Verfahren fürs Beispiel. $\diamond$

20.9.2 Computerexperiment ohne Kenntnis der Verteilung der Zufallsvariablen unter der Nullhypothese

Dabei gibt es im wesentlichen zwei Verfahren: (1) Wir schätzen die Verteilung der Stichprobendaten aus den Daten und simulieren die Verteilung der interessierenden Statistik für Datensätze, welche die geschätzte Verteilung aufweisen. (2) Wir schätzen unmittelbar eine Verteilung der interessierenden Statistik, indem wir Stichproben aus den Daten ziehen.

1. Wir schätzen die Verteilung der Daten mit Hilfe einer empirischen Verteilungsfunktion. Dazu bestimmen wir die Verteilungsfamilie und schätzen anschliessend die Parameter, um eine spezifische Verteilung zu erhalten. Wir simulieren wiederum mit einem Computer Datensätze, welche die geschätzte Verteilung der Stichprobe aufweisen. Für jeden der Datensätze berechnen wir die interessierende Kennzahl. Wir erhalten wiederum eine Verteilung von Parametern. Allerdings wird hier nicht die Verteilung der Kennzahl unter der Nullhypothese simuliert, sondern die Verteilung des Parameters der Stichprobe. Man dreht den Test also gleichsam um: Liegt die Kennzahl der Nullhypothese bei einem rechtsseitigen Test unter dem 0.05-Quantil des Kennzahlvektors? Liegt sie - bei einem linksseitigen Test - oberhalb des 0.95-Quantils? Liegt sie bei einem zweiseitigen Test unter- oder oberhalb des 0.025-Quantils beziehungsweise des 0.975-Quantils des Kennzahlvektors (Signifikanzniveau jeweils 0.05). Das Verfahren mit R verläuft analog zum Beispiel mit der Lebensdauer von Produkten.
2. Wir ziehen aus der Stichprobe der Grösse n mit Zurücklegen $B \geq 200$ Stichproben der Grösse n , berechnen jeweils den Parameter und gehen dann wie bei (2) vor. Diese Verfahren wird „Bootstrap“ genannt (s. R-Skript für eine effizientes Vorgehen). Es gibt weitere ähnliche Verfahren.

20.9.3 Übungen

1. Eine Firma untersucht, ob ein neuer Typ von Glühlampen für ihre Fabrikationsgebäude sie billiger zu stehen kommt. Die neuen Lampen konsumieren gleichviel Strom wie die alten, kosten jedoch etwas mehr. Der Lieferant verspricht für die neuen Lampen eine längere Lebensdauer. Die Firma kennt die durchschnittliche Lebensdauer der alten Glühlampensorte: 13 Monate (Standardabweichung: 2). Bevor Sie auf die neuen Lampen umsteigt, will Sie in einer der Hallen diese testen. 35 Lampen brennen durchschnittlich 14 Monate lang. Ist der Unterschied signifikant ($\alpha = 0.01$)? Wieviel darf der Preisunterschied betragen, damit sich die neuen Lampen lohnen? Wir gehen davon aus, dass sie gleichviel Energie verbrauchen.
2. Eine Eisenwarenfabrik möchte bei der Produktion von Schrauben eine tiefere Varianz der Länge der Schrauben herbringen, um den Ausschuss zu verringern. Dazu werden verschiedene neue Maschinen getestet. Die Länge von Schrauben sei normalverteilt. Die Varianz der Zufallsvariablen unter der Nullhypothese sei 0.1 cm^2 (gegeben durch die Daten einer grossen Menge von Schrauben, die mit der alten Maschine produziert wurden). Es wird eine Zufallsstichprobe von 50 Schrauben aus einem von einer neuen Maschine produzierten Haufen gezogen ($\alpha = 0.01$). Für die Messresultate siehe Exceldatei „Einstichprobentests“ (in cm). Testen Sie!
3. Ein Produktionsstätte hat bei der Schraubenproduktion für Schrauben, die genau 10 cm haben sollten, einen mittleren Wert von 9.9. Sie justiert die Maschine und zieht eine Stichprobe von 50 Schrauben (für die Messresultate siehe Exceldatei „Einstichprobentests“). Weicht die mittlere Schraubenlänge noch signifikant von 10 ab? Testen Sie ($\alpha = 0.01$). Testen Sie auch mit dem Wilcoxon-Test.
4. Der Median der Noten einer Klasse (ohne das Fach Buchhaltung) betrage 4.45. In Buchhaltung werden folgende Noten erzielt: 5.5; 4.5; 4; 3; 3.5; 3.7; 3.3; 5; 4.5; 4.7; 4.8; 4.1; 4.2; 4.3; 4.25. Passen die Daten zum Median der Gesamtnoten? ($\alpha = 0.01$)
5. Anton behauptet, 45% der Bevölkerung seien für eine bestimmte politische Vorlage. Es wird eine Zufallsstichprobe gezogen (50 Personen). Dabei sprechen sich 15 Personen für die Vorlage aus. Ist die These Antons haltbar? ($\alpha = 0.05$)
6. Es soll überprüft werden, ob die Käuferstruktur (Alter) eines Produktes der Verteilung in der Gesamtbevölkerung entspricht, um eventuell durch die gezielte Bewerbung bestimmter Altersschichten mehr an Absatz des Produktes zu erzielen. Es wird eine Stichprobe von 81 Personen gezogen. Es ergeben sich folgende Daten (s. Tabelle 20.9.12):

	Anteile Grundgesamtheit	Häufigkeiten Stichprobe
< 20	0.35	20
20 - 30	0.28	18
30 - 40	0.2	18
40 - 60	0.12	15
> 60	0.05	10
Summen	1	81

Tabelle 20.9.12: Ausgangslage für den Test der Übung

Unterscheiden sich die beiden Häufigkeitsverteilungen ? ($\alpha = 0.01$)

7. Zwecks einer Forschungsstudie wurde bei den Vertretern eines Berufes im französischsprachigen Wallis eine Umfrage durchgeführt. Allen Berufsvertretern wurde ein Fragebogen zugeschickt. Der Rücklauf betrug 20.962%. Es soll untersucht werden, ob die zurückgelaufenen Fragebogen allenfalls als Zufallsstichprobe aus der Walliser Population betrachtet werden können. Dazu werden verschiedene Anteile der Stichprobe mit den entsprechenden Anteilen der Walliser Population verglichen, unter anderem die Altersanteile „20 - 30“ sowie „älter als 30“ in der Stichprobe mit den entsprechenden Anteilen in der Walliser Population. Es ergaben sich folgende Daten (s. Tabelle 20.9.13)

Altersanteile	Fragebogen	Grundgesamtheit
20 - 30	18	74
> 30	43	217

Tabelle 20.9.13: Anteilsvergleich zwecks Überprüfung der Voraussetzung, dass Zufallsstichprobe vorliegt

Führen Sie einen Chi-Quadrat-Test durch und vergleichen Sie das Resultat mit dem exakten Binomialtest (Anteilsvergleiche Einstichprobentests). Können die zurückgelaufenen Fragebogen als Zufallsstichprobe betrachtet werden?

8. In der Gratis-Zeitung 20minuten vom 24.2.2011 steht, Menschen vom Sternzeichen Wassermann würden die besseren Autofahrer sein. Dazu werden die folgenden Daten geliefert (s. Tabelle 20.9.14):

Schadenfrequenz in %	Sternzeichen
25.0	Wassermann
27.6	Steinbock
27.7	Jungfrau
27.8	Fisch
27.8	Wider
28.6	Krebs
29.4	Löwe
29.5	Skorpion
29.6	Zwilling
29.8	Schütze
30.3	Waage
30.4	Stier

Tabelle 20.9.14: Daten Sternzeichen und Schadensfälle Autofahren

Es handelt sich um Schadensfälle in % pro Sternzeichen (100% sind die Versicherten mit dem entsprechenden Sternzeichen) einer Versicherung, die 400 000 Versicherungspolice verwal- tet. Sind die Unterschiede signifikant? Setzen Sie dabei für die Nullhypothese voraus, dass in der Bevölkerung auf die Sternzeichen eine Gleichverteilung vorliegt.

9. Lösen Sie die Aufgaben 9 bis der Exceltabelle „Einstichprobentests“
10. Finden Sie zu jeder der obigen Problemstellungen eine betriebswirtschaftliche Aufgabe und lösen Sie diese.

20.9.4 Lösungen

1. H_0 : Die durchschnittliche Lebensdauer der neuen Birnen ist nicht grösser als die der alten Birnen

H_A : Die durchschnittliche Lebensdauer der neuen Birnen ist grösser als jene der alten Birnen ($\implies$ rechtseitiger Test).

Wir nehmen an, dass die Lebensdauer der Birnen voneinander unabhängig ist und dass identische Verteilungen vorliegen. Wir können somit den Z-Test verwenden:

$$P(Z \geq \frac{14-13}{\frac{2}{\sqrt{35}}}) = 1 - \Phi\left(\frac{14-13}{\frac{2}{\sqrt{35}}}\right) = 1 - \Phi(2.958) = 1 - 0.998451721 = 0.0015483 \leq 0.01.$$

Der Unterschied in der Lebensdauer ist signifikant.

Durchschnittlich kostet eine neue Birne pro Monat $\frac{x}{14}$ und eine alte $\frac{y}{13}$ (x und y sind die Preise der jeweiligen Birnen). Die neuen Birnen lohnen sich, wenn $\frac{x}{14} < \frac{y}{13} \implies \frac{x}{y} < \frac{14}{13}$ (da $y > 0$). Somit muss das Preisverhältnis $\frac{x}{y} < \frac{14}{13} = 1.0769$ sein (die neue Birne darf nicht mehr als 7.69% teurer sein).

2. H_0 : Die Varianzen der Stichprobe ist nicht kleiner als die Varianz s_o^2 (der durch die alte Maschine produzierten Schrauben).

H_A : Die Varianz der Stichprobe ist kleiner als s_o^2 ($\implies$ linksseitiger Test).

Wir überprüfen die Daten auf Normalverteilung (mit P-P-Plot; s. Abbildung 20.9.8):

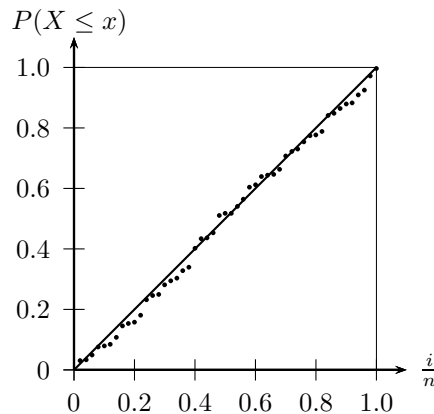


Abbildung 20.9.8: P-P-Plot der Daten der Übung; graphische Überprüfung auf Normalverteilung mit aus den Daten geschätzten Parametern

Die Daten sind deutlich normalverteilt

Die Varianz der Stichprobe beträgt: $s^2 = 0.010117436$

Wir erhalten: $\frac{S^2}{s_0^2}(n-1)$ nimmt den Testwert $\frac{0.010117436}{0.1}(50-1) = 4.9575$ an.

Da wir linksseitig testen, ist der Testwert $P(X \leq 4.9575) = \text{chiqu.vert}(4.9575; 49; 1) = 1.37061 \cdot 10^{-16} \leq 0.1$.

Die Differenz ist signifikant.

3. H_0 : Der neue Mittelwert weicht nicht von 10 ab.

H_A : Der neue Mittelwert weicht von 10 ab ($\implies$ zweiseitiger Test).

Da wir die Varianz der Grundmenge nicht haben, müssen wir den t -Test verwenden.

Wir überprüfen die Daten auf Normalverteilung (mit P-P-Plot; s. Abbildung 20.9.9).

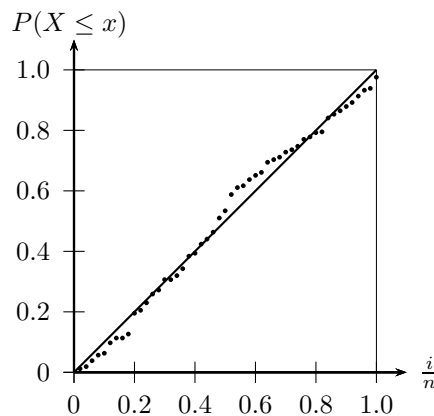


Abbildung 20.9.9: P-P-Plot auf Normalverteilung der Daten mit aus den Daten geschätzten Parametern

Die Daten können als normalverteilt betrachtet werden.

Die Varianz der Stichprobe beträgt: $s^2 = 0.001652244$; der Mittelwert: 9.99.

Die Zufallsvariable $\frac{\bar{X} - \bar{x}_0}{\frac{s}{\sqrt{n}}}$ nimmt im Beispiel als Testwert den folgenden Wert an: $\frac{9.99 - 10}{\frac{\sqrt{0.001652244}}{\sqrt{50}}} =$

-1.7396 ; $df = 49$, p-Wert zweiseitig $= 2 * \min\{P(X \leq -1.7396); P(X \geq -1.7396)\} = 2 \cdot t.\text{vert}(-1.7396; 49; 1) = 0.088207941$

(offensichtlich ist $P(X \leq -1.7396) < P(X \geq -1.7396)$!!!!)

Die Abweichung von 10 ist nicht mehr signifikant.

Mit dem Wilcoxon-Test: Rangsumme: $\frac{50 \cdot 51}{2} = 1275$; Summe der Ränge der negativen Differenzen: 781. Entsprechend ist die Summe der Ränge der positiven Differenzen: $1275 - 781 = 494$

$$\frac{781 - \frac{50 \cdot 51}{4}}{\sqrt{\frac{50 \cdot 51 \cdot (2 \cdot 50 + 1)}{24}}} = 1.3852$$

 $\Phi(1.3852) = 0.91700434$ p-Wert = $2 \min\{0.91700434, 1 - 0.91700434\} = 0.165991321$. Im Beispiel unterscheiden sich die beiden p-Werte stark.

4. Noten sind ordinalskaliert. Somit kommen der z -Test und der t -Test nicht in Frage.
 H_0 : Die Buchhaltungsnoten sind nicht tiefer als die übrigen Noten.
 H_A : Die Buchhaltungsnoten liegen tiefer.
 Wir verwenden den Vorzeichentest: Wir zählen die Noten, die kleiner als 4.45 sind. Es sind dies 9 auf 15. Wir berechnen für $X \sim B(15, 0.5)$: $P(X \geq 9) = 1 - P(X < 9) = 0.303619385 > 0.01$ (rechtsseitiger Test). Die Buchhaltungsnoten unterscheiden sich nicht-signifikant von den übrigen Noten.
5. Wir testen zweiseitig:
 H_0 : Die Zufallsstichprobe passt zu einem Anteil von 0.45.
 H_A : Die Zufallsstichprobe passt nicht zu einem Anteil von 0.45.
 Wir erhalten: $P(X \geq 15) = 1 - 0.010383581$ für $X \sim B(50; 0.45)$
 Wir erhalten: $P(X \leq 15) = 0.02195068$.
 Der p-Wert ist somit: $2 \cdot 0.02195068 = 0.043901 < 0.05$.
 Die Abweichung ist signifikant. Die Erhebungsergebnisse sind mit der Meinung Antons nicht verträglich.
6. H_0 : Die Käuferstruktur in der Stichprobe entspricht der Häufigkeitsverteilung der betrachteten Population
 H_A : Die Käuferstruktur in der Stichprobe entspricht nicht der Häufigkeitsverteilung der betrachteten Population.
 Wir rechnen die Prozente in absolute Häufigkeiten bezüglich der Stichprobengröße um und berechnen den Testwert (s. Tabelle 20.9.15):

	Anteile Grundgesamtheit	Häufigkeiten Stichprobe	umgerechnete Häufigkeiten Grundgesamtheit	quadrierte gewichtete Abweichungen
< 20	0.35	20	28.35	2.459347443
20 - 30	0.28	18	22.68	0.965714286
30 - 40	0.2	18	16.2	0.2
40 - 60	0.12	15	9.72	2.868148148
> 60	0.05	10	4.05	8.741358025
Summen	1	81	81	15.2345679

Tabelle 20.9.15: Tabelle zur Übung

$$1 - F_{\chi^2_4}(15.2345679) = 1 - \text{chiqu.vert}(15.2345679; 4; 1) = 0.004238635 < 0.05.$$

Die Käuferstruktur entspricht nicht der Häufigkeitsverteilung der betrachteten Population.

Mit R:

```
hau=c(20,18,18,15,10)
```

```
p=c(0.35,0.28,0.2,0.12,0.05)
```

```
test=chisq.test(hau,p=p)
```

```
ergibt: Chi-squared test for given probabilities
```

```
data: hau
```

X-squared = 15.235, df = 4, p-value = 0.004239

Residuenanalyse:

```
tab=cbind(test$observed,test$expected,test$observed-test$expected, test$residuals,test$stdres)
```

```
colnames(tab)=c("beob","erwart","abs.Resid","Pearson-Resid","stand.Resid")
```

```
tab
```

ergibt:

	beob	erwart	abs.Resid	Pearson-Resid	stand.Resid
[1,]	20	28.35	-8.35	-1.5682307	-1.945151
[2,]	18	22.68	-4.68	-0.9827076	-1.158132
[3,]	18	16.20	1.80	0.4472136	0.50000
[4,]	15	9.72	5.28	1.6935608	1.80534
[5,]	10	4.05	5.95	2.9565788	3.033386

Es gibt eine signifikante Abweichung bei der 5. Kategorie (reproduzieren sie die Tabelle mit Excel!).

7. H_0 : Die Anteile der Altersgruppen in der Stichprobe entspricht dem Altersanteil in der betrachteten Population.

H_A : Die Anteile der Altersgruppen in der Stichprobe entspricht dem Altersanteil in der betrachteten Population nicht.

Wir errechnen die erwarteten Häufigkeiten und den Testwert (s. Tabelle 20.9.16):

Alter	Häufigkeiten Fragebogen	Häufigkeiten Grundgesamtheit	Anteile Grundgesamtheit	auf Stichproben- grösse umgerechnete Grundgesamtheit	gewichtete quadrierte Abweichungen
20 - 30	18	74	0.254295533	15.51202749	0.399045657
> 30	43	217	0.745704467	45.48797251	0.136080086
Summen	61	291	1	61	0.535125743

Tabelle 20.9.16: Lösungstabelle für die Übung

$$1 - F_{\chi^2_1}(0.53512574) = 1 - \text{chiqu.vert}(0.53512574; 1; 1) = 0.464459828 > 0.05.$$

Auf Grund dieses Vergleichs kann man die zurückgelaufenen Fragebogen als Zufallsstichprobe betrachten.

Mit dem Binomialtest: Der Anteil der 20-30-jährigen an der Berufsgruppe beträgt 25.4295533%.

Damit ergibt sich für $X \sim B(61; 0.254295533)$ (zweiseitiger Test):

$$2 * \min(P(X \geq 18), P(X \leq 18)) = 0.5485132$$

Die beiden p-Werte unterscheiden sich somit um $0.5485132 - 0.464459829 = 0.0840533$ (8%).

Es ist besser, bei 2X1-Tabellen den exakten Test zu verwenden.

8. Lösungshinweise. Die erwarteten Häufigkeiten berechnet man, indem man den Anteil a der Schadensfälle an der Gesamtpopulation berechnet und dann allen Kategorien dieselben absoluten erwarteten Häufigkeiten zuordnet ($400000a/12$). Für die Lösung s. Exceltabelle.

20.10 Lernziele

- Den Begriff des „Einstichprobentests“ kennen und die Situation beschreiben können, für die Einstichprobentests verwendet werden. Folgende Einstichprobentests situationsgerecht auswählen und durchführen können: z- Test, t-Test, Wilcoxon-Test, Vorzeichentest, Test auf abweichende Varianz, Test auf Anteile (Binomialtests und Chi-Quadrat-Tests).
- Begriff der „Mächtigkeit eines Tests“ kennen und wissen, welche Tests mächtiger sind (wenn mehrere Tests anwendbar sind).

- Übungen von der Art des Übungsblocks lösen können (von Hand, mit Excel, mit einer Statistiksoftware).
- Test auf abweichende Varianz - Herleitung verstehen und selber durchführen können (im Allgemeinen müssen Sie ja Herleitungen nicht kennen. Es lohnt sich aber, an einem einfachen Beispiel die Grundgedanken der Testtheorie nochmals zu wiederholen und im Detail zu verstehen).

Kapitel 21

Verbundene Zweistichprobentests

Von *verbundenen Stichproben* (man spricht auch von „gepaarten Stichproben“) reden wir, wenn zwei Stichproben vorliegen und Messungen x_i und y_i beider Stichproben $\mathbf{x}$ und $\mathbf{y}$ jeweils zu denselben Objekten i gehören (z.B. 17 Personen werden darauf hin untersucht, wie sie auf Alkohol reagieren, indem die jeweilige Reaktionszeit in Sekunden gemessen wird. Für jede der 17 Personen wird die Reaktionszeit vor und nach dem Alkoholgenuss gemessen, wobei jeder Person $0.0005 \cdot G$ (G = Gewicht in Kilogramm) in Gramm Alkohol verabreicht wird). Die ersten drei der folgenden Tests können auf den Einstichprobenfall zurückgeführt werden, indem man statt der Daten die Differenz der beiden Stichproben verwendet. Trotzdem behandeln wir diese drei Tests ausführlich, um weitere Beispiele zu haben und um das Verständnis zu fördern.

21.1 t-Test

Ziel: Die Mittelwerte zweier verbundene Stichproben werden verglichen, um zu entscheiden, ob sie als Werte einer Statistik mit derselben Verteilung zu betrachten sind.

Voraussetzungen: Die Differenzen der Daten sind unabhängig und identisch verteilt. Die Varianz der Differenzen wird aus den Daten geschätzt. Die Differenzen müssen normalverteilt sein (metrisch skalierte Variablen)

Wir können diesen Fall auf den Einstichprobenfall zurückführen, indem wir die Differenzen der gepaarten Daten als Werte von Zufallsvariablen betrachten und den Parameter unter der Null-Hypothese mit 0 identifizieren. Wir erhalten unter der Voraussetzung normalverteilter Differenzen die t_{n-1} - verteilte Teststatistik:

$T = \frac{\bar{D}}{\frac{s_d}{\sqrt{n}}}$ ist t_{n-1} -verteilt.
$\bar{D}$: Zufallsvariable mit dem Mittelwert $\bar{d}$ der Differenzen d_i als Wert
s_d : Zufallsvariable mit der empirischen Standardabweichung s_d der Differenzen als Wert
n : Anzahl Differenzen

Beispiel 21.1.1. Wir möchten überprüfen, wie sich Alkoholkonsum auf die Reaktionsfähigkeit auswirkt, wobei wir testen wollen, ob diese durch Alkoholkonsum länger wird. Wir nehmen an, es würden sich für die Personen die folgenden Reaktionszeiten vor und nach dem Alkoholgenuss ergeben: (s. Tabelle 21.1.1; die Daten finden Sie auch unter „VerbundeneZweistichproben.xls“). Wir berechnen die Differenzen „nachher - vorher“ (umgekehrt ist auch möglich, dann kehren sich die Vorzeichen der Differenzen und die „Seitigkeit“ des Testes um, s. unten). Mit der vorliegenden Berechnung der Differenzen „nachher - vorher“ liegt ein rechtsseitiger Test vor.

Proband	vorher	nachher	Differenzen
1	1.009445701	1.889979837	0.880534136
2	1.143801786	2.11894739	0.975145604
3	1.135230599	2.01103206	0.875801461
4	0.944542635	1.797204105	0.85266147
5	1.068052231	1.93721726	0.869165029
6	0.859071169	1.668888872	0.809817703
7	0.873441448	1.720319665	0.846878217
8	1.075975254	1.965755195	0.889779941
9	1.019241093	1.892808534	0.873567442
10	0.881345433	1.828933293	0.94758786
11	1.003952275	1.910366456	0.906414181
12	0.954190138	1.853237182	0.899047044
13	1.055894064	1.977400597	0.921506533
14	1.018117503	1.882383448	0.864265944
15	1.229650215	2.136534769	0.906884554
16	0.90778402	1.813363515	0.905579494
17	0.952582039	1.887155042	0.934573003

Tabelle 21.1.1: Beispieldaten für verbundenen Zweistichprobentest

Wir müssen die Differenzen auf Normalverteilung prüfen (s. Abbildung 21.1.1).

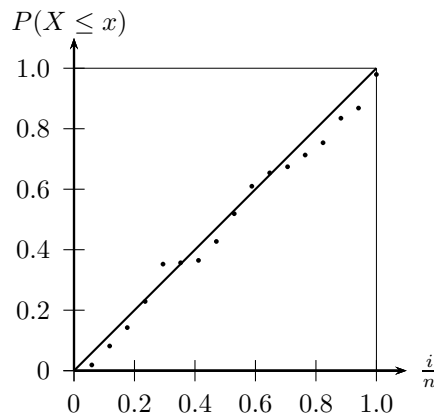


Abbildung 21.1.1: P-P-Plot der Differenzen von Tabelle 21.1.1 bei Vermutung von Normalverteilung mit aus Differenzen geschätzten Parametern

Die Differenzen können als normalverteilt angesehen werden.

Der Mittelwert der Differenzen beträgt: 0.891718213

Die Standardabweichung der Differenzen beträgt: 0.040089677

H_0 : Die Reaktionszeit verändert sich im Mittel nicht.

H_A : Die Reaktionszeit ist im Mittel länger. ($\alpha = 0.05$, rechtsseitiger Test)

Wir berechnen: Die Teststatistik $\frac{\bar{D}}{s_d/\sqrt{n}}$ nimmt den Testwert $\frac{0.891718213}{\frac{0.040089677}{\sqrt{17}}} = 91.711$ an:

p -Wert: $1 - F_{t_{16}}(91.711) \approx 1.65978E - 23$ (mit Excel: $=1-t.vert(91.711;16;1)$). Die Nullhypothese kann verworfen werden.

Zählen wir „vorher-nachher“ ab, so erhalten wir als Mittel der Differenzen -0.891718213 und es liegt ein linksseitiger Test vor - die Reaktionszeit ist ja dann länger, wenn wir bei dieser Art abzuzählen im Mittel eine negative Zahl erhalten. Wir erhalten denselben p -Wert: $F_{t_{16}}(-91.711) \approx 1.65978E - 23$ (mit Excel: $=t.vert(-91.711;16;1)$). $\diamond$

21.2 Wilcoxon-Test

Ziel: Zwei verbundene Stichproben werden verglichen, um zu entscheiden, ob sich die Mediane der beiden Datensätze unterscheiden.

Voraussetzungen: Die Differenzen der Daten sind unabhängig und identisch verteilt. (metrisch skalierte Variablen) (es sollte nicht zu viele gleiche Differenzen geben und diese Gruppen von Differenzen sollten nicht zu gross sein). Die Differenzen sollten symmetrisch um ihren Median liegen (bei symmetrischen Verteilungen stimmt der Median allerdings mit dem Mittelwert überein).

Wir ordnen den Betrag der Differenzen der Grösse nach. 0-Differenzen lassen wir weg. In einem ersten Durchlauf numerieren wir die verbleibenden Beträge der Grösse nach. In einem zweiten Durchlauf ordnen wir gleichen Beträgen (= verbundenen Rängen) das arithmetische Mittel der zugeordneten Nummern zu (siehe analoges Verfahren bei der Berechnung der Spearman-Korrelation).

Beispiel 21.2.1. *Es seien die folgenden Daten gegeben (z.B. Umsatz von Personen vor und nach einer Fortbildung, s. Tabelle 21.2.2):*

Person	vorher	nachher
1	2	7
2	15	22
3	5	7
4	12	11
5	5	3
6	7	8
7	9	13
8	12	5
9	15	20
10	6	9
11	6	12
12	2	10

Tabelle 21.2.2: Daten für Wilcoxon-Test

Wir berechnen die Differenzen z.B. nachher-vorher, den Betrag der Differenzen, ordnen den Betrag der Differenzen, ordnen Ränge zu und berechnen die Rangsumme der positiven (oder negativen) Differenzen - die Summe der beiden Rangsummen entspricht der Summe aller Ränge, also $\frac{n(n+1)}{2}$ mit n = Anzahl Daten. Damit kann nach der Berechnung einer der Rangsummen die andere leicht berechnet werden (s. Tabelle 21.2.3):

Person	vorher	nachher	Differenzen	Betrag der Differenzen	Ränge des Betrags der Differenzen	Ränge negativer Differenzen
4	12	11	-1	1	1.5	1.5
6	7	8	1	1	1.5	0
3	5	7	2	2	3.5	0
5	5	3	-2	2	3.5	3.5
10	6	9	3	3	5	0
7	13	9	4	4	6	0
1	2	7	5	5	7.5	0
9	15	20	5	5	7.5	0
11	6	12	6	6	9	0
2	15	22	7	7	10.5	0
8	12	5	-7	7	10.5	10.5
12	2	10	8	8	12	0

Tabelle 21.2.3: Berechnung der Ränge der negativen Differenzen

Das Beispiel enthält für die Anwendung des Wilcoxon-Testes zu viele verbundenen Ränge. Es ging hier aber um ein Beispiel für die Berechnung der Ränge bei verbundenen Rängen. Wir nehmen an, wir möchten beim obigen Beispiel testen, ob durch die Weiterbildung der Umsatz gestiegen ist. Da wir „nachher-vorher“ gerechnet haben, trifft die Hypothese zu, wenn eine kleine Rangsumme der negativen Differenzen haben und eine grosse der positiven Differenzen (mit $\alpha = 0.025$). Bezüglich der Summe der Ränge der positiven Differenzen liegt also ein rechtsseitiger Test vor. Dabei gilt: die Rangsumme der positiven Differenzen ist genau dann gross, wenn die Rangsumme der negativen Differenzen klein ist (da die Summe der beiden konstant ist!). In der Tabelle mit den kritischen Werten werden deshalb nur die kleinen kritischen Werte angegeben, da die Verteilung der Teststatistik Rangsumme symmetrisch ist. Beim obigen Beispiel erhalten wir eine Rangsumme der negativen Ränge von 15.5 ($n = 12$, Rangsumme der positiven Differenzen ist also $\frac{12 \cdot 13}{2} - 15.5 = 62.5$). Der kritische Wert ist 14. Es gilt $15.5 > 14$. Entsprechend ist das Ergebnis auf dem fixierten Signifikanzniveau nicht signifikant. Hätten wir „vorher-nachher“ gerechnet, würde die Hypothese bestätigt, wenn wir eine grosse Rangsumme der negativen Differenzen erhalten. $\diamond$

Beispiel 21.2.2. Die SUVA macht eine Unfallverhütungskampagne in Betrieben und möchte deren Wirkung überprüfen. In 21 zufällig ausgewählten Betrieben wird die Unfallhäufigkeit vor und nach der Kampagne während eines Jahres erhoben.

H_0 : Die Kampagne hat nichts bewirkt.

H_A : Die Kampagne hat die Unfallhäufigkeit gesenkt ($\implies$ linksseitiger Test). $\alpha = 0.05$

Bei der Erhebung ergaben sich folgende Daten (s. Tabelle 21.2.4):

Betrieb	vorher	nachher
1	7	2
2	22	15
3	7	5
4	11	12
5	3	5
6	8	7
7	13	9
8	5	12
9	20	15
10	9	6
11	12	6
12	10	2
13	11	14
14	12	20
15	15	10
16	22	21
17	18	15
18	3	13
19	5	11
20	30	20
21	15	3

Tabelle 21.2.4: Frei erfundene Beispieldaten für einen Wilcoxon-Test

Wenn die Kampagne die Unfallhäufigkeit senkt, muss die Anzahl der positiven Differenzen die Anzahl der negativen Differenzen übersteigen. Je grösser die Abweichung der Unfallhäufigkeiten ist, desto grösser werden die positiven Differenzen ausfallen und desto kleiner die negativen Differenzen. Dies wird sich auf die entsprechenden Rangsummen auswirken.

Wir ordnen die Beträge der Differenzen und erhalten die folgende Rangsumme der positiven Differenzen: 151.5. Um die Tabelle verwenden zu können, berechnen wir $\frac{21 \cdot 22}{2} - 151.5 = 79.5$ (= Testwert der Teststatistik R^-). Wäre die Kampagne wirkungslos gewesen, müsste sich für die Summe der Ränge der positiven und damit der negativen Differenzen ein Wert geben, der nicht zu stark von $\frac{231}{2} = 115.5 = E(R^i)$ abweicht. Je mehr die Summe der Ränge der positiven (negativen) Differenzen davon abweicht, desto unwahrscheinlicher ist das Ergebnis unter der Nullhypothese. Laut Tabelle ist 59 der kritische Wert für den einseitigen Test mit $\alpha = 0.025$. Entsprechend ist eine Rangsumme von 79.5 nicht genügend. Die Nullhypothese kann nicht abgelehnt werden. $\diamond$

21.3 Vorzeichentest

Ziel: Zwei verbundene Stichproben werden verglichen, um zu entscheiden, ob sie sich unterscheiden.
Voraussetzungen: Die Differenzen der Daten sind unabhängig und identisch verteilt. (metrisch oder ordinal skalierte Variablen)

Der Vorzeichentest berücksichtigt im Gegensatz zum Wilcoxon-Test nicht, wie stark die Abweichungen der Differenzen vom Median sind. Allerdings setzt er auch nicht voraus, dass die Daten metrisch skaliert sind. Bei metrisch skalierten Daten berechnen wir beim Vorzeichentest die Differenz der Daten und zählen die negativen oder positiven Vorzeichen. Dieses Verfahren können wir auch bei ordinalskalierten Daten anwenden, obwohl die Differenzbildung für solche Daten eigentlich keinen Sinn macht. Es geht uns ja nur um das Vorzeichen der Abweichung. Unter der Nullhypothese setzen wir eine binomialverteilte Teststatistik ($p = 0.5$) voraus. Sind Differenzen mit 0 identisch, so lassen wir solche Objekte (Datenpaare) weg. Solche Differenzen werden „Bindungen“ genannt.

Entsprechend reduziert sich n in $B(n, 0.5)$ um die Anzahl der Bindungen.

$X \sim B(n, 0.5)$
n : Anzahl der von 0 verschiedenen Differenzen

Beispiel 21.3.1. 10 Personen werden darauf hin untersucht, wie ihre Konzentrationsfähigkeit bei der Arbeit von Musik abhängt. Die Konzentrationsfähigkeit wird mit in Geld messbarem Output pro Tag gemessen. An einem Tag arbeiten die Personen mit Musik ihrer Wahl, am anderen Tag ohne Musik (dabei ist auf Gleichheit der übrigen Bedingungen zu achten. Es wäre ungünstig, eine Messung bei einem Azorenhoch und die andere an einem Föhntag vorzunehmen!). Wir nehmen an, wir würden die folgenden Daten erhalten (s. Tabelle 21.3.5; die Daten finden Sie auch unter "VerbundeneZweiStichproben.xls").

mit Musik	ohne Musik	Differenzen
349	365	16
358	370	12
364	365	1
351	365	14
342	350	8
403	390	-13
412	400	-12
377	380	3
374	380	6
381	370	-11

Tabelle 21.3.5: Daten für verbundenen Zweistichprobentest

Da wir nicht wissen, ob sich Musik positiv oder negativ auf die Konzentrationsfähigkeit (so wie sie gemessen wird) auswirkt, machen wir einen zweiseitigen Test:

H_0 : Musik wirkt sich nicht auf die Konzentration aus.

H_A : Musik macht einen Unterschied ($\Rightarrow$ zweiseitiger Test) ($\alpha = 0.05$)

Da die Daten metrisch skaliert sind, denken wir zuerst an einen t -Test. Deshalb überprüfen wir die Differenzen auf Normalverteilung:

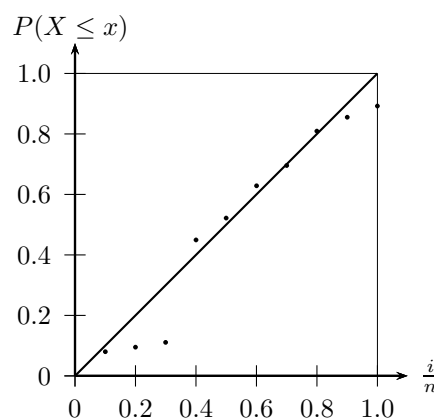


Abbildung 21.3.2: P-P-Plot der Differenzen von Tabelle 21.3.5 bei Vermutung von Normalverteilung mit aus Differenzen geschätzten Parametern

Wir kommen zum Schluss, dass die Differenzen eventuell nicht normalverteilt sind (wenig Daten!). Deshalb verwenden wir sicherheitshalber einen Vorzeichentest. Wir haben 3 negative Differenzen (auf 10). Damit nimmt die Teststatistik den Testwert 3 an. Als p -Wert erhalten wir

somit: $2 \cdot \min\{P(X \leq 3), P(X \geq 3)\} = 2 \cdot P(X \leq 3) = 2 \cdot 0.171875 = 0.34375 > 0.05$ (oder $0.171875 > 0.025$). Die Nullhypothese kann nicht verworfen werden. $\diamond$

Beispiel 21.3.2. Wir überprüfen, ob sich die Probanden mit oder ohne Musik bei der Arbeit subjektiv besser fühlen. Durch Befragung haben wir folgende Resultate erhalten. (Ich fühle mich: sehr gut (5), gut (4), ziemlich gut (3), ungut (2), schlecht (1)) (s. Tabelle 21.3.6).

Proband	mit Musik	ohne Musik
1	4	3
2	1	2
3	2	4
4	3	1
5	2	1
6	5	2
7	5	4
8	2	3
9	3	1
10	3	5
11	3	1
12	3	2

Tabelle 21.3.6: Daten für verbundenen Zweistichprobentest

H_0 : Musik hat keinen Einfluss auf die Befindlichkeit. H_A : Musik hat einen positiven Einfluss auf die Befindlichkeit ($\implies$ rechtsseitiger Test) ($\alpha = 0.05$)
Es ergeben sich die folgenden Resultate (+ für "es geht besser mit Musik" und - für "es geht besser ohne") (s. Tabelle 21.3.7).

Proband	mit Musik	ohne Musik	Vorzeichen der Änderung
1	4	3	+
2	1	2	-
3	2	4	-
4	3	1	+
5	2	1	+
6	5	2	+
7	5	4	+
8	2	3	-
9	3	1	+
10	3	5	-
11	3	1	+
12	3	2	+

Tabelle 21.3.7: Vorzeichen der Änderung von Tabelle 21.3.6

8 auf 12 Personen fühlen sich mit Musik bei der Arbeit besser, 4 hingegen schlechter. Die Teststatistik nimmt den Testwert 8 an. Wir erhalten als p-Wert: $P(X \geq 8) = 0.193847656$ für $X \sim B(12, 0.5)$. Der p-Wert ist grösser als das Signifikanzniveau. Das Resultat ist nicht signifikant. $\diamond$

21.4 McNemar-Test

Ziel: Zwei verbundene Stichproben werden verglichen, um zu entscheiden, ob sich eine signifikante Änderungen der absoluten Häufigkeiten ergeben haben (Variablen mit zwei Ausprägungen).

Voraussetzungen: i.i.d. Zufallsvariablen

Wir betrachten ein

Beispiel 21.4.1. *Es wird eine Zufallsstichprobe gezogen, um die Wirksamkeit einer Werbekampagne zu überprüfen. Dieselbe Stichprobe wird vor und nach der Werbekampagne untersucht. Dabei ergibt sich, dass vor der Werbekampagne 84 Personen das Produkt kauften, die es auch nachher noch kauften. 26 Personen kauften nach der Kampagne das Produkt, die es vorher nicht kauften. 4 Personen, die das Produkt vor der Kampagne kauften, kauften es nach der Kampagne nicht mehr. 90 Personen kauften das Produkt vorher und nachher nicht. War die Werbekampagne wirksam? Man erhält die folgende Kreuztabelle:*

		Nachher	
		Käufer	Nichtkäufer
vorher	Käufer	84	4
	Nichtkäufer	26	90

◇

Im McNemar-Test werden nur die Fälle berücksichtigt, bei denen es eine Änderung gibt: Im Beispiel sind dies 26 und 4 Fälle. Die Anzahl der sich ins positive (oder negative) ändernden Fälle sind als Werte einer Zufallsvariablen X zu betrachten, die unter der Nullhypothese - keine Veränderung - binomialverteilt ist: $X \sim B(n, 0.5)$, n = Anzahl der sich ändernden Fälle:

$X \sim B(n, 0.5)$
n : Anzahl der Objekte, bei denen eine Änderung erfolgte.

Im Beispiel: $n = 30$ und der Testwert 4 (einseitiger Test);

p-Wert: $P(X \leq 4) = \sum_{i=0}^4 \binom{30}{i} 0.5^i 0.5^{30-i} = 2.9738 \times 10^{-5} \leq 0.05$. Zu überprüfen ist die

Richtung der Abweichung (mehr positive als negative Änderungen oder umgekehrt. Im Beispiel gibt es mehr Personen, die nachher kaufen und vorher nicht kauften als umgekehrt. Es gibt also signifikant mehr Personen, die nachher kaufen und die vorher nicht kauften als Personen, die vorher kauften und nachher nicht mehr kauften).

Bemerkung 21.4.2. *Oft wird für den McNemar-Test die folgende Teststatistik angegeben:*

$$\frac{(X - Y)^2}{X + Y} \text{ ist chi-quadrat-verteilt mit einem Freiheitsgrad}$$

wobei X die Zufallsvariable ist, welche die erste Anzahl sich ändernder Objekte als Wert annimmt und Y die Zufallsvariable ist, welche die zweite Anzahl sich ändernder Objekte als Wert annimmt. Es handelt sich um eine Näherung, die mit den heutigen technischen Möglichkeiten an Relevanz verloren hat. Allerdings lässt sich der χ_1^2 -McNemar-Test auf den Bowker-Test für $m \times n$ -Kreuztabellen erweitern (s. [Bortz \(2008\)](#), S. 165). Für den χ_1^2 -McNemar-Test wird oft eine Teststatistik mit Diskretheitskorrektur verwendet:

$$\frac{(|X - Y| - 1)^2}{X + Y} \text{ ist chi-quadrat-verteilt mit einem Freiheitsgrad}$$

◇

21.4.1 Übungen

- Wir nehmen an, bei der Verabreichung eines Medikamentes würden sich die Reaktionszeiten wie folgt verändern (s. Tabelle 21.4.8; Daten: VerbundeneZweiStichproben.xls) Testen Sie, ob die Verabreichung des Medikamentes einen Einfluss hat ($\alpha = 0.05$).

Proband	Reaktionszeit vorher	Reaktionszeit nachher
1	1.449261794	2.537742155
2	1.518605621	2.511347788
3	1.549301781	2.399154523
4	1.511863908	2.472282873
5	1.554923399	2.29781149
6	1.7348597	1.61129094
7	1.357404224	2.732258003
8	1.567440311	2.529558664
9	1.563293219	2.507353587
10	1.398282057	2.607298354
11	2.1989989	1.8989809
12	1.480371925	2.396040653
13	1.635326175	2.298971185
14	1.441021049	2.267727047
15	1.490376591	2.355039862

Tabelle 21.4.8: Beispieldaten der Übung

- In einem Lieferdienst wird der Fahrzeugpark bezüglich der Effizienz zweier Schmieröle überprüft (Anzahl gefahrener Kilometer pro Liter Schmieröl). Für die Daten siehe "VerbundeneZweiStichproben.xls". Testen Sie, ob sich die Schmieröle unterscheiden ($\alpha = 0.05$).
- Die Suva macht eine Kampagne für Unfallverhütung. Im Berufsleben sollen Leitern nur mehr gesichert erstiegen werden. Eine Stichprobe von 200 Personen wird vor und nach der Kampagne auf ihr Verhalten untersucht. 120 Personen haben ihr Verhalten nicht geändert. 60 verwenden zusätzlich Sicherungsmassnahmen bei der Verwendung von Leitern, 20 verwenden nach der Kampagne keine solche Sicherungsmassnahmen mehr (was nicht unbedingt der Kampagne zuzuschreiben ist, sondern vielleicht einer gewissen Nachlässigkeit, die ohne Kampagne vielleicht noch grösser gewesen wäre). War die Kampagne wirksam? ($\alpha = 0.05$)
- Denken Sie sich für jede der obigen Fragestellungen ein betriebswirtschaftliches Problem aus und lösen Sie es.

21.4.2 Lösungen

- H_0 : Das Medikament hat keinen Einfluss.
 H_A : Das Medikament hat einen Einfluss ($\implies$ zweiseitiger Test). ($\alpha = 0.05$) Wir möchten einen t-Test anwenden und überprüfen deshalb zuerst die Normalverteilungsannahme (s. Abbildung 21.4.3):

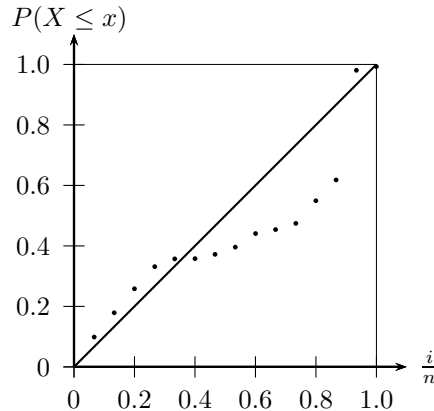


Abbildung 21.4.3: P-P-Plot der Differenzen bei Vermutung von Normalverteilung mit aus den Differenzen geschätzten Parametern

Die Daten sind kaum normalverteilt. Entsprechend müssen wir einen Vorzeichen oder eine Rangsummentest durchführen (s. Exceltabelle). Es gibt zwei positive Differenzen mit den Rängen 1 und 2.

Die Teststatistik nimmt damit den Testwert 3 an (kleinere Rangsumme): Dieser Wert liegt weit unter jeglichen kritischen Werten für $n = 15$ (nämlich 25, 20, 16).

Wir führen noch den Vorzeichentest durch. Mit Hilfe des Vorzeichentests erhalten wir: 2 (= Testwert) auf 15 sind positiv. Dies ergibt den p-Wert (für $X \sim B(15, 0.5)$): $P(X \leq 2) = 0.003692627 < 0.025$

Somit ist das Ergebnis signifikant.

2. H_0 : Die beiden Öle unterscheiden sich nicht.

H_A : Die beiden Öle unterscheiden sich. ($\implies$ zweiseitiger Test).

Wir denken an einen t-Test und überprüfen die Differenzen auf Normalverteilung (s. Abbildung 21.4.4).

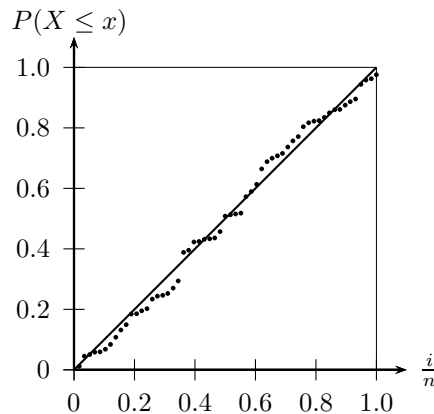


Abbildung 21.4.4: P-P-Plot der Differenzen auf Normalverteilung mit aus den Differenzen geschätzten Parametern

Die Normalverteilungsannahme ist offenbar haltbar. Wir können somit den t-Test anwenden. Wir berechnen die Differenzen, den Mittelwert der Differenzen und deren Standardabweichung: wir erhalten als Wert der Zufallsvariable „Mittelwert der Differenzen“ $\bar{d} = -5.018708587$ sowie als Wert der Zufallsvariable „Standardabweichung“ $s_d = 3.832234822$. Die Teststatistik T nimmt somit den Wert $\frac{-5.018708587}{\frac{3.832234822}{\sqrt{58}}} = -9.9736$ an.

p -Wert: $2 \cdot \min\{F_{t_{57}}(-9.9736), 1 - F_{t_{57}}(-9.9736)\} = 2 \cdot F_{t_{57}}(-9.9736) = 4.1426E-14 < 0.05$.
Das Testergebnis ist hoch signifikant.

3. H_0 : Die Kampagne zeigt keine positive Wirkung.

H_A : Die Kampagne bewirkt eine positive Änderung (einseitiger Test)

$X \sim B(80, 0.5)$:

$$P(X \leq 20) = \sum_{i=0}^{20} \binom{80}{i} 0.5^i 0.5^{80-i} = 4.2903 \times 10^{-6} < 0.05.$$

Die Nullhypothese kann verworfen werden, da signifikant mehr Personen nach der Kampagne die Sicherheitsvorschriften beachten als umgekehrt.

21.5 Lernziele

- Die obigen Tests anwenden können
- Für konkrete Beispiele die richtigen Tests auswählen können
- Übungen der obigen Art lösen können

Kapitel 22

Unverbundene Zweistichprobentests

Es werden zwei Stichproben gezogen, so dass die Daten unverbunden sind, d.h. die Daten der zwei Stichproben gehören nicht zu jeweils gleichen Objekten. Dabei wird überprüft, ob es haltbar ist, dass die zwei Stichproben als Realisierungen von Zufallsvariablen betrachtet werden können, die identische Parameter aufweisen - im Falle parametrischer Tests. Sonst wird überprüft, ob die Verteilungen der beiden Datensätze vereinbar sind. Unverbundene Zweistichprobentests sind wie die verbundenen Zweistichprobentests formal gesehen bivariate Tests. Bei den unverbundenen Tests ist eine der Variablen eine Gruppenvariable: sie gibt an, zu welcher Stichprobe ein Datum gehört. Sind die Daten nicht in einer Urliste gegeben (pro Zeile Daten eines Objektes), so muss man die Tabelle entsprechend umstellen.

22.1 Testen von Varianzen

Ziel: Die Varianzen zweier Stichproben werden verglichen, um zu entscheiden, ob sie bezüglich Varianzen verträglich sind.

Voraussetzungen: Die Daten sind unabhängig und identisch normalverteilt (metrisch skaliert).

Sind die Daten normalverteilt, dann gilt - wie bereits gezeigt - für die Zufallsvariable

$$\frac{(n_1 - 1)S_1^2}{s_1^2} \text{ ist } \chi_{n_1-1}^2 - \text{verteilt.}$$

Ebenso gilt für

$$\frac{(n_2 - 1)S_2^2}{s_2^2} \text{ ist } \chi_{n_2-1}^2 - \text{verteilt.}$$

Laut Definition der $F_{m,n}$ -Verteilung gilt dann

$$\frac{\frac{1}{n_1-1} \frac{(n_1-1)S_1^2}{s_1^2}}{\frac{1}{n_2-1} \frac{(n_2-1)S_2^2}{s_2^2}} = \frac{\frac{S_1^2}{s_1^2}}{\frac{S_2^2}{s_2^2}} \text{ ist } F_{(n_1-1, n_2-1)} - \text{verteilt.}$$

Unter der Nullhypothese gilt $s_1^2 = s_2^2$. Somit ist die Teststatistik $\frac{S_1^2}{S_2^2}$ unter der Nullhypothese $F_{(n_1-1, n_2-1)}$ -verteilt:

$\frac{S_1^2}{S_2^2}$ ist $F_{(n_1-1, n_2-1)}$ -verteilt.
S_1^2 : Zufallsvariable mit s_1^2 als Wert
S_2^2 : Zufallsvariable mit s_2^2 als Wert
n_1 : Umfang der ersten Stichprobe
n_2 : Umfang der zweiten Stichprobe

Testen wir $s_1^2 < s_2^2$ und berechnen wir den Testwert von $\frac{S_1^2}{S_2^2}$, so liegt ein linksseitiger Test vor. Berechnen wir jedoch den Testwert von $\frac{S_2^2}{S_1^2}$, liegt ein rechtsseitiger Test vor. Testen wir $s_1^2 > s_2^2$ und berechnen wir den Testwert von $\frac{S_1^2}{S_2^2}$, liegt ein rechtsseitiger Test vor. Berechnen wir jedoch den Testwert von $\frac{S_2^2}{S_1^2}$, liegt ein linksseitiger Test vor.

Beispiel 22.1.1. Zwei Maschinen werden auf die Varianz des Outputs hin untersucht, wobei eine tiefere Varianz vorzuziehen ist. Wir nehmen an, die Daten seien normalverteilt (Wir überprüfen jede der Stichproben auf Normalverteilung)

Wir ziehen für beide Maschinen eine Stichprobe von je 100, d.h. $n_1 = n_2$ und hätten als Varianzen die folgenden Werte erhalten: $s_1^2 = 90$; $s_2^2 = 80$. Wir überprüfen, ob der Unterschied signifikant ist.

H_0 : Die Varianzen unterscheiden sich nicht.

H_A : Die Varianzen unterscheiden sich ($\implies$ zweiseitiger Test) ($\alpha = 0.05$).

Die Teststatistik $\frac{S_1^2}{S_2^2}$ nimmt den Testwert $\frac{90}{80} = 1.125$ an. $F_{F_{99,99}}(1.125) = f.vert(1.125; 99; 99; 1) = 0.72049648$; $1 - 0.72049648 = 0.279503524$

$2 \min\{0.72049648, 0.279503524\} = 2 \cdot 0.279503524 = 0.55901 > 0.05$.

Die Varianzen des Outputs der beiden Maschinen unterscheiden sich nicht signifikant. $\diamond$

22.2 t-Test bei nicht signifikant verschiedenen Varianzen

Ziel: Es wird überprüft, ob sich zwei Stichproben bezüglich der Mittelwerte unterscheiden.

Voraussetzungen: Die Daten sind unabhängig und identisch normalverteilt (metrisch skaliert). Die Varianzen werden aus den Daten geschätzt und unterscheiden sich nicht signifikant.

Das Vorliegen verschiedener Varianzen wird auch „heteroskedastisch“, des identischer „homoskedastisch“ genannt. Ohne Beweis halten wir fest, dass die folgende Teststatistik T t -verteilt ist.

$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1) + (n_2-1)} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \text{ ist } t_{n_1+n_2-2}\text{-verteilt.}$
$\bar{X}_1$: Zufallsvariable mit dem Mittelwert $\bar{x}_1$ der Stichprobe 1 als Wert $\bar{X}_2$: Zufallsvariable mit dem Mittelwert $\bar{x}_2$ der Stichprobe 2 als Wert
S_1^2 : Zufallsvariable mit der Varianz s_1^2 der Stichprobe 1 als Wert.
S_2^2 : Zufallsvariable mit der Varianz s_2^2 der Stichprobe 2 als Wert
n_1 : Anzahl Daten der Stichprobe 1
n_2 : Anzahl Daten der Stichprobe 2

Beispiel 22.2.1. Zwei Produktionsstätten produzieren folgende Mengeneinheiten an Produkten pro Tag (Daten: UnverbundeneZweiStichprobentests.xls). Ist der Unterschied signifikant?

($n_1 = 52$; $s_1^2 = 7.375786375$, $\bar{x}_1 = 1000.22908$;

$n_2 = 75$, $s_2^2 = 7.55534469$, $\bar{x}_2 = 1005.509322$)

H_0 : Die zweite Produktionsstätte produziert nicht mehr.

H_A : Die zweite Produktionsstätte produziert mehr ($\implies$ einseitiger Test; $\alpha = 0.05$).

Wir denken an einen t -Test. Entsprechend überprüfen wir die Daten zuerst auf Normalverteilung (s. Abbildungen 22.2.1 und 22.2.2).

Die Normalverteilungsannahme ist wohl haltbar.

Als nächstes testen wir die Varianzen. $\frac{7.375786375}{7.55534469} = 0.97623$;

$F_{F_{(51,74)}}(0.97623) = f.vert(0.97623; 51; 74; 1) = 0.469232962$

(Wir treffen diese Entscheidung mit dem obigen F -Test. Oft werden auch andere, geeignete Tests

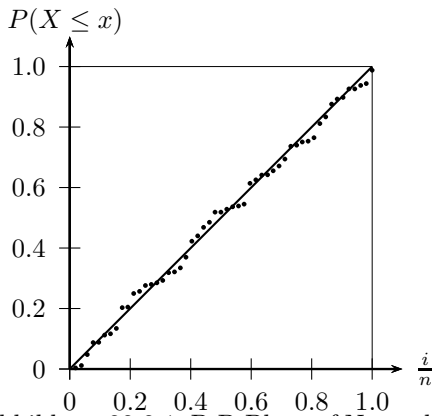


Abbildung 22.2.1: P-P-Plot auf Normalverteilung Stichprobe 1

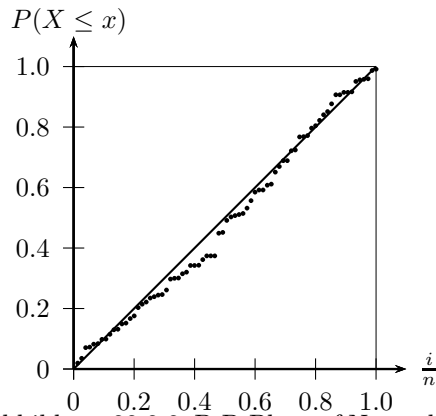


Abbildung 22.2.2: P-P-Plot auf Normalverteilung Stichprobe 2

verwendet).

Die Varianzen unterscheiden sich nicht signifikant. Entsprechend verwenden wir den t-Test:

Die Teststatistik

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1) + (n_2-1)} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

nimmt den Wert

$$\frac{1000.22908 - 1005.509322}{\sqrt{\frac{51 \cdot 7.375786375 + 74 \cdot 7.55534469}{51+74} \left(\frac{1}{52} + \frac{1}{75} \right)}} = -10.697$$

an.

$$F_{t_{52+75-2}}(-10.697) = t.\text{vert}(-10.697; 52 + 75 - 2; 1) = 1 - 1.16665E - 19$$

p-Wert: $1.16665E - 19 \leq 0.05$.

Das Testergebnis ist hochsignifikant.

Es liegt ein linksseitiger Test vor. Würde man die Subtraktion im Zähler umkehren, so würden wir einen rechtsseitigen Test erhalten. $\diamond$

22.3 t-Test bei signifikant verschiedenen Varianzen (Welch-Test)

Unterscheiden sich die Varianzen signifikant, so können wir für die obige Situation nicht den eingeführten t-Test verwenden. Sind die Daten normalverteilt oder liegen mehr als 50 Daten für jede Stichprobe vor, so bietet sich der sogenannte Welch-t-Test an.

Ziel: Es wird überprüft, ob die zwei Stichproben verträgliche Mittelwerte aufweisen.

Voraussetzungen: Die Daten sind unabhängig und identisch normalverteilt (oder es liegen für beide Stichproben mehr als 50 Daten vor; metrisch skaliert).

Teststatistik:

$T = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ ist näherungsweise t -verteilt mit
$df = \frac{(1+R)^2}{\frac{R^2}{(n_1-1)} + \frac{1}{n_2-1}}$, wobei $R = \frac{\frac{S_1^2}{n_1}}{\frac{S_2^2}{n_2}} = \frac{S_1^2 n_2}{S_2^2 n_1}$
$\bar{X}_1$ = ZV Mittelwert der Variable 1
$\bar{X}_2$ = ZV Mittelwert der Variable 2
S_1^2 = Zufallsvariable Varianz der Variable 1
S_2^2 = Zufallsvariable Varianz der Variable 2
n_1 = Stichprobengröße der Stichprobe 1
n_2 = Stichprobengröße der Stichprobe 2

Beispiel 22.3.1. (Daten s. *UnverbundeneZweiStichprobentests.xls*) Verkäufer 1 erzielt einen durchschnittlichen Tagesumsatz von 100.7251579 (Anzahl Tage: 95, $s_1 = 4.142825345$) und Verkäufer 2 (Anzahl Tage 112, $s_2 = 2.88585642$) einen durchschnittlichen Tagesumsatz von 104.7855357 Geldeinheiten

H_0 : Die durchschnittlichen Umsätze der beiden Verkäufer unterscheiden sich nicht.

H_A : Die durchschnittlichen Umsätze der beiden Verkäufer unterscheiden sich ($\Rightarrow$ zweiseitiger Test, $\alpha = 0.05$).

Wir überprüfen die Normalverteilungsannahme (s. Abbildungen 22.3.3 und 22.3.4):

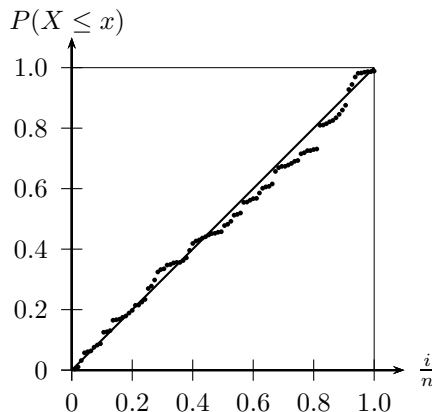


Abbildung 22.3.3: P-P-Plot auf Normalverteilung Stichprobe 1

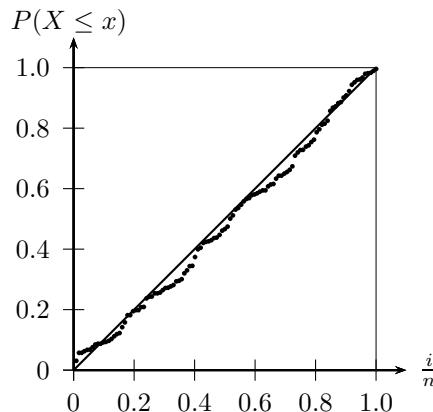


Abbildung 22.3.4: P-P-Plot auf Normalverteilung Stichprobe 2

Die Normalverteilungsannahme kann akzeptiert werden.

$n_1 = 95; n_2 = 112 > 50$;

Die Teststatistik $F = \frac{S_2^2}{S_1^2}$ nimmt den Wert $\frac{2.88585642^2}{4.142825345^2} = 0.48524$ und ist $F_{111,94}$ -verteilt.

$2 \cdot F_{F_{111,94}}(0.48524) = 2 \cdot 0.00013552 \leq 0.05$.

Wir verwerfen die Hypothese, dass die beiden Populationen die selbe Varianz haben und verwenden den Welch-Test:

$$T = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{100.7251579 - 104.7855357}{\sqrt{\frac{4.142825345^2}{95} + \frac{2.88585642^2}{112}}} = -8.0404$$

$$R = \frac{\frac{S_1^2}{n_1}}{\frac{S_2^2}{n_2}} = \frac{\frac{4.142825345^2}{95}}{\frac{2.88585642^2}{112}} = 2.4296$$

$$df = \frac{(1 + 2.4296)^2}{\frac{2.4296^2}{(95-1)} + \frac{1}{112-1}} = 163.8$$

p -Wert: $8.65974E-14$ (mit Excel: $=1-t.vert(8.0404;163.8;1)$)

◇

Bemerkung 22.3.2. Liegen mehr als 50 Daten für beide Stichproben vor, so könnte man auch statt der t -Verteilung die Normalverteilung verwenden: die Fehlerwahrscheinlichkeit 1. Ordnung wird dadurch etwas grösser, was allerdings in der Praxis kaum ins Gewicht fällt. ◇

22.4 Rangtest von Mann-Whitney

Ziel: Es wird überprüft, ob sich die Mediane zweier unverbundener Stichproben unterscheiden.

Voraussetzungen: Die Daten sind unabhängig und identisch verteilt (metrisch oder ordinal skaliert).

Der Mann-Whitney-Test wird auch „U-Test von Mann-Whitney“ genannt. Der Rangtest hat bedeutend schwächere Voraussetzungen als der t -Test bei nicht signifikant verschiedener Varianz oder der Welch- t -Test. Wir ordnen die Daten der beiden Stichproben untereinander an, indem wir eine Variable „Stichprobe“ schaffen. Wir ordnen danach die Daten der Grösse nach und ordnen ihnen nach derselben Methode wie beim Wilcoxon-Test Ränge zu. Von verbundenen Rängen sprechen wir, wenn Daten die selbe Ausprägung haben. Im Gegensatz zum Wilcoxon-Test werden im Test selber nicht die Rangsummen verwendet. Es wird gezählt, wie viele Male u_1 die Ränge der Stichprobe S_1 Ränge der Stichprobe S_2 übertreffen oder wie viele Male u_2 die Ränge der Stichprobe S_2 die Ränge der Stichprobe S_1 übertreffen. Formal

$u_1 := \{(x_k, x_j) \mid r(x_k) \geq r(x_j) \text{ und } x_k \in S_1, x_j \in S_2\} $
$u_2 := \{(x_k, x_j) \mid r(x_k) \geq r(x_j) \text{ und } x_k \in S_2, x_j \in S_1\} $
$S_i :=$ Menge der Objekte der Stichprobe i
$r(x_j) :=$ Rang des Objektes x_j .

U_i sind die Teststatistiken, welche die $u_i, i \in \{1, 2\}$, als Werte annehmen. Diese Teststatistiken sind identisch mit

$U_1 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$
$U_2 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$
R_i Zufallsvariable, die als Wert $\sum_{x_j \in S_i} r(x_j)$ annimmt

Die Zufallsvariable R_i nimmt also als Wert die Rangsumme der Daten x_j der i -ten Stichprobe an. Für einen Beweis siehe Satz 22.4.3.

Beispiel 22.4.1. Wir betrachten die zwei Ministichproben

Ränge	Stichprobe
1	2
2	1
3	2
4	2
5	1

Offenbar gilt: $n_1 = 2$; $n_2 = 3$; Wert von $R_1 = 7$; Wert von $R_2 = 8$; Es gibt zwei Datenpaare, so dass die Ränge der Stichprobe 2 grösser als die Ränge der Stichprobe 1 sind $((3, 2)$ und $(4, 2)$).

Es ist also $u_2 = 2$. Es gibt 4 Datenpaare, so dass Ränge der Stichprobe 1 grösser als Ränge der Stichprobe 2 sind $((2,1), (5,1), (5,3), (5,4))$. Damit ist $u_1 = 4$.

Mit den obigen Formeln: $u_1 = 2 \cdot 3 + \frac{3 \cdot 4}{2} - 8 = 4$; $u_2 = 2 \cdot 3 + \frac{2 \cdot 3}{2} - 7 = 2$ $\diamond$

Es gilt

Satz 22.4.2.

$$u_2 = n_1 n_2 - u_1.$$

Beweis. Gilt für alle $r(x_i)$ mit $x_i \in S_1$ und $r(x_j)$ mit $x_j \in S_2$, dass $r(x_i) > r(x_j)$, beträgt die Anzahl der Paare mit $r(x_i) > r(x_j)$ offensichtlich $n_1 n_2$ (für jedes $x_i \in S_1$ gibt es n_2 $x_j \in S_2$, so dass $r(x_i) > r(x_j)$). Die Gesamtanzahl Paare bleibt unverändert, auch wenn die Ränge der Stichproben nicht sauber in Gruppen getrennt sind. Die Summe der Anzahl Paare u_1 und u_2 ist damit mit $n_1 n_2$ identisch. $\square$

Satz 22.4.3. $U_1 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$; $U_2 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$

Beweis. Aus Satz 22.4.2 folgt:

$$U_2 + U_1 = n_1 n_2$$

Es muss also

$$n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 + n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = n_1 n_2 \quad (22.4.1)$$

gelten. In der Tat folgt aus (22.4.1)

$$\begin{aligned} \frac{n_2(n_2+1)}{2} - R_2 + n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 &= 0 \\ \frac{n_2(n_2+1)}{2} + n_1 n_2 + \frac{n_1(n_1+1)}{2} &= R_1 + R_2 = \frac{(n_1+n_2)(n_1+n_2+1)}{2} \\ \frac{n_2(n_2+1) + 2n_1 n_2 + n_1(n_1+1)}{2} &= \frac{(n_1+n_2)(n_1+n_2+1)}{2} \\ n_2(n_2+1) + 2n_1 n_2 + n_1(n_1+1) &= (n_1+n_2)(n_1+n_2+1) \\ n_1^2 + 2n_1 n_2 + n_1 + n_2^2 + n_2 &= n_1^2 + 2n_1 n_2 + n_1 + n_2^2 + n_2 \end{aligned}$$

$\square$

Im Test verwenden wir bei der Verwendung von Tabellen $u = \min\{u_1, u_2\}$, d.h. den kleineren der Testwerte u_i - wobei zu überprüfen ist, ob die Richtung der Abweichung der Rangsumme vom Mittel mit der Alternativ-Hypothese übereinstimmt. Die Rangsummen der Stichproben berechnen wir in Excel, indem wir mit „wenn-dann“-Befehlen für jede Stichprobe die Ränge herausfiltern und diese dann zusammenzählen.

In Tabellen werden die kritischen u -Werte angegeben (zur Erinnerung: die kritischen Werte sind die Grenzen des Verwerfungsbereichs. Die Grenzen gehören zum Verwerfungsbereich; s. Tabellen 22.4.1, 22.4.2 und 22.4.13.)

n_2	9	10	11	12	13	14	15	16	17	18	19	20
n_1												
2					0	0	0	0	0	0	1	1
3	1	1	1	2	2	2	3	3	4	4	4	5
4	3	3	4	5	5	6	7	7	8	9	9	10
5	5	6	7	8	9	10	11	12	13	14	15	16
6	7	8	9	11	12	13	15	16	18	19	20	22
7	9	11	12	14	16	17	19	21	23	24	26	28
8	11	13	15	17	20	22	24	26	28	30	32	34
9	14	16	18	21	23	26	28	31	33	36	38	40
10	16	19	22	24	27	30	33	36	38	41	44	47
11	18	22	25	28	31	34	37	41	44	47	50	53
12	21	24	28	31	35	38	42	46	49	53	56	60
13	23	27	31	35	39	43	47	51	55	59	63	67
14	26	30	34	38	43	47	51	56	60	65	69	73
15	28	33	37	42	47	51	56	61	66	70	75	80
16	31	36	41	46	51	56	61	66	71	76	82	87
17	33	38	44	49	55	60	66	71	77	82	88	93
18	36	41	47	53	59	65	70	76	82	88	94	100
19	38	44	50	56	63	69	75	82	88	94	101	107
20	40	47	53	60	67	73	80	87	93	100	107	114

Tabelle 22.4.1: Kritische Werte für den U-Test von Mann und Whitney für den einseitigen Test bei $\alpha = 0.01$, für den zweiseitigen Test bei $\alpha = 0.02$; der kritische Wert gehört zum Verwerfungsbereich

n_2	9	10	11	12	13	14	15	16	17	18	19	20
n_1												
2	0	0	0	1	1	1	1	1	2	2	2	2
3	2	3	3	4	4	5	5	6	6	7	7	8
4	4	5	6	7	8	9	10	11	11	12	13	13
5	7	8	9	11	12	13	14	15	17	18	19	20
6	10	11	13	14	16	17	19	21	22	24	25	27
7	12	14	16	18	20	22	24	26	28	30	32	34
8	15	17	19	22	24	26	29	31	34	36	38	41
9	17	20	23	26	28	31	34	37	39	42	45	48
10	20	23	26	29	33	36	39	42	45	48	52	55
11	23	26	30	33	37	40	44	47	51	55	58	62
12	26	29	33	37	41	45	49	53	57	61	65	69
13	28	33	37	41	45	50	54	59	63	67	72	76
14	31	36	40	45	50	55	59	64	67	74	78	83
15	34	39	44	49	54	59	64	70	75	80	85	90
16	37	42	47	53	59	64	70	75	81	86	92	98
17	39	45	51	57	63	67	75	81	87	93	99	105
18	42	48	55	61	67	74	80	86	93	99	106	112
19	45	52	58	65	72	78	85	92	99	106	113	119
20	48	55	62	69	76	83	90	98	105	112	119	127

Tabelle 22.4.2: Kritische Werte für den U-Test von Mann und Whitney für den einseitigen Test bei $\alpha = 0.025$, für den zweiseitigen Test bei $\alpha = 0.05$; der kritische Wert gehört zum Verwerfungsbereich

n_2	9	10	11	12	13	14	15	16	17	18	19	20
n_1												
1											0	0
2	1	1	1	2	2	2	3	3	3	4	4	4
3	3	4	5	5	6	7	7	8	9	9	10	11
4	6	7	8	9	10	11	12	14	15	16	17	18
5	9	11	12	13	15	16	18	19	20	22	23	25
6	12	14	16	17	19	21	23	25	26	28	30	32
7	15	17	19	21	24	26	28	30	33	35	37	39
8	18	20	23	26	28	31	33	36	39	41	44	47
9	21	24	27	30	33	36	39	42	45	48	51	54
10	24	27	31	34	37	41	44	48	51	55	58	62
11	27	31	34	38	42	46	50	54	57	61	65	69
12	30	34	38	42	47	51	55	60	64	68	72	77
13	33	37	42	47	51	56	61	65	70	75	80	84
14	36	41	46	51	56	61	66	71	77	82	87	92
15	39	44	50	55	61	66	72	77	83	88	94	100
16	42	48	54	60	65	71	77	83	89	95	101	107
17	45	51	57	64	70	77	83	89	96	102	109	115
18	48	55	61	68	75	82	88	95	102	109	116	123
19	51	58	65	72	80	87	94	101	109	116	123	130
20	54	62	69	77	84	92	100	107	115	123	130	138

Tabelle 22.4.3: Kritische Werte für den U-Test von Mann und Whitney für den einseitigen Test bei $\alpha = 0.05$, für den zweiseitigen Test bei $\alpha = 0.1$; der kritische Wert gehört zum Verwerfungsbereich

Bei $n_1 + n_2 \geq 30$ ist die Verteilung der Teststatistik U_i näherungsweise normal.

$Z = \frac{U_i - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0, 1)$ (asymptotisch)
U_1 : Zufallsvariable $= n_1 n_2 + \frac{n_2 (n_2 + 1)}{2} - R_2$
U_2 : Zufallsvariable $= n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1$
R_i : Zufallsvariable (Wert von $R_i = \sum_{j=1}^{n_i} r_i(x_j)$)
$r_i(x_j)$ = Rang des Datums x_j der Stichprobe i
n_1 : Umfang der ersten Stichprobe
n_2 : Umfang der zweiten Stichprobe

Bei verbundenen Rängen wird eine andere Standardabweichungsformel verwendet:

$Z = \frac{U_i - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2}{n(n-1)} \left(\frac{n^3 - n}{12} - \sum_{i=1}^k \frac{t_i^3 - t_i}{12} \right)}} \sim N(0, 1)$
t_i =: Anzahl der Daten mit dem jeweils selben Rang i
k : Anzahl der Datengruppen mit verbundenen Rängen
$n := n_1 + n_2$
Übrige Interpretationen wie oben.

Wir machen die Verwendung der Formel

$$\frac{n^3 - n}{12} - \sum_{i=1}^k \frac{t_i^3 - t_i}{12}$$

mit einem Beispiel deutlich (das Beispiel hätte für den Test etwas viele verbundene Ränge. Es geht hier allerdings darum, die obige Formel anwenden zu können; s. Tabelle 22.4.4).

Gruppe	Daten	Ränge	no
1	1	1.5	1
2	1	1.5	2
1	2	3.5	3
2	2	3.5	4
1	3	6	5
1	3	6	6
2	3	6	7
1	4	9.5	8
1	4	9.5	9
2	4	9.5	10
2	4	9.5	11
1	5	12.5	12
2	5	12.5	13
2	6	14	14
2	7	15	15

Tabelle 22.4.4: Beispiel für die Berechnung der Varianz bei verbundenen Rängen beim U-Test

Wir erhalten

$$\sum_{i=1}^k \frac{t_i^3 - t_i}{12} = \frac{2^3 - 2}{12} + \frac{2^3 - 2}{12} + \frac{3^3 - 3}{12} + \frac{4^3 - 4}{12} + \frac{2^3 - 2}{12} = 8.1667$$

Bemerkung 22.4.4. *Es empfiehlt sich, beim Mann-Whitney-Test jeweils die Rangmittel zu berechnen, um den allfälligen Unterschied zwischen den Stichproben auch inhaltlich interpretieren zu können.* $\diamond$

Bemerkung 22.4.5. *Es gibt einen mit dem Mann-Whitney-Test äquivalenten Wilcoxon-Test, der statt mit den Statistiken U_i unmittelbar mit den Stichprobenrängen R_i arbeitet.* $\diamond$

Beispiel 22.4.6. *(für metrisch skalierte Daten)*

Für ein Produkt wird ein Zusatz getestet, der sich positiv auf die Lebensdauer des Produktes (in Monaten) auswirken soll.

H_0 : Der Zusatz hat keine Wirkung

H_A : Der Zusatz verlängert die Lebensdauer des Produktes ($\implies$ rechtsseitiger Test, $\alpha = 0.01$)

Es wurden dabei folgende Daten erhoben (s. Tabelle 22.4.5):

Lebensdauer ohne Zusatz	Lebensdauer mit Zusatz
85	96
106	105
118	104
81	108
138	86
90	84
112	99
119	101
107	78
95	124
88	121
103	97
	129
	87
	109

Tabelle 22.4.5: Beispieldaten für unverbundenen Zweistichprobentest

Im Beispiel gibt es keine Bindungen. Entsprechend gilt dann: $no(i) = r(i)$ für alle i . Wir erhalten (s. Tabelle 22.4.6):

Lebensdauer geordnet	Zusatz (ohne = 1; mit = 2)	Nummerierung (= Ränge)	Ränge Stichprobe 1	Ränge Stichprobe 2
78	2	1	0	1
81	1	2	2	0
84	2	3	0	3
85	1	4	4	0
86	2	5	0	5
87	2	6	0	6
88	1	7	7	0
90	1	8	8	0
95	1	9	9	0
96	2	10	0	10
97	2	11	0	11
99	2	12	0	12
101	2	13	0	13
103	1	14	14	0
104	2	15	0	15
105	2	16	0	16
106	1	17	17	0
107	1	18	18	0
108	2	19	0	19
109	2	20	0	20
112	1	21	21	0
118	1	22	22	0
119	1	23	23	0
121	2	24	0	24
124	2	25	0	25
129	2	26	0	26
138	1	27	27	0
Summen		378	172	206

Tabelle 22.4.6: Berechnung der Rangsummen für den U-Test - für die Daten der Tabelle 22.4.5

Um zu wissen, welche Daten zu welcher Stichprobe gehören, kennzeichnen wir diese Zugehörigkeit vor dem Ordnen durch 1 und 2. In der zweitletzten Spalte sind die Ränge für die Daten der ersten Stichprobe angeführt, in der letzten Spalte die Ränge der Daten der zweiten Stichprobe. Die Summen der beiden Rangsummen $R_1 + R_2$ muss mit der Summe der $r(i)$ identisch sein. Das Rangmittel der ersten Stichprobe beträgt $\frac{172}{12} = 14.333$, das der zweiten Stichprobe $\frac{206}{15} = 13.733$. Die Rangmittel unterscheiden sich nicht stark.

Wir berechnen u_1 : Die Teststatistik

$$U_2 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

nimmt den Wert

$$u_2 = \left((12 \cdot 15) + \frac{12(12 + 1)}{2} \right) - 172 = 86$$

an.

Die Teststatistik $U_1 = n_1 n_2 - U_2$ nimmt damit den Wert $(12 \cdot 15) - 86 = 94 = u_1$ an.

Da $n_1 + n_2 < 30$, müssen wir den Wert in einer Tabelle nachschauen. Wir erhalten als kritische Grenze 42. Da $86 > 42$, ist das Resultat nicht signifikant.

Übungshalber verwenden wir die Normalverteilung, um den grob genäherten p -Wert zu berechnen:

$$s = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{12 \cdot 15 (12 + 15 + 1)}{12}} = 20.494.$$

$$\frac{n_1 n_2}{2} = \frac{12 \cdot 15}{2} = 90$$

Wir erhalten: $Z = \frac{86-90}{20.495} = -0.19517$

$\Phi(-0.19517) = 1 - \Phi(0.19517) = 1 - 0.577370049 = 0.42263 > 0.01.$

◇

Beispiel 22.4.7. (für ordinal skalierte Daten): (Daten s. *UnverbundeneZweiStichprobentests.xls*). Bei einem betriebsinternen Rating werden die Gesamtleistungen (Zusammenarbeit, Mitdenken, Leistung, etc.) von Mitarbeitern der zwei Standorte von einem Vorgesetzten auf einer Ordinalskala von 1 - 10 bewertet. Es ergibt sich die folgende Tabelle 22.4.7:

Standort 1	Standort 2
5	5.5
6	6
4	4.5
6.5	6.75
7	7.5
8	8.25
9.5	9
10	6.25
8.5	8.75
	9.25
	9.75
	4.75

Tabelle 22.4.7: Beispieldaten für unverbundenen Zweistichprobentest

H_0 : Die beiden Produktionsstätten unterscheiden sich nicht.

H_A : Die beiden Produktionsstätten unterscheiden sich ($\implies$ zweiseitiger Test, $\alpha = 0.05$).

Die Mediane unterscheiden sich wie folgt: erste Stichprobe hat einen Median von 7, die zweite einen Median von 7.125.

Wir ordnen die Daten und rangieren sie (s. Tabelle 22.4.8):

Leistung geordnet	Standort (1 = Standort 1)	Nummerierung	Ränge	Ränge Standort 1	Ränge Standort 2
4	1	1	1	1	0
4.5	2	2	2	0	2
4.75	2	3	3	0	3
5	1	4	4	4	0
5.5	2	5	5	0	5
6	1	6	6.5	6.5	0
6	2	7	6.5	0	6.5
6.25	2	8	8	0	8
6.5	1	9	9	9	0
6.75	2	10	10	0	10
7	1	11	11	11	0
7.5	2	12	12	0	12
8	1	13	13	13	0
8.25	2	14	14	0	14
8.5	1	15	15	15	0
8.75	2	16	16	0	16
9	2	17	17	0	17
9.25	2	18	18	0	18
9.5	1	19	19	19	0
9.75	2	20	20	0	20
10	1	21	21	21	0
Summen		231	231	99.5	131.5

Tabelle 22.4.8: Berechnung der Rangsummen der Daten der Tabelle 22.4.7 für den U-Test

Rangmittel der ersten Stichprobe: $\frac{99.5}{9} = 11.056$, Rangmittel der zweiten Stichprobe $\frac{131.5}{12} = 10.958$. Die Rangmittel liegen nicht weit auseinander.

Wir berechnen den Wert von

$$U_2 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1 : \left(9 \cdot 12 + \frac{9(9 + 1)}{2} \right) - 99.5 = 53.5$$

Wert von $U_1 : 9 \cdot 12 - 53.5 = 54.5$. (kritischer Wert aus der Tabelle: 26). Der Unterschied ist nicht signifikant, was schon der Umstand zeigt, dass U_1 und U_2 sehr nahe beieinander liegen. $\diamond$

22.5 Test auf Anteile bei zwei Ausprägungen

Ziel: Es sollen überprüft werden, ob die Anteile in zwei Stichproben zueinander passen.

Voraussetzungen: Stichprobengröße: $n_i \geq 50$ (manche Autoren verlangen auch nur $m_{ij} \geq 5$, wobei m_{ij} die erwarteten Häufigkeiten der 2×2 -Tabelle sind. Diese werden berechnet mit $n_{i+} \frac{n_{+j}}{n}$, wobei n_{i+} die Zeilensumme der i -ten Zeile ist, n_{+j} die Spaltensumme der j -ten Spalte ist und $n = n_1 + n_2$. Eine Begründung für diese Formel $n_{i+} \frac{n_{+j}}{n}$ erfolgt im Zusammenhang mit dem bivariaten Chi-Quadrattest. Die Daten sind Werte von unabhängigen, identisch verteilten Zufallsvariablen.

Ohne Beweis halten wir fest:

$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\frac{X_1+X_2}{n_1+n_2} \left(1 - \frac{X_1+X_2}{n_1+n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1) \quad (\text{asymptotisch für } n_i \geq 50)$	
$\frac{X_1}{n_1}$	: Zufallsvariable mit dem Anteil $\frac{x_1}{n_1}$ an der Stichprobe 1 als Wert
$\frac{X_2}{n_2}$	: Zufallsvariable mit dem Anteil $\frac{x_2}{n_2}$ an der Stichprobe 2 als Wert
n_1	: Grösse der Stichprobe 1
n_2	: Grösse der Stichprobe 2

Beispiel 22.5.1. Es wird untersucht, ob sich der Jugendanteil (< 30 Jahre) bei den Konsumenten eines Produktes in zwei Kantonen unterscheiden (Zürich; Waadt). In beiden Kantonen wird eine Zufallsstichprobe gezogen (je 50). Wir nehmen an, es hätten sich folgende Anteile ergeben: 0.3 und 0.46. Ist zu vermuten, dass die Konsumentenstruktur der beiden Kantone sich unterscheidet?

H_0 : Die Jugendanteile unterscheiden sich nicht.

H_A : Die Jugendanteile unterscheiden sich ($\implies$ zweiseitiger Test, $\alpha = 0.05$).

$$\frac{X_1}{n_1} : 0.3 \implies X_1 = 50 \cdot 0.3 = 15$$

$$\frac{X_2}{n_2} : 0.46 \implies X_2 = 50 \cdot 0.46 = 23$$

$$\frac{X_1+X_2}{n_1+n_2} : \frac{15+23}{50+50} = 0.38$$

Testwert:

$$\frac{0.3 - 0.46}{\sqrt{0.38(1 - 0.38)\left(\frac{1}{50} + \frac{1}{50}\right)}} = -1.6482$$

$$\Phi(-1.6482) = 0.049655802$$

Da wir zweiseitig testen gilt: $0.049655802 > 0.025$.

Der Unterschied ist somit nicht signifikant. ◇

22.5.1 Übungen

- Bei der Produktion mit zwei verschiedenen Maschinen wurden folgende Daten erhoben (UnverbundeneZweiStichprobentests.xls). Überprüfen Sie, ob die Varianzen signifikant voneinander abweichen ($\alpha = 0.01$)
- Daten: unter UnverbundeneZweiStichprobentests.xls
Zwei Anlagen produzieren die pro Tag angegebenen Mengen eines chemischen Feinproduktes (in Liter). Unterscheiden sich die beiden Anlagen? ($\alpha = 0.05$)
- Zwei Versicherungs-Vertreter werden bezüglich ihrer Abschlüsse verglichen (Daten: UnverbundeneZweiStichprobentests.xls). Überprüfen Sie, ob die beiden Vertreter in ihren Leistungen signifikant voneinander abweichen. ($\alpha = 0.05$)
- Zwei Fonds-Manager werden bezüglich Ihrer Performance getestet (Gewinne pro Tag). (Daten: UnverbundeneZweiStichprobentests.xls).
- Zwei Diäten werden bei Dauerläufern bezüglich der Leistung geprüft. (Daten: UnverbundeneZweiStichprobentests.xls). Unterscheiden sich die Leistungen?
- Im April 02 ergeben sich in der 2. HSW die folgenden Resultate in einer Mathematikprüfung (s. Tabelle 22.5.9):

Geschlecht	Punkte
w	19.25
w	19
m	18
m	18
w	17.5
w	17
m	16.5
w	16.25
m	16
m	15.75
m	13.25
m	12.75
m	11.875
m	11
m	7.5
m	1.5

Tabelle 22.5.9: Beispieldaten für unverbundenen Zweistichprobentest

(w für „weiblich“, „m“ für „männlich“). Die Hypothese sei, dass die Frauen signifikant besser abgeschnitten haben. Überprüfen Sie diese Hypothese. $\alpha = 0.01$

- Berechnen Sie das Beispiel 22.5.1 mit Hilfe der Urdaten (s. Exceltabelle UnverbundeneZwei-Stichprobentests.xls). Sie können diese Urdaten auch in R direkt erzeugen durch:
`m=matrix(c(rep(1,50),rep(2,50), rep(1,15),rep(2,35),rep(1,23),rep(2,27)), ncol=2)`
`alter=data.frame(m);names(alter)=c("Alter","Kaeufer")`
- Die politische Einstellung der Bevölkerungen Genfs und Zürichs bezüglich eines spezifischen Themas sollen verglichen werden. Dabei werden 60 Personen in Genf und 70 Personen in Zürich befragt. Es ergibt sich folgendes Bild: 15 Personen in Genf sind positiv eingestellt und 10 Personen in Zürich. Handelt es sich um einen statistisch gesehen relevanten Unterschied?
 $\alpha = 0.05$
- Finden Sie je ein betriebswirtschaftliches Problem, das mit den behandelten Methoden gelöst werden kann.

22.5.2 Lösungen

- 1) H_0 : Die Varianzen unterscheiden sich nicht.
 H_A : Die Varianzen unterscheiden sich ($\implies$ zweiseitiger Test, $\alpha = 0.01$)
Wir überprüfen die Daten auf Normalverteilung (s. Abbildungen 22.5.5 und 22.5.6):

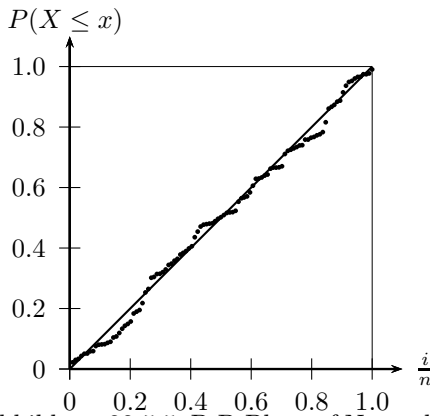


Abbildung 22.5.5: P-P-Plot auf Normalverteilung Stichprobe 1

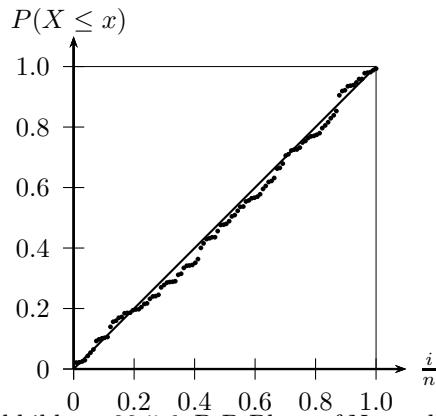


Abbildung 22.5.6: P-P-Plot auf Normalverteilung Stichprobe 2

Die Daten können als normalverteilt betrachtet werden. Wir können somit den eingeführten F-Test verwenden. Die Teststatistik $\frac{S_1^2}{S_2^2}$ nimmt den Wert $\frac{6.017110135}{13.09697008} = 0.45943$ an.
 $F_{103,106}(0.45943) = 0.000046284 < 0.005$
 Die Abweichung ist hochsignifikant.

2. H_0 : Die Anlagen unterscheiden sich nicht.

H_A : Die Anlagen unterscheiden sich ($\implies$ zweiseitiger Test).

Wir denken an einen t - oder Welch- t -Test. Wir überprüfen die Daten auf Normalverteilung (s. Abbildungen 22.5.7 und 22.5.8):

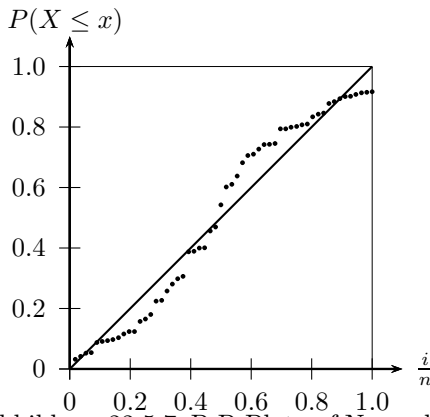


Abbildung 22.5.7: P-P-Plot auf Normalverteilung Stichprobe 1

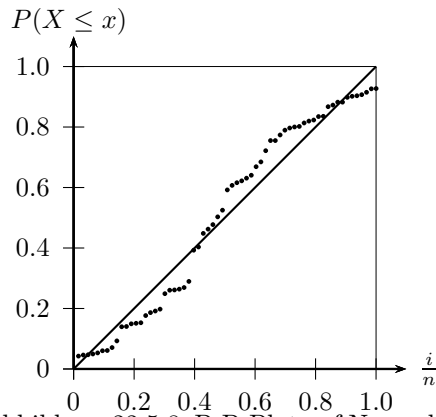


Abbildung 22.5.8: P-P-Plot auf Normalverteilung Stichprobe 2

Die Daten sind nicht normalverteilt. Entsprechend verwenden wir den U-Test von Mann und Whitney.

Wir erhalten für die tiefere Rangsumme 1596 (alle tiefsten Ränge gehören zur ersten Stichprobe; entsprechend wird das Resultat hochsignifikant sein). Wir berechnen den asymptotischen Z-Wert:

Da R_1 den Wert $\frac{n_1(n_1+1)}{2}$ annimmt. (d.h. die Daten der ersten Stichprobe besetzen alle ersten Ränge), nimmt U_2 den Wert $n_1n_2 + \frac{n_1(n_1+1)}{2} = 56 \cdot 63 + \frac{56 \cdot 57}{2} - \frac{56 \cdot 57}{2} = 56 \cdot 63 = 3528$ an. Entsprechend ist der Wert von $U_1 = n_1n_2 - U_2 : 0 \implies u_1 = 0$.

$$s = \sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{56 \cdot 63 \cdot (56+63+1)}{12}} = 187.83$$

$$\frac{n_1n_2}{2} = \frac{56 \cdot 63}{2} = 1764.$$

Z nimmt den Wert $\frac{0-1764}{187.83} = -9.3915$ an.

p-Wert: $\Phi(-9.3915) \approx 0 < 0.025$.

Das Ergebnis ist hoch signifikant.

3. H_1 : Die Vertreter unterscheiden sich nicht.

H_A : Die Vertreter unterscheiden sich ($\Rightarrow$ zweiseitiger Test)

Wir überprüfen auf Normalverteilung (s. Abbildungen 22.5.9 und 22.5.10):

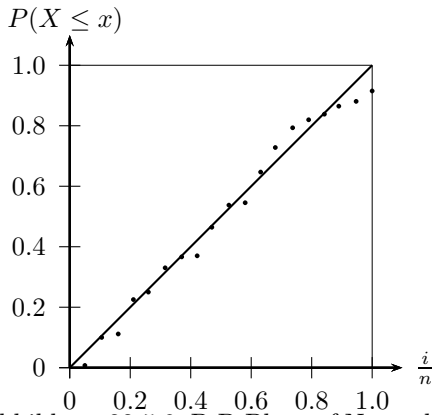


Abbildung 22.5.9: P-P-Plot auf Normalverteilung Stichprobe 1

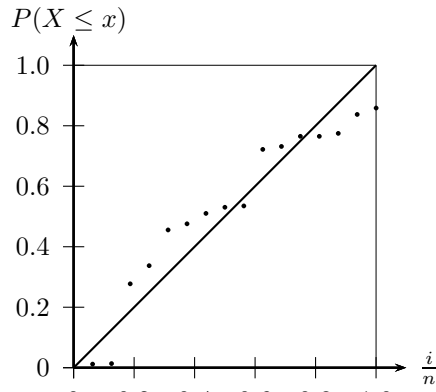


Abbildung 22.5.10: P-P-Plot auf Normalverteilung Stichprobe 2

Die Daten des zweiten Vertreters sind kaum normalverteilt. Wir müssen einen U-Test verwenden.

R_2 nimmt den Wert 250 an.

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_2 : \left(16 \cdot 19 + \frac{16(16+1)}{2} \right) - 250 = 190$$

$$U_2 : 16 \cdot 19 - 190 = 114 \Rightarrow u_2 = 114$$

Mit der Tabelle erhalten wir als kritischen Wert 92. Der Unterschied ist nicht signifikant.

Verwenden wir die Normalverteilung, erhalten wir:

$$s = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{16 \cdot 19 (16 + 19 + 1)}{12}} = 30.199$$

$$\frac{n_1 n_2}{2} = \frac{16 \cdot 19}{2} = 152$$

Z nimmt den Wert $\frac{190-152}{30.199} = 1.2583$ an.

$$\Phi(1.2583) = 1 - 0.895858294 = 0.104141 > 0.025 \text{ (oder: p-Wert: } 2 \cdot 0.10414 = 0.20828 > 0.05).$$

Das Resultat ist nicht signifikant.

Die beiden Vertreter unterscheiden sich nicht signifikant.

4. H_0 : Die Fonds-Manager unterscheiden sich nicht.

H_A : Die Fonds-Manager unterscheiden sich. ($\Rightarrow$ zweiseitiger Test, $\alpha = 0.05$)

Wir versuchen einen t -oder Welch- t -Test anzuwenden. Wir überprüfen die Daten auf Normalverteilung (s. Abbildungen 22.5.11 und 22.5.12):

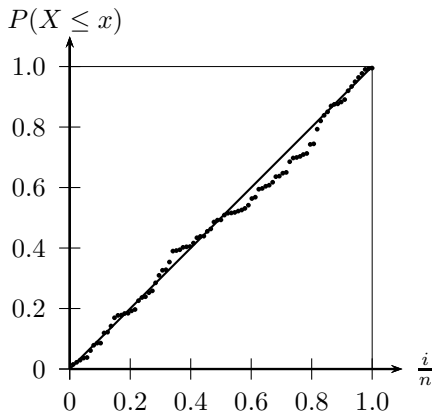


Abbildung 22.5.11: P-P-Plot auf Normalverteilung Stichprobe 1

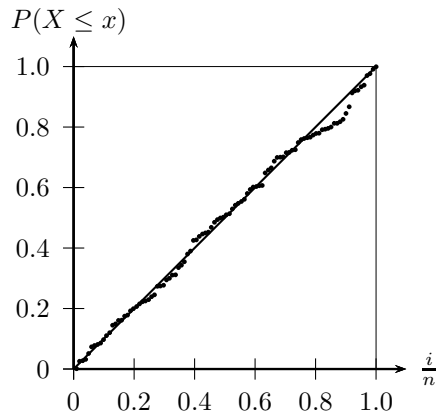


Abbildung 22.5.12: P-P-Plot auf Normalverteilung Stichprobe 2

Die Daten können als normalverteilt betrachtet werden.

Wir überprüfen die Homogenität der Varianzen: $\frac{25295974.46}{26889353} = 0.94074$

$F_{(87,100)}(0.94074) = 0.61363406$. Die Varianzen unterscheiden sich nicht signifikant. Wir können den t -Test verwenden:

Testwert:

$$\frac{4571.521962 - 3731.514532}{\sqrt{\frac{87 \cdot 25295974.46 + 100 \cdot 26889353}{87+100} \left(\frac{1}{88} + \frac{1}{101} \right)}} = 1.1265$$

$F_{t_{187}}(1.1265) = 0.8693 = 1 - 0.130698735$; p-Wert (zweiseitig): $2 \cdot 0.130698735 = 0.26140$

Der Unterschied ist nicht signifikant. Die Nullhypothese kann nicht abgelehnt werden.

5. H_0 : Die Diäten unterscheiden sich nicht.

H_A : Die Diäten unterscheiden sich ($\implies$ zweiseitiger Test; $\alpha = 0.05$)

Wir denken an einen t - oder Welch- t -Test. Wir überprüfen die Daten auf Normalverteilung: (s. Abbildungen 22.5.13 und 22.5.14):

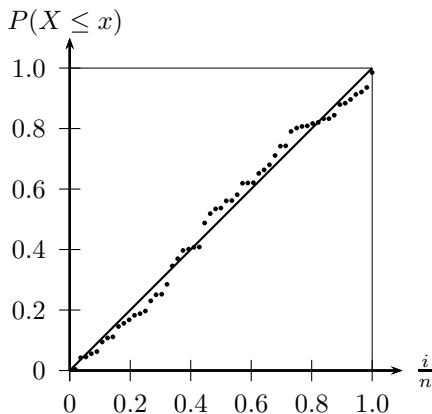


Abbildung 22.5.13: P-P-Plot auf Normalverteilung Stichprobe 1

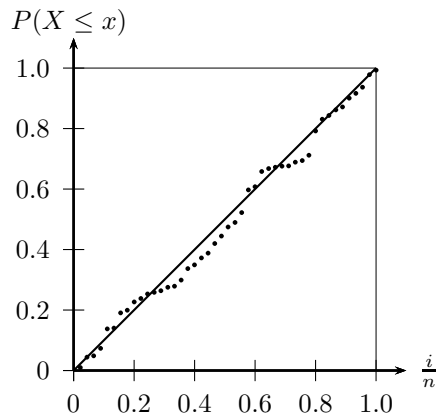


Abbildung 22.5.14: P-P-Plot auf Normalverteilung Stichprobe 2

Wir entscheiden, dass die Daten normalverteilt sind.

Wir überprüfen die Homogenität der Varianz:

$$\frac{37.22076984}{167.2132034} = 0.22259$$

$F_{55,44}(0.22259) = 1.31577E-07$ (mit Excel: $=F.VERT(0.22259;55;44;1)$)

Die Varianzen sind signifikant verschieden. Wir verwenden den Welch-Test:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \text{ nimmt Testwert } \frac{29.57668317 - 36.78387577}{\sqrt{\frac{37.22076984}{56} + \frac{167.2132034}{45}}} = -3.4435 \text{ an.}$$

$$R = \frac{S_x^2 m}{S_y^2 n} = \frac{37.22076984 \cdot 45}{167.2132034 \cdot 56} = 0.17887$$

$$df = \frac{(1+R)^2}{\frac{R^2}{(n-1)} + \frac{1}{m-1}} = \frac{(1+0.17887)^2}{\frac{0.17887^2}{(55)} + \frac{1}{44}} = 59.622$$

p -Wert = 0.001056. Der Unterschied ist signifikant.

Da beide Stichproben grösser als 50 sind, könnte man auch mit der Normalverteilung rechnen: Hälfte des p -Wertes: $\Phi(-3.4435) = 1 - 0.999712833 = 0.00028717 < 0.025$
Der Unterschied ist signifikant.

6. H_0 : Die Resultate von Männern und Frauen unterscheiden sich nicht.
 H_A : Die Frauen schneiden besser ab als die Männer. (einseitiger Test)
 Das Noten ordinalskaliert sind, müssen wir den Mann-Whitney-Test verwenden. Wir erhalten (s. Tabelle 22.5.10):

Geschlecht	Punkte geordnet	Nummerierung	Ränge	Ränge W	Ränge M
w	19.25	1	1	1	0
w	19	2	2	2	0
m	18	3	3.5	0	3.5
m	18	4	3.5	0	3.5
w	17.5	5	5	5	0
w	17	6	6	6	0
m	16.5	7	7	0	7
w	16.25	8	8	8	0
m	16	9	9	0	9
m	15.75	10	10	0	10
m	13.25	11	11	0	11
m	12.75	12	12	0	12
m	11.875	13	13	0	13
m	11	14	14	0	14
m	7.5	15	15	0	15
m	1.5	16	16	0	16
Summen		136	136	22	114

Tabelle 22.5.10: Lösungstabelle für die Übungsdaten

Rangmittel Stichprobe 1: $\frac{22}{5} = 4.4$

Rangmittel Stichprobe 2: $\frac{114}{11} = 10.364$. Die Rangmittel unterscheiden sich recht stark. Wir überprüfen, ob die Differenz gut oder schlecht durch Zufall erklärbar ist:

$$U_2 : 5 \cdot 11 + \frac{5(6)}{2} - 22 = 48$$

$$U_1 : 5 \cdot 11 - 48 = 7$$

$$\min(U_1, U_2) : \min(7, 48) = 7.$$

Wir schauen in der Tabelle nach: unter $n_1 = 5$ und $n_2 = 11$ steht 7. Da der kritische Wert zum Verwerfungsbereich gehört, kann die Nullhypothese verworfen werden.

7. Die Lösung können Sie am Beispiel 22.5.1 kontrollieren.
8. Für Genf: $x_1 = 15$
 $n_1 = 60$
 Für Zürich: $x_2 = 10$
 $n_2 = 70$
 $\frac{X_1 + X_2}{n_1 + n_2} = \frac{15 + 10}{60 + 70} = 0.19231$

Die Teststatistik $\frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\frac{X_1+X_2}{n_1+n_2} \left(1 - \frac{X_1+X_2}{n_1+n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ nimmt den Wert $\frac{\frac{15}{60} - \frac{10}{70}}{\sqrt{0.19231 \cdot (1-0.19231) \cdot \left(\frac{1}{60} + \frac{1}{70}\right)}} = 1.5452$ an.
 p-Wert: $2 \cdot \min\{\Phi(1.5452), 1 - \Phi(1.5452)\} = 2 \cdot (1 - 0.938851038) = 0.12230 > 0.05$
 Der Unterschied ist nicht signifikant

22.6 Lernziele

- Die angegebenen Zweistichprobentests für Datenmaterial korrekt auswählen und durchführen können (mit Excel und einer Statistiksoftware). Bei der Auswahl richtige Beurteilung der Voraussetzungen vornehmen können.
- Die Bemerkung (1) bei der letzten Aufgabe des letzten Übungsblockes kennen und anwenden können.
- Übungen von der Art der angegebenen lösen können.

Kapitel 23

Vertrauensintervalle bei Testwerten

Wir nehmen an, in einem spezifischen Test würde ein konkreter Testwert b zu einem signifikanten Resultat führen. So hätte etwa eine Mastmethode zu einem mittleren Gewicht von $\bar{x}$ geführt, das zur Verwerfung der Nullhypothese mit dem Parameter $\bar{x}_0$ geführt hat. Würden wir die neue Mastmethode auf alle Tiere der Art anwenden, würden wir somit ein neues $\bar{x}_A$ erhalten, da es unwahrscheinlich ist, dass $\bar{x}$ aus Zufallsvariablen mit dem Parameter $\bar{x}_0$ stammt. Wir können den konkreten Wert $\bar{x}$ des Testergebnisses als Wert der Stichprobe berechnet aus Zufallsvariablen mit dem Parameter $\bar{x}_A$ betrachten. Entsprechend ist $\bar{x}$ als Wert einer Zufallsvariable zu betrachten und schwankt um den tatsächlichen, aber unbekannten Parameter $\bar{x}_A$. Es wäre nun nützlich zu wissen, in welchem Intervall um $\bar{x}$ sich $\bar{x}_A$ mit einer angebbaren Wahrscheinlichkeit β befindet.

Erhalten wir einen spezifischen Testwert b , können wir uns fragen, welche Verteilungsparameter b' mit diesem Wert verträglich sind (= **nicht** zur Verwerfung der jeweiligen Nullhypothese, dass die Stichprobe mit dem Testwert b aus Zufallsvariablen mit dem Parameter b' stammt). Es sind ja genau diese Werte, zu denen b als Testresultat passt. Bei anderen Werten wäre b als Testresultat unwahrscheinlich. Deshalb ist es angemessen, „Vertrauensintervall“ wie folgt zu definieren:

Definition 23.0.1. $]T_0, T_1[= \{b' : b' \text{ führt bei einem Testergebnis } b \text{ der Teststatistik } B \text{ bei einem Signifikanzniveau von } \alpha = 1 - \beta \text{ nicht zur Verwerfung der Nullhypothese, dass die Stichprobe, welche zum Wert } b \text{ der Teststatistik } B \text{ führt, aus Zufallsvariablen mit dem Parameter } b' \text{ resultiert}\}$ (bei zweiseitigem Testen)

$]T_0, T_1[$ wird das β -Vertrauensintervall oder β -Konfidenzintervall von b genannt. T_0, T_1 sind Zufallsvariablen, welche als konkrete Werte die Intervallgrenzen t_i annehmen.

Die Grenzen T_i des Vertrauensintervalls sind die Parameter, die genau zur Verwerfung der besagten Null-Hypothese führen. Die Intervallgrenzen T_i gehören bei dieser Definition nicht zum Vertrauensintervall.

Die Angabe eines Testwertes ohne die Angabe eines Intervalls, in dem der tatsächliche, aber unbekannte Parameter mit angebbarer Wahrscheinlichkeit liegt, ist eigentlich nicht besonders informativ. Ist dieses Intervall gross, so ist der Unterschied zwischen der Nullhypothese und der Alternativhypothese unter Umständen ökonomisch nicht relevant. Ein Wechsel der Produktionsmethoden würde in einem solchen Fall nur Kosten verursachen, ohne ökonomisch interessante Verbesserungen. Wir leiten Vertrauensintervalle nicht für alle behandelten Tests her, sondern beschränken uns auf drei Beispiele.

23.1 Vertrauensintervalle für den Einstichproben z-Test

Es sei X_i normalverteilt ($1 \leq i \leq n$) oder n genügend gross und s_0^2 bekannt. Wir hätten im Test einen konkreten Mittelwert $\bar{x}$ als Wert der Zufallsvariable $\bar{X}$ erhalten. Wir können uns fragen, wel-

che $\bar{x}'$ mit diesem $\bar{x}$ verträglich sind (= bei einem Signifikanzniveau von α bei zweiseitigem Testen **nicht** zur Verwerfung der Nullhypothese $\bar{x} = \bar{x}'$ führen; dies entspricht einem Vertrauensintervall von $\beta = 1 - \alpha$). Laut Voraussetzungen gilt:

$$\frac{\bar{X} - \bar{x}'}{\frac{s_0}{\sqrt{n}}} \sim N(0, 1).$$

$\bar{x}'$ führt nicht zur Verwerfung der Nullhypothese, dass $\bar{x} = \bar{x}'$ ist genau dann, wenn

$$\frac{1 - \beta}{2} < \Phi\left(\frac{\bar{x} - \bar{x}'}{\frac{s_0}{\sqrt{n}}}\right) < \beta + \frac{1 - \beta}{2}$$

($\beta = 1 - \alpha$; α = Signifikanzniveau, z.B. $\beta = 0.95 \implies \alpha = 0.05$; $\frac{1-\beta}{2} = 0.025$; $\beta + \frac{1-\beta}{2} = 0.975$)

Da Φ eine strikt monoton steigende Funktion ist, können wir die Werte der Umkehrfunktion berechnen:

$$\Phi^{-1}\left(\frac{1 - \beta}{2}\right) < \frac{\bar{x} - \bar{x}'}{\frac{s_0}{\sqrt{n}}} < \Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right)$$

Um die beiden Intervallgrenzen t_i zu finden, ersetzen wir $\bar{x}'$ durch t_i - es sind genau die $\bar{x}'$, welche zur Verwerfung der Nullhypothese $\bar{x}' = \bar{x}$ führen, und ersetzen das Ungleichheitszeichen durch das Gleichheitszeichen. Wir erhalten:

$$\Phi^{-1}\left(\frac{1 - \beta}{2}\right) = \frac{\bar{x} - t_i}{\frac{s_0}{\sqrt{n}}}$$

und

$$\frac{\bar{x} - t_i}{\frac{s_0}{\sqrt{n}}} = \Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right).$$

Diese Gleichungen lösen wir nach t_i auf:

(i)

$$\bar{x} - t_1 = \Phi^{-1}\left(\frac{1 - \beta}{2}\right) \frac{s_0}{\sqrt{n}} \implies t_i = \bar{x} - \Phi^{-1}\left(\frac{1 - \beta}{2}\right) \frac{s_0}{\sqrt{n}}$$

(ii)

$$\bar{x} - t_i = \Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right) \frac{s_0}{\sqrt{n}} \implies t_i = \bar{x} - \Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right) \frac{s_0}{\sqrt{n}}$$

Nun gilt wegen der y-Achsen-Symmetrie der Standardnormalverteilung:

$$\Phi^{-1}\left(\frac{1 - \beta}{2}\right) = -\Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right).$$

Wir können somit für die Grenzen des Vertrauensintervalls zusammenfassen:

$$t_i = \bar{x} \pm \Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right) \frac{s_0}{\sqrt{n}}$$

$\bar{x}$ ist Wert der Zufallsvariablen $\bar{X}$. Setzen wir für $\bar{x}$ die entsprechende Zufallsvariable, erhalten wir die Intervallgrenzen als Zufallsvariablen:

$T_i = \bar{X} \pm \Phi^{-1} \left(\beta + \frac{1-\beta}{2} \right) \frac{s_0}{\sqrt{n}}$
β = Wahrscheinlichkeit, dass wirklicher Wert in Intervall ist
$\bar{X}$ = Zufallsvariable mit dem Stichprobenmittel $\bar{x}$ als Wert
T_i = Zufallsvariable Intervallgrenze des Vertrauensintervalls, die als Wert t_i annimmt.
s_0 = Standardabweichung der Zufallsvariablen unter der Nullhypothese
$\Phi^{-1} \left(\beta + \frac{1-\beta}{2} \right) = \beta + \frac{1-\beta}{2}$ – Quantil der Standardnormalverteilung

Für ein gesuchtes Vertrauensintervall $\beta \Phi^{-1} \left(\beta + \frac{1-\beta}{2} \right)$ berechnen wir mit Excel:

$$= \text{norm.inv}(\beta + \frac{1-\beta}{2}, 0, 1).$$

Für $\beta = 0.95$ (95%-Vertrauensintervalle) müssen wir somit $\Phi^{-1}(0.975) = 1.96$ berechnen (=norm.inv(0.975;0;1)). Entsprechend erhalten wir als Grenze des 95%-Vertrauensintervalls

$$T_i = \bar{X} \pm \Phi^{-1} \left(\beta + \frac{1-\beta}{2} \right) \frac{s_0}{\sqrt{n}} = \bar{X} \pm 1.96 \frac{s_0}{\sqrt{n}}$$

da

$$\Phi^{-1} \left(0.95 + \frac{1-0.95}{2} \right) \approx 1.96.$$

Wir können davon ausgehen, dass der tatsächliche Parameter sich mit einer Wahrscheinlichkeit β im Vertrauensintervall um den Testwert herum befindet.

Beispiel 23.1.1. Wir nehmen an, wir hätten den Testwert $\bar{x} = 10$ (Mittelwert der Stichprobe) erhalten und wir kennen die Standardabweichung der Zufallsvariablen unter der Nullhypothese: $s_0 = 0.5$ ($n = 30$): Wir erhalten als Grenzen des Vertrauensintervalls: $10 + 1.96 \frac{0.5}{\sqrt{30}} = 10.179$

$10 - 1.96 \frac{0.5}{\sqrt{30}} = 9.8211$. Das 95%-Vertrauensintervall ist somit $]9.8211, 10.179[$.

Das 98%-Vertrauensintervall ist demgegenüber mit $\Phi^{-1}(0.98 + \frac{1-0.98}{2}) = \Phi^{-1}(0.99) = 2.326347874$

$10 + 2.326347874 \frac{0.5}{\sqrt{30}} = 10.212$ und

$10 - 2.326347874 \frac{0.5}{\sqrt{30}} = 9.7876$:

$]9.7876, 10.212[$.

Einleuchtender Weise gilt $]9.8211, 10.179[\subset]9.7876, 10.212[$

◇

Bemerkung 23.1.2. Man sieht an der Formel $\bar{X} \pm 1.96 \frac{s_0}{\sqrt{n}}$ unmittelbar, dass das Vertrauensintervall grösser wird, wenn die Standardabweichung grösser ist. Zudem ist das Vertrauensintervall kleiner, wenn n grösser ist: je grösser die Stichprobe, desto kleiner das Vertrauensintervall. ◇

Am Beispiel 23.1.1 können wir die Grundüberlegungen bezüglich der Vertrauensintervalle noch etwas anschaulicher machen. Wir können uns die Verteilung um 10 herum denken. Die Grenzen des Vertrauensintervalls sind dann die T_i , die symmetrisch um 10 zwischen der x -Achse und der Dichtefunktion der Normalverteilung mit $\bar{x} = 10$ und $s_0 = \frac{0.5}{\sqrt{30}}$ eine Fläche von 0.95 ausschneiden (s. Abbildung 23.1.1):

Bemerkung 23.1.3. Vertrauensintervalle sind informativer als das Testresultat „signifikant“ oder „nicht-signifikant“. Es gilt offenbar: die Abweichung eines Testwertes vom Parameter der

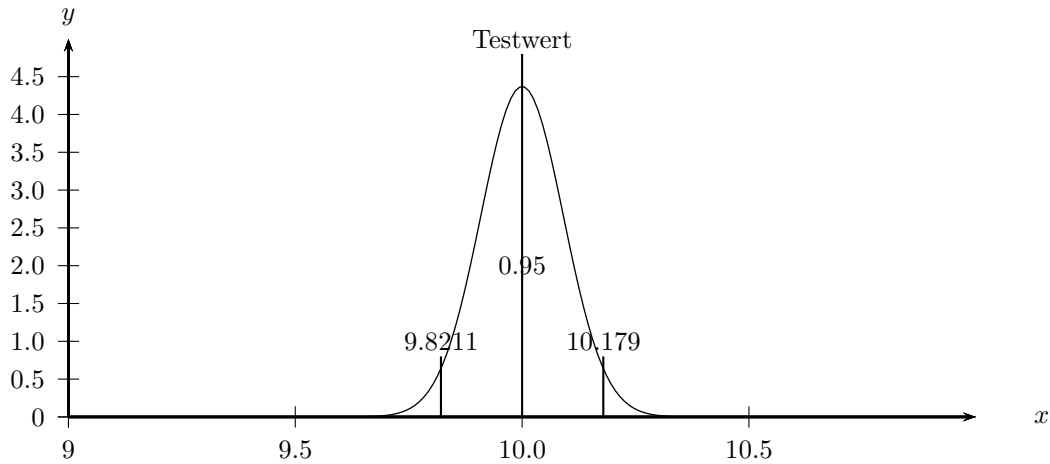


Abbildung 23.1.1: Vertrauensintervall am Beispiel einer normalverteilten Statistik

Nullhypothese ist genau dann signifikant, wenn bei zweiseitigem Testen der Parameter ausserhalb des Vertrauensintervalls um den Testwert liegt. Somit kann man aus Vertrauensintervallen Signifikanz unmittelbar herauslesen, während man aus dem Ergebnis „signifikant“ allein keine Vertrauensintervalle herleiten kann. $\diamond$

23.2 Vertrauensintervalle für den Einstichproben t-Test

Es gelte $X_i \sim N(\mu, \sigma^2)$ mit unbekanntem σ^2 und für $1 \leq i \leq n$.

Wir hätten einen t-Test verwendet: $\frac{\bar{X} - \bar{x}_0}{\frac{s}{\sqrt{n}}}$ ist t_{n-1} -verteilt.

Bei einem Testergebnis von $\bar{x}$ und einer geschätzten Standardabweichung von s wird die Nullhypothese mit $\bar{x}'$ bei zweiseitigem Testen nicht verworfen, wenn

$$F_{t_{(n-1)}}^{-1} \left(\frac{1-\beta}{2} \right) < \frac{\bar{x} - \bar{x}'}{\frac{s}{\sqrt{n}}} < F_{t_{(n-1)}}^{-1} \left(\beta + \frac{1-\beta}{2} \right).$$

Wir formen wie oben um, ersetzen s durch die Zufallsvariable S , wobei s Wert von S ist, $\bar{x}$ durch $\bar{X}$, $\bar{x}'$ durch T_i , und erhalten:

$$T_i = \bar{X} \pm F_{t_{(n-1)}}^{-1} \left(\beta + \frac{1-\beta}{2} \right) \frac{S}{\sqrt{n}},$$

Um die Werte von T_i zu berechnen, setzen wir für $\bar{X}$ den konkreten Testwert $\bar{x}$ und für S die konkrete Standardabweichung s der Stichprobe ein. Excel liefert bei der Eingabe von „=t.inv($\beta + \frac{1-\beta}{2}$;n-1)“ den Wert von

$$F_{t_{(n-1)}}^{-1} \left(\beta + \frac{1-\beta}{2} \right)$$

Beispiel 23.2.1. Sei $n = 50$; $\bar{x} = 20$; $s = 2$.

Wir suchen mit Hilfe von Excel $F_{t_{(50-1)}}^{-1}(0.975) := 2.009574018$. (Im Beispiel müssen wir mit Excel „=t.inv(0.975; 49)“ Wir erhalten somit: $20 \pm 2.009574018 \frac{2}{\sqrt{50}} = 20.568$ für „+“ und 19.432 für „-“.

Das Vertrauensintervall ist somit $]19.432, 20.568[$

Das 99%-Vertrauensintervall erhalten wir analog mit $F_{t(50-1)}^{-1}(0.995)$. Zeigen Sie im Beispiel, dass das 95%-Vertrauensintervall Teilmenge des 99%-Vertrauensintervalls ist. $\diamond$

Bei asymmetrischen Verteilungen (z.B. $F_{n,m}$ - und χ^2 -Verteilungen) müssen wir die rechte und die linke Intervallgrenze separat berechnen. Wir betrachten noch ein Beispiel mit einer diskreten Verteilung:

23.3 Vertrauensintervalle bei poisson-verteilter Teststatistik

Bei der Messung von 10 Proben Luft würden wir 45 Asbestfasern erhalten (Grenzwert 30). Wir berechnen: $P(X \geq 45) = 1 - 0.993731385 = 6.2686 \times 10^{-3}$ für $X \sim P(30)$. Wir fragen uns, welcher Parameter λ in $X \sim P(\lambda)$ mit dem Testwert 45 verträglich wären, d.h. nicht zur Verwerfung der entsprechenden Nullhypothese führen (auf einem spezifischen Signifikanzniveau, z.B. 0.05, zweiseitig).

Wir verwenden ein Näherungsverfahren. Wir berechnen $P(X \geq 45)$ für $\lambda = 30, 31, 32, 33, \text{etc.}$ um eine erste Näherung zu finden. Auf der anderen Seite berechnen wir $P(X \leq 45)$ für $\lambda = 56, 57, 58, 59, \text{etc.}$ (s. Tabelle 23.3.1) :

$\lambda =$	$P(X \geq 45)$	$\lambda =$	$P(X \leq 45)$
30	0.006268615	56	0.076688247
31	0.010646775	57	0.059814016
32	0.017296896	58	0.046139355
33	0.026961979	59	0.035207352
34	0.040438338	60	0.026582065
35	0.058509359	61	0.019862704
36	0.081865376	62	0.014692073
37	0.111019769	63	0.010760263

Tabelle 23.3.1: Berechnung des Vertrauensintervalls einer poisson-verteilten Statistik

Beispiel 23.3.1. Wir erhalten: $P(X \geq 45) = 0.017296896$ für $X \sim P(32)$ als letztem λ , das zu einem Wert unter 0.025 und damit zur Verwerfung der Nullhypothese $\lambda = 45$ führt. Auf der anderen Seite von 45 gilt: $P(X \leq 45) = 0.0198627$ für $X \sim P(61)$ als erstem λ , das zu einem Wert unter 0.025 und damit zur Verwerfung der Nullhypothese $\lambda = 61$ führt. Wir können die Grenzen nun verfeinern, indem wir rechnen:

$\lambda =$	$P(X \geq 45)$	$\lambda =$	$P(X \leq 45)$
32.1	0.018114821	60.1	0.025830616
32.2	0.018963636	60.2	0.025097813
32.3	0.019844132	60.3	0.02438328
32.4	0.020757107	60.4	0.023686648
32.5	0.021703359	60.5	0.023007554
32.6	0.022683692	60.6	0.022345637
32.7	0.023698913	60.7	0.021700542
32.8	0.024749828	60.8	0.021071918
32.9	0.025837247	60.9	0.020459419
33	0.026961979	61	0.019862704

Tabelle 23.3.2: Fortführung der Berechnung des Vertrauensintervalls einer poisson-verteilten Statistik, s. Tabelle 23.3.1

Das Vertrauensintervall für 45 wäre entsprechend zwischen den Grenzen 32.8 und 60.3. Die Grenzen könnten noch weiter verfeinert werden. Wir können, wenn die Genauigkeit genügt, sagen: das tatsächliche λ befindet sich mit einer Wahrscheinlichkeit von 95% im Intervall $]32.8, 60.3[$, wenn der Testwert 45 beträgt. $\diamond$

Bemerkung 23.3.2. Man könnte auch für Chi-Quadratunabhängigkeitstests und Chi-Quadratanteilstests Vertrauensintervalle bezüglich des Testwertes berechnen. Allerdings macht das dort wenig Sinn, da man für die Praxis solche für die Häufigkeiten bräuchte, nicht jedoch für den Testwert. Die Berechnung von Vertrauensintervallen für die Häufigkeiten könnten mittels der Formeln für die standardisierten Pearson-Residuen angegeben werden. Für die übrigen behandelten Tests werden Vertrauensintervalle von den Computerprogrammen ausgegeben. Es genügt, den entsprechenden Output interpretieren zu können. $\diamond$

23.3.1 Übungen

1. Berechnen Sie das 95%-Vertrauensintervall für $n = 10$; empirische Varianz: $s^2 = 11.55433333$; $\bar{x}_0 = 55$, $\bar{x} = 49.61$ (siehe Beispiel 20.2.1 für Einstichproben-t-Test)
2. Berechnen Sie das 95%-Vertrauensintervall für das Beispiel 20.1.1, Einstichprobentests (z-Test).
3. Berechnen Sie das 95%-Vertrauensintervall für das Beispiel 21.1.1, verbundene Stichproben.

23.3.2 Lösungen

1. $F_{t(9)}^{-1}(0.975) \approx 2.262$

Wir erhalten somit: $49.61 + 2.262 \sqrt{\frac{11.5543333}{10}} = 52.041$

$$49.61 - 2.262 \sqrt{\frac{11.5543333}{10}} = 47.179$$

Das Vertrauensintervall ist somit $]47.179; 52.041[$

2. Indem wir die obigen Formeln verwenden gilt: $\bar{X} \pm 1.96 \frac{s_0}{\sqrt{n}} \Rightarrow$
 $20.5 + 1.96 \frac{2}{\sqrt{50}} = 21.052$
 $20.5 - 1.96 \frac{2}{\sqrt{50}} = 19.948$
 $\Rightarrow]19.9456283, 21.0548717[$

3. Indem wir die obigen Formeln verwenden, gilt: $\bar{X} \pm F_{t_{(n-1)}}^{-1}(0.975) \frac{s}{\sqrt{n}}$;
 $F_{t_{(16)}}^{-1}(0.975) = 2.12$
 Wir erhalten somit: $-0.891718213 + 2.12 \cdot \frac{0.040089677}{\sqrt{17}} = -0.87111$
 $-0.891718213 - 2.12 \cdot \frac{0.040089677}{\sqrt{17}} = -0.91233$
 $] -0.91233, -0.87111[.$

23.4 Lernziele

- Vertrauensintervalle für den Einstichproben-z-Test und -t-Test berechnen und interpretieren können.
- Vertrauensintervalle auf Computerausgaben bei den übrigen behandelten Tests interpretieren können.

Teil V

Bivariate Einstichprobentests

Einleitung

In der bivariaten schliessenden Statistik geht es um den Zusammenhang von zwei Variablen. Rauchen Männer mehr als Frauen (Variable „Geschlecht“ und Variable „gerauchte Zigaretten pro Tag“)? Kaufen Junge mehr Süssigkeiten (Variable „Alter“, Variable „Süssigkeiten in Franken“). Gibt es einen Zusammenhang zwischen Einkommen und Ausgaben für Schmuck?

Gewöhnlich wird eine der Variablen als unabhängige (beeinflussende) Variable betrachtet, während die andere als die zu erklärende Variable festgelegt wird. Dabei ergibt sich das Problem, dass Variablen statistisch zusammenhängen können, ohne dass ein „wirklicher“ und „kausaler“ Zusammenhang besteht. Ein klassisches Beispiel kann dies klar machen: der Geburtenrückgang in Westeuropa im 20. Jahrhundert erfolgte im Gleichschritt mit dem Rückgang der Störche. Es scheint aber unvernünftig zu sein, den Rückgang der Störche als „Ursache“ für den Rückgang der Geburten zu betrachten, oder den Rückgang der Geburten als „Ursache“ für den Rückgang der Störche. Es gibt im vorliegenden Fall eine dritte „Ursache“ - die Industrialisierung, die auf beide Variablen einwirkt. Allerdings ist der Begriff der Ursache nicht präzise zu fassen. A ist Ursache von B , wenn wir A auf dem Hintergrund eines bestimmten historischen Wissenstandes zur Erklärung von B heranziehen. Das wissenschaftliche Fortschreiten kann jedoch dazu führen, dass solches Erklären als Scheinerklärung erkannt wird und man neue „Ursachen“ für B anführt (Wir wissen genug über die Fortpflanzung, so dass wir nicht in Versuchung geraten, den Geburtenrückgang durch den Storchrückgang zu erklären. Es gibt aber genügend Wissensgebiete, wo wir wenig wissen, und wo wir auf Grund statistischer Zusammenhänge Gefahr laufen, Kausalität zu postulieren, wo diese vielleicht nicht vorliegt - wenn man wüsste, was „Kausalität“ überhaupt genau ist).

Je nach der Skalierung der Variablen gibt es verschiedene Methoden, um die Unabhängigkeit oder die Abhängigkeit zweier Variablen zu testen.

- Untersuchen wir den Einfluss einer nominalskalierten Variable mit n Ausprägungen auf eine nominalskalierte Variablen mit m Ausprägungen, so wird der χ^2 -Test verwendet (z.B. Frauen, Männer - Raucher, Nichtraucher). Als Alternative bieten sich sogenannte exakte Tests an.
- Untersuchen wir den Einfluss einer nominalskalierten Variable mit mehr als zwei Ausprägungen auf eine metrisch skalierte Variable, so verwenden wir oft die Varianzanalyse (z.B. drei Sorten Weizen werden auf ihre Erträge hin untersucht).
- Untersuchen wir den Einfluss einer metrisch skalierten Variable auf eine metrisch skalierte Variable, verwenden wir die Regressionsanalyse (Einkommen in Franken und Ausgaben für Schmuck in Franken).

Damit haben wir noch nicht alle möglichen Kombinationen angeführt: möglich sind noch:

- ordinalskaliert $\rightarrow$ ordinalskaliert
- ordinalskaliert $\rightarrow$ metrisch skaliert
- ordinalskaliert $\rightarrow$ nominalskaliert
- nominalskaliert $\rightarrow$ ordinalskaliert
- metrisch skaliert $\rightarrow$ ordinalskaliert
- metrisch skaliert $\rightarrow$ nominalskaliert

In der zweiten Hälfte des 20. Jahrhunderts wurden bivariate Methoden entwickelt, die solche Probleme zu lösen vermögen, auf die wir jedoch im Allgemeinen nicht eingehen können (weiterführende Literatur: Agresti, A. (2013). *Categorical Data Analysis*, New York, John Wiley. Das Buch führt auch in die multivariate Statistik nominal- und ordinalskalierter Daten ein, ist allerdings recht anspruchsvoll).

Die in Teil IV behandelten Zweistichprobentests sind übrigens formal betrachtet bivariate Tests, da jeweils bei der Auswertung zwei Variablen ins Spiel kommen. Inhaltlich gesehen, handelt es sich aber eher um univariate Analysen.

Kapitel 24

Testen auf Unabhängigkeit in Kreuztabellen und bivariate Anpassungstests

24.1 χ^2 –Unabhängigkeitstests

Ziel: Es soll überprüft werden, ob zwei statistische Variablen unabhängig sind (d.h. ob die Verteilung der Häufigkeit von Objekten den laut Randverteilung erwarteten Häufigkeiten entspricht oder ob es bedeutsame Abweichungen von den erwarteten Häufigkeiten gibt).

Voraussetzungen: Die erwartete Häufigkeit pro Feld sollte nicht bei mehr als 20% der Felder kleiner sein als 5. (Vor allem für nominalskalierte Daten verwendet). Die Zufallsvariablen sind identisch verteilt und unabhängig.

Da nur ein nominales Skalenniveau vorausgesetzt wird, wird der χ^2 –Test (auch χ^2 –Unabhängigkeitstest genannt) in den Sozial- und Wirtschaftswissenschaften häufig angewendet. Da auch höher skalierte Daten als nominalskalierte aufgefasst werden können (bei Inkaufnahme von Informationsverlusten), sind χ^2 –Test auch auf Daten mit höheren Skalenniveaus anwendbar.

Beispiel 24.1.1. Wir betrachten zwei Variablen Rauchverhalten und Geschlecht mit je zwei Ausprägungen (in diesem Falle wäre der exakte Test von Fisher, der später behandelt wird, besser). Wir ziehen aus der schweizerischen Bevölkerung eine Zufallsstichprobe. Wir nehmen an, man hätte eine Stichprobe von 100 Personen gezogen, darunter 46 Männer und 54 Frauen. Davon rauchen 12 Männer und 11 Frauen. Kann der Unterschied in der Stichprobe zwischen Frauen und Männer durch Zufall erklärt werden oder weisen die Frauen und Männer insgesamt unterschiedliches Rauchverhalten auf?

H_0 : Männer und Frauen weisen nicht ein unterschiedliches Rauchverhalten auf.

H_A : Männer und Frauen weisen ein unterschiedliches Rauchverhalten auf. ($\alpha = 0.05$).

(s. Tabelle 24.1.1):

		Rauchverhalten		Summen
		Raucher	Nichraucher	
Geschlecht	Männer	12	34	46
	Frauen	11	43	54
Summen		23	77	100

Tabelle 24.1.1: Beispielkreuztabelle für Berechnung eines bivariaten Chi-Quadrattestes

Dies stellt eine bivariate absolute Häufigkeitsverteilung dar ($I \times J$ Kreuztabelle mit $I = 2$ Zeilen und $J = 2$ Spalten). Die Zahlen in den einzelnen Zellen kürzen wir mit k_{ij} ab (i -te Zeile und j -te

Spalte, so ist $k_{12} = 34, k_{11} = 12; k_{21} = 11; k_{22} = 43$). Rechts stehen in der Spalte unter „Summen“ die jeweiligen Zeilensummen

$$k_{i+} := \sum_{j=1}^J k_{ij}$$

(k_{i+} für Randsumme der Zeile i). Unten stehen in der Zeile neben „Summen“ die jeweiligen Spaltensummen

$$k_{+j} := \sum_{i=1}^I k_{ij}$$

(= Randsumme der Spalte j).

Unterscheidet sich das Rauchverhalten der Frauen nicht vom Rauchverhalten der Männer, so müssten sich bei einer Tabelle, die genau den Randverteilungen entspricht, folgende erwartete absolute Häufigkeiten ergeben: Bei jeweils anteilmässiger Verteilung der Objekte sollte der Anteil der Raucher unter den Männern dem Anteil der Männer an der Stichprobe entsprechen. Da 46/100 Männer sind, sollten sie unter den Rauchern (23) genau diesen Anteil ausmachen. Dies ergibt:

für die rauchenden Männer $\frac{46}{100} \cdot 23 = 10.58$

für die rauchenden Frauen ergibt sich $\frac{54}{100} \cdot 23 = 12.42$

für die nichtrauchenden Männer $\frac{46}{100} \cdot 77 = 35.42$

für die nichtrauchenden Frauen $\frac{54}{100} \cdot 77 = 41.58$

Dies gilt, wenn die Verteilung in den Zellen der Kreuztabelle durch die Randverteilungen bestimmt ist. Wir sagen in diesem Fall: die Variablen sind unabhängig.

Allgemein gilt damit für die bei Unabhängigkeit der Variablen erwartete, absolute Häufigkeit der Zelle i, j einer $I \times J$ Kreuztabelle:

$$k_{ij}^e = \frac{k_{i+} k_{+j}}{n}$$

mit n = Stichprobengrösse total. Wir erhalten für das Beispiel folgende Tabelle 24.1.2 erwarteter absoluter Häufigkeiten:

		Rauchverhalten		Summen
		Raucher	Nichtraucher	
Geschlecht	Männer	10.58	35.42	46
	Frauen	12.42	41.58	54
Summen		23	77	100

Tabelle 24.1.2: Tabelle der erwarteten Häufigkeiten bei Unabhängigkeit für Kreuztabelle 24.1.1

◇

Je mehr die Verteilung der Daten von den erwarteten Häufigkeiten abweicht, desto weniger sind sie mit der Hypothese der Unabhängigkeit der beiden Variablen verträglich. Es gilt:

$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(k_{ij} - k_{ij}^e)^2}{k_{ij}^e}$ ist asymptotisch $\chi_{(I-1) \cdot (J-1)}^2$ -verteilt für $k_{ij}^e < 5$ für
höchstens 20% der Felder
$K_{i,j}$: Zufallsvariable, die als Werte die empirischen absolute Häufigkeiten k_{ij} annimmt.
k_{ij}^e : erwartete absolute Häufigkeiten
I : Anzahl der Zeilen
J : Anzahl der Spalten

Dies gilt näherungsweise und für $k_i^e \geq 5$ in mindestens 80% der Felder. Die Anzahl der Freiheitsgrade der χ^2 -Verteilung beträgt: $(I-1) \cdot (J-1)$ (Anzahl der Zeilen weniger 1 multipliziert mit der Anzahl der Spalten weniger 1).

Im obigen Beispiel erhalten wir:

$$\begin{aligned} X^2 &= \sum_{i=1}^{I=2} \sum_{j=1}^{J=2} \frac{(k_{i,j} - k_{i,j}^e)^2}{k_{i,j}^e} \\ &= \frac{(12 - 10.58)^2}{10.58} + \frac{(11 - 12.42)^2}{12.42} + \frac{(34 - 35.42)^2}{35.42} + \frac{(43 - 41.58)^2}{41.58} \\ &= 0.45836 \end{aligned}$$

Die $\chi^2_{(2-1)(2-1)}$ -Verteilung ordnet $[0.45836, \infty[$ die Wahrscheinlichkeit 0.498391429 (= p-Wert) zu. Dieser ist grösser als 0.05. Die Unterschiede zwischen Frauen und Männern bezüglich Rauchverhalten sind damit (im Beispiel) nicht signifikant.

Um die Abweichungen der erwarteten von den faktischen absoluten Häufigkeiten zu untersuchen, betrachten wir eine Tabelle mit den sogenannten *Residuen*

$$k_{ij} - k_{ij}^e.$$

Es handelt sich um die Differenzen der faktischen und der erwarteten absoluten Häufigkeiten. Im obigen Beispiel würden wir erhalten (s. Tabelle 24.1.3): Auch hier ist es besser, die Pearson-

		Rauchverhalten		Summen
		Raucher	Nichtraucher	
Geschlecht	Männer	1.42	-1.42	0
	Frauen	-1.42	1.42	0
Summen		0	0	0

Tabelle 24.1.3: Tabelle der Residuen $k_{ij} - k_{ij}^e$ der Tabelle 24.1.1

Residuen

$$\frac{k_{ij} - k_{ij}^e}{\sqrt{k_{ij}^e}}$$

oder die standardisierten Pearson-Residuen

$$\frac{k_{ij} - k_{ij}^e}{\sqrt{k_{ij}^e (1 - p_{i+}) (1 - p_{+j})}}$$

mit $p_{i+} := \frac{k_{i+}}{n}$ und $p_{+j} := \frac{k_{+j}}{n}$ zu berechnen. Die Standardisierten Pearson-Residuen sind näherungsweise standardnormalverteilt. Abweichungen von mehr als 2 oder -2 von 0 können auf dem 0.5-Niveau als signifikant gelten, wenn im Voraus entsprechende Hypothesen formuliert wurden. Im Beispiel würde man für die Pearson-Residuen erhalten (für die Zelle (1,1) : $\frac{k_{11} - k_{11}^e}{\sqrt{k_{11}^e}} = \frac{12 - 10.58}{\sqrt{10.58}} = 0.4365616$):

	Raucher	Nichtraucher
Männer	0.4365616	-0.2385965
Frauen	-0.4029281	0.2202146

und die standardisierten Pearson-Residuen (für die Zelle (1,1): $\frac{k_{11} - k_{11}^e}{\sqrt{k_{11}^e (1 - p_{1+}) (1 - p_{+1})}} = \frac{12 - 10.58}{\sqrt{10.58 (1 - \frac{46}{100}) (1 - \frac{23}{100})}} = 0.6770228$):

	Raucher	Nichtraucher
Männer	0.6770228	-0.6770228
Frauen	-0.6770228	0.6770228

Keine Abweichung ist signifikant.

Beispiel 24.1.2. In einer Firma wird mit verschiedenen Arbeitsformen experimentiert (keine Gruppenarbeit, mittlere Gruppenarbeit und vorwiegend Arbeit in Gruppen). Die Arbeitnehmer der verschiedenen Arbeitsformen werden befragt, wie sich ihre Arbeitssituation verändert habe. Dabei wurden folgende Daten erhoben (s. 3×3 Kreuztabelle 24.1.4).

		Veränderung der Arbeitssituation			Summen
		verbessert	unverändert	verschlechtert	
Realisierungs-niveau	hoch	101	48	17	166
von	mittel	41	43	8	92
Gruppenarbeiten	gering	20	36	15	71
Summen		162	127	40	329

Tabelle 24.1.4: Kreuztabelle mit Beispieldaten für (suboptimale) Anwendung des Chi-Quadratstest

H_0 : Die faktischen Häufigkeiten weichen nicht von den erwarteten Häufigkeiten ab (d.h. dass die Ausprägungen des Realisierungsniveaus von Gruppenarbeit keinen Einfluss auf die Veränderung der Arbeitssituation haben).

H_A : Die faktischen Häufigkeiten weichen von den erwarteten Häufigkeiten ab. (Die Ausprägungen des Realisierungsniveaus von Gruppenarbeit haben mindestens auf manchen Ausprägungsstufen einen Einfluss auf die Veränderung der Arbeitssituation).

Wir erhalten die folgende Tabelle 24.1.5 für die erwarteten Häufigkeiten:

		Veränderung der Arbeitssituation			Summen
		verbessert	unverändert	verschlechtert	
Realisierungs-niveau	hoch	81.73860182	64.07902736	20.18237082	166
von	mittel	45.30091185	35.51367781	11.18541033	92
Gruppenarbeiten	gering	34.96048632	27.40729483	8.632218845	71
Summen		162	127	40	329

Tabelle 24.1.5: Erwartete Häufigkeiten für die Kreuztabelle 24.1.4

Wir berechnen den Wert der Teststatistik:

$$\begin{aligned}
 X^2 &= \sum_{i=1}^I \sum_{j=1}^J \frac{(k_{i,j} - k_{i,j}^e)^2}{k_{i,j}^e} \\
 &= \frac{(101 - 81.73860182)^2}{81.73860182} + \frac{(48 - 64.07902736)^2}{64.07902736} + \frac{(17 - 20.18237082)^2}{20.18237082} + \\
 &\quad \frac{(41 - 45.30091185)^2}{45.30091185} + \frac{(43 - 35.51367781)^2}{35.51367781} + \frac{(8 - 11.18541033)^2}{11.18541033} + \\
 &\quad \frac{(20 - 34.96048632)^2}{34.96048632} + \frac{(36 - 27.40729483)^2}{27.40729483} + \frac{(15 - 8.632218845)^2}{8.632218845} \\
 &= 25.762
 \end{aligned}$$

Die χ^2_4 -Verteilung ordnet dem Intervall $[25.762, \infty[$ den Wert $3.53406E - 05$ zu ($= p$ -Wert). $3.53406E - 05 < 0.05$. Die Abweichung der faktischen Häufigkeiten von den erwarteten Häufigkeiten ist hochsignifikant.

Die Analyse der Residuen erlaubt es, das Ergebnis des eigentlich recht undifferenzierten Chi-Quadrat-Test genauer zu untersuchen (s. Tabelle 24.1.6):

		Veränderung der Arbeitssituation			Summen
		verbessert	unverändert	verschlechtert	
Realisierungsniveau	hoch	19.26139818	-16.07902736	-3.182370821	0
von	mittel	-4.300911854	7.486322188	-3.185410334	0
Guppenarbeiten	gering	-14.96048632	8.592705167	6.367781155	0
Summen		0	0	0	0

Tabelle 24.1.6: Unstandardisierte Residuen für die Tabelle 24.1.4

Bedeutende Abweichungen der faktischen von den erwarteten Häufigkeiten scheinen etwa in den Zellen (1,1), (1,2) und (3,1) vorzuliegen. Wir betrachten die standardisierten Pearson-Residuen

	verbessert	unverändert	verschlechtert
hoch	4.248328	-3.641902	-1.073786
mittel	-1.056746	1.888934	-1.197325
gering	-4.010373	2.365414	2.611344

Signifikante Abweichungen der faktischen von den erwarteten Häufigkeiten kommen also in den Zellen (1,1), (1,2), (3,1), (3,2) und (3,3) vor. $\diamond$

Bemerkung 24.1.3. (1) Die Daten der Kreuztabelle des Beispiel 24.1.2 sind in beide Dimensionen ordinalskaliert. Für solche Daten gibt es Tests, die mächtiger sind als der χ^2 -Test. Einen davon werden wir in Kürze behandeln (Testen von Gamma).

(2) Bei einem χ^2 -Test vergleichen wir die Häufigkeiten zweier Tabellen. Je mehr sich die jeweiligen absoluten Häufigkeiten pro Zelle unterscheiden, desto grösser wird der Testwert. Wir lehnen die Nullhypothese ab, wenn diese Unterschiede ein gewisses Mass überschreiten. Entsprechend testen wir jeweils rechtsseitig. Liegt linksseitig ein signifikantes Ergebnis vor, bedeutet dies eine Überanpassung der empirischen Daten an die erwarteten Häufigkeiten. Es könnte dann z.B. eine Fälschung der Daten vorliegen.

(3) Manchmal wird im 2×2 -Felder Fall noch eine sogenannte Kontinuitätskorrektur berechnet. Diese stellt eine Näherung an den exakten Test von Fisher dar. Die Korrektur erfolgt durch die Verwendung der folgenden Formel:

$$\sum_{i=1}^2 \sum_{j=1}^2 \frac{(|K_{i,j} - k_{i,j}^e| - 0.5)^2}{k_{i,j}^e} \text{ ist asymptotisch } \chi_1^2 - \text{verteilt.}$$

 $\diamond$

Am Beispiel:

$$\frac{(|12 - 10.58| - 0.5)^2}{10.58} + \frac{(|11 - 12.42| - 0.5)^2}{12.42} + \frac{(|34 - 35.42| - 0.5)^2}{35.42} + \frac{(|43 - 41.58| - 0.5)^2}{41.58} = 0.1924$$

$1 - F_{\chi_{(2-1)(2-1)}^2}(0.1925) = 0.660843984$. Die Kontinuitätskorrektur ist erst von Bedeutung, wenn der exakte Test von Fisher aus Kapazitätsgründen bei grossen Stichproben nicht mehr berechnet wird. Ob sie überhaupt nützlich ist, ist umstritten.

(4) Liegen die erwarteten Häufigkeiten in mehr als 20% der Zellen unter 5, so kann man Monte-Carlo-Verfahren anwenden oder exakte Tests, die auf Kombinatorik beruhen.

24.2 Fishers exakter Test

Ziel: Es soll überprüft werden, ob die Verteilung von Objekten in einer 2×2 Kreuztabelle zufällig ist oder ob es bedeutsame Abweichungen von einer zufälligen Verteilung gibt.

Voraussetzungen: (vor allem für nominalskalierte Daten verwendet). Die Zufallsvariablen sind identisch verteilt und unabhängig.

Der Vorteil des exakten Tests nach Fisher besteht darin, dass er auch mit kleinen Stichproben durchgeführt werden kann. Es wird von folgender Idee ausgegangen: Im Vierfelderfall ist bei fixierten Randverteilungen die Teststatistik, die den Testwert k_{11} annimmt, hypergeometrisch verteilt - wie wir ohne Beweis festhalten, d.h.

$$P(X = k_{11}) = \frac{\binom{k_{1+}}{k_{11}} \binom{n-k_{1+}}{k_{+1}-k_{11}}}{\binom{n}{k_{+1}}}$$

mit

		Variable 2		Summen
		Ausprägung 1	Ausprägung 2	
Variable 1	Ausprägung 1	k_{11}	k_{12}	k_{1+}
	Ausprägung 2	k_{21}	k_{22}	k_{2+}
Summen		k_{+1}	k_{+2}	$k_{++} = n$

Betrachten wir wiederum das Beispiel 24.1.1, um dies zu verstehen:

		Rauchverhalten		Summen
		Raucher	Nichtraucher	
Geschlecht	Männer	12	34	46
	Frauen	11	43	54
Summen		23	77	100

Dabei ist zu beachten, dass die Notation $X \sim H(n, k_{+1}, k_{1+})$ der früher verwendeten Notation $H(N, n, M)$ entspricht. Bei der Alternativhypothese, dass Objekte der Kategorie k_{11} übervertreten sind, wird ein rechtsseitiger Test durchgeführt, d.h. es wird unter der Nullhypothese, dass kein Zusammenhang zwischen den Variablen besteht, die Wahrscheinlichkeit berechnet, zufällig Tabellen zu erhalten, für die gilt $k'_{11} \geq k_{11}$:

$$\sum_{i=k_{11}}^{k_{+1}} \frac{\binom{k_{1+}}{i} \binom{n-k_{1+}}{k_{+1}-i}}{\binom{n}{k_{+1}}}$$

Bei der Alternativhypothese, dass die Kategorie der Häufigkeiten k_{11} untervertreten ist, wird

$$\sum_{i=0}^{k_{11}} \frac{\binom{k_{1+}}{i} \binom{n-k_{1+}}{k_{+1}-i}}{\binom{n}{k_{+1}}}$$

berechnet. Für die obige Kreuztabelle 24.1.1 haben wir eine ungerichtete Alternativhypothese: wir denken, dass sich das Rauchverhalten der Männer und Frauen unterscheidet. Es ist also ein zweiseitiger Test durchzuführen

Wir berechnen

$$P(X \geq 12) = \sum_{i=12}^{23} \frac{\binom{46}{i} \binom{100-46}{23-i}}{\binom{100}{23}} = 0.329\,75$$

$$P(X \leq 12) = \sum_{i=0}^{12} \frac{\binom{46}{i} \binom{100-46}{23-i}}{\binom{100}{23}} = 0.820\,05$$

mit $X \sim H(100, 46, 23)$. Für den zweiseitigen Test erhalten wir den p -Wert $2 \cdot \min\{0.329\,75, 0.820\,05\} = 0.659\,5 > 0.05$. Der Unterschied im Rauchverhalten der Männer und Frauen ist nicht signifikant (für die Berechnung mit Excel $=2 \cdot \min\{HYPGEOM.VERT(12; 23; 46; 100; 1), 1 - HYPGEOM.VERT(11; 23; 46; 100; 1)\}$)

Statt dieses zweiseitigen p -Wertes wird oft der folgende nicht-äquivalente p -Wert angegeben: sei q die Wahrscheinlichkeit, dass sich genau die vorliegende Tabelle bei gegebener Randverteilung zufälliger Weise ergibt. Dann wird die Eintreffenswahrscheinlichkeit q_i für jede der möglichen Tabellen i berechnet. Der zweiseitige p -Wert wird dann berechnet durch

$$p = \sum_{q_i \leq q} q_i. \quad (24.2.1)$$

Im obigen Beispiel ergibt sich, wie Sie selber nachrechnen können: p -Wert = 0.634297374.

Um die Interpretation von 2×2 -Tabellen zu erleichtern, wird oft das Quoten-Verhältnis (Odds-Ratio) θ samt Vertrauensintervallen angegeben. Die näherungsweise Standardabweichung von $\ln \theta$ ist

$$s(\ln \theta) = \sqrt{\frac{1}{k_{11}} + \frac{1}{k_{12}} + \frac{1}{k_{21}} + \frac{1}{k_{22}}}$$

und das asymptotische Vertrauensintervall

$$\exp \left(\ln \theta \pm \Phi^{-1} \left(\beta + \frac{1-\beta}{2} \right) s(\ln \theta) \right),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist und β die Breite des Vertrauensintervalls in Anteilen (Agresti, 2002, S. 71). Im obigen Beispiel ergibt sich folgendes asymptotisches 95%-Vertrauensintervall (d.h. $\beta = 0.95$) mit

$$\begin{aligned} \Phi^{-1} \left(\beta + \frac{1-\beta}{2} \right) &= \Phi^{-1} \left(0.95 + \frac{1-0.95}{2} \right) = \Phi^{-1}(0.975) \approx 1.96. \\ \hat{\theta} &= \frac{12.5 \cdot 43.5}{11.5 \cdot 34.5} = 1.370510397 \\ \ln \hat{\theta} &= 0.315183223 \\ s(\ln \theta) &= \sqrt{\frac{1}{12} + \frac{1}{11} + \frac{1}{34} + \frac{1}{43}} = 0.476350714 \end{aligned}$$

$$\text{Obere Grenze:} \quad \exp(0.315183223 + 1.96 \cdot 0.476350714) = 3.486204004$$

$$\text{Untere Grenze} \quad : \quad \exp(0.315183223 - 1.96 \cdot 0.476350714) = 0.538780503$$

Bemerkung 24.2.1. (1) Mit R ergeben sich die folgenden, oft etwas vorsichtigeren Resultate fürs Beispiel: „ p -value = 0.6343; 95 percent confidence interval: 0.4877661 3.9167187; sample estimates: odds ratio 1.375193“. Der p -Wert stimmt also mit der zweiten obigen Berechnungsmethode (24.2.1) überein. Die anderen Werte weichen etwas ab. In der Literatur (z.B. (Agresti, 2002, S. 91 - 101)) werden unterschiedliche Berechnungsverfahren angegeben, z.B. exakte Vertrauensintervalle für das Quoten-Verhältnis (das obige Vertrauensintervall gilt asymptotisch).

(2) Im Vierfelder-Fall ist der Fisher-Test dem χ^2 -Test vorzuziehen. Die Einschränkung, dass nicht mehr als 20% der Felder eine kleinere erwartete Häufigkeit als 5 haben müssen, braucht nicht beachtet zu werden. Es gibt auch für $n \times m$ -Kreuztabellen ($n, m > 2$) exakte Tests. Diese beruhen auf einer verallgemeinerten Hypergeometrischen Verteilung. Es ist allerdings mühsam, diese von Hand zu berechnen. Statistikprogramme bieten diese manchmal an, oft in Zusatzmodulen, die speziell gekauft werden müssen. R berechnet einen exakten Test auch für grössere Tabellen (`fisher.test`), allerdings nur, wenn die Häufigkeiten nicht zu gross sind. $\diamond$

24.3 Vergleich mit vorgegebener Häufigkeitsverteilung

χ^2 -Tests können ebenfalls verwendet werden, um die Häufigkeiten zweier Tabellen zu vergleichen, wenn die erwarteten Häufigkeiten nicht aus der Randverteilung der empirischen Kreuztabelle berechnet werden. Es handelt sich im Prinzip um den Anteilstest, den wir bei den Einstichprobentests bereits gesehen haben (Unterkapitel 20.8 auf der Seite 395). Die Freiheitsgrade werden in diesem Fall wie folgt berechnet

$$I \cdot J - 1,$$

mit der Anzahl I an Zeilen und der Anzahl J an Spalten - im Gegensatz zum vorangegangenen χ^2 -Unabhängigkeits-Test, bei dem die erwarteten Häufigkeiten aus den Randsummen berechnet wurden, wodurch man Freiheitsgrade verliert (Anzahl unabhängiger Zufallsvariablen). Wir betrachten ein Beispiel zum Anpassungstest:

Beispiel 24.3.1. Wir nehmen an, in der Schweiz gilt: 51% sind Frauen, 49% Männer, 24% der Frauen rauchen, 28% der Männer rauchen. Wir wollen überprüfen, ob im Oberwallis eine ähnliche Verteilung auf die Raucher vorliegt und ziehen eine Zufallsstichprobe von 100 Personen. Dabei würden wir folgendes Resultat erhalten (s. Tabelle 24.3.7):

		Rauchverhalten		Summen
		Raucher	Nichraucher	
Geschlecht	Männer	28	20	48
	Frauen	6	46	52
Summen		34	66	100

Tabelle 24.3.7: Beispielkreuztabelle für Berechnung eines bivariaten Anteilsvergleichs

H_0 : Die Anteile in der Stichprobe entsprechen den Anteilen in der CH-Bevölkerung.

H_A : Die Anteile in der Stichprobe entsprechen den Anteilen in der CH-Bevölkerung nicht.

Wir müssen nun eine entsprechende Tabelle mit absoluten Häufigkeiten für die Schweiz berechnen, indem wir die Häufigkeitsverteilung auf Stichprobengrösse umrechnen. Dabei gilt, dass der Anteil:

- der rauchenden Frauen in der Schweiz 24% von 51%: $0.24 \cdot 0.51 = 0.1224$
- der nicht-rauchenden Frauen in der Schweiz 76% von 51%: $0.76 \cdot 0.51 = 0.3876$
- der rauchenden Männer in der Schweiz 28% von 49%: $0.28 \cdot 0.49 = 0.1372$
- der nicht rauchenden Männer in der Schweiz 72% von 49%: $0.72 \cdot 0.49 = 0.3528$

Indem man diese relativen Häufigkeiten mit der Stichprobengrösse 100 multipliziert, erhält man die folgende Tabelle 24.3.8 der erwarteten Häufigkeiten:

		Rauchverhalten		Summen
		Raucher	Nichraucher	
Geschlecht	Männer	13.72	35.28	49
	Frauen	12.24	38.76	51
Summen		25.96	74.04	100

Tabelle 24.3.8: Erwartete Häufigkeiten - auf Stichprobengrösse umgerechnete Häufigkeiten der Grundgesamtheit

Beachten Sie, dass die Randsummen nicht mit den Randsummen der Stichprobe übereinstimmen. Im Test gehen wir gleich vor wie bei einem univariaten Anteilsvergleich, d.h. die Freiheitsgrade sind die Anzahl der zu vergleichenden Zellen minus 1.

Wir erhalten den Testwert:

$$\frac{(28 - 13.72)^2}{13.72} + \frac{(20 - 35.28)^2}{35.28} + \frac{(6 - 12.24)^2}{12.24} + \frac{(46 - 38.76)^2}{38.76} = 26.014$$

$F_{\chi^2_3}(26.014) = \text{chiqu.vert}(26.014; 3; 1) = 0.999990527$. Der p -Wert ist somit: $1 - 0.999990527 = 9.47325 \cdot 10^{-6} < 0.05$. Die Häufigkeiten unterscheiden sich signifikant. $\diamond$

Bemerkung 24.3.2. Untersuchungen sollten jeweils durch Hypothesen, die vor der Untersuchung schriftlich niedergelegt werden, begleitet werden. Testen wir bei 15 Variablen jede gegen jede, erhalten wir $\binom{15}{2} = 105$ Tests (bei Fragebogen ist es ein leichtes, mehr als 15 Variablen zu gewinnen). Dabei erhalten wir mit einer hohen Wahrscheinlichkeit ein paar signifikante Ergebnisse (Fehler erster Ordnung! Bei einem Signifikanzniveau von 5% erhalten wir durchschnittlich 5 signifikante Ergebnisse auf 100, bei einem Signifikanzniveau von 1% durchschnittlich ein signifikantes Ergebnis auf 100). Es ist deshalb sinnlos, solche Ergebnisse, die man ohne vorgängige Hypothese erhalten hat, verwenden zu wollen: man wird mit hoher Wahrscheinlichkeit Opfer des Fehlers erster Ordnung. Dieser Fehler wird in Marketingstudien oder früher sogar in medizinischen Studien oft gemacht. Entdeckt man derart signifikante Ergebnisse, können diese jedoch für eine neue Erhebung als Hypothesen dienen. Die erste Erhebung dient dann der sogenannten „explorativen“ Phase, während die zweite Erhebung die Ergebnisse der explorativen Phase testet. Erst wenn in einer zweiten Studie die Hypothesen bestätigt werden können, darf man den Resultaten der ersten Studie Glauben schenken. $\diamond$

24.3.1 Übungen

1. Erstellen Sie die Tabelle der erwarteten Häufigkeiten und der Residuen für die folgenden Kreuztabellen 24.3.9 - 24.3.10:

		Variable 1		Summen
		Ausprägung 1	Ausprägung 2	
Variable 2	Ausprägung 1	20	30	
	Ausprägung 2	10	80	
Summen				

Tabelle 24.3.9: Beispielkreuztabelle für Berechnung der erwarteten Häufigkeiten

Variable 2	Variable 1				Summen
	Auspr. 1	Auspr. 2	Auspr. 3	Auspr. 4	
Ausprägung 1	10	50	30	8	
Ausprägung 2	11	22	40	20	
Ausprägung 3	14	7	8	40	
Ausprägung 4	4	8	20	60	
Summen					

Tabelle 24.3.10: Beispielkreuztabelle für Berechnung der erwarteten Häufigkeiten

2. Ein Marktforschungsinstitut überprüft mit Hilfe einer Umfrage die Hypothese, ob Frauen tatsächlich weniger als Männer im Internet surfen. Es wird eine Zufallsstichprobe von 80 Personen gezogen. Dabei ergeben sich die folgenden bivariaten Häufigkeiten (s. Tabelle 24.3.11):

		Internetnutzung		Summen
		nutzt	nutzt nicht	
Geschlecht	Männer	12	26	
	Frauen	5	37	
Summen				

Tabelle 24.3.11: Beispielkreuztabelle für Berechnung eines Unabhängigkeitstest in Kreuztabellen

Testen Sie von Hand, mit Excel und einer Statistiksoftware ($\alpha = 0.05$)

3. Eine Marktforschungsfirma macht im Auftrag einer KMU eine Umfrage bei Konsumenten und kommt zu folgenden Resultaten: siehe Daten chiquadratbivariat.xls. Lassen sich Gruppen finden, die sich bezüglich ihrer Einstellung zum Produkt signifikant von anderen unterscheiden. Analysieren Sie unter Berücksichtigung der Bemerkung 24.3.2. ($\alpha = 0.05$)
4. Die Einstellung bezüglich einer politischen Fragestellung werden untersucht. Dabei ergibt sich folgendes Bild (Daten: chiquadratbivariat.xls). Unterscheiden sich die Einstellungen der Gruppen? Es wird vermutet, dass sich altersspezifische Unterschiede vorliegen ($\alpha = 0.05$).
5. Es wird eine Stichprobe gezogen (Resultate siehe chiquadratbivariat.xls). In der untersuchten Population kommen folgende Anteile vor (s. Tabelle 24.3.12):

		Alter			
		≤ 30	$]30 - 50]$	$]50, 100]$	
Wohnort	Stadt	0.05	0.15	0.2	0.4
	Aglo	0.2	0.03	0.235	0.465
	Land	0.1	0.025	0.01	0.135
Summen		0.35	0.205	0.445	1

Tabelle 24.3.12: Tabelle der Anteile an der Grundgesamtheit - gemeinsame relative Häufigkeitsverteilung

Kann die Stichprobe als Zufallsstichprobe aus der untersuchten Population betrachtet? ($\alpha = 0.05$)

24.3.2 Lösungen

1. Wir berechnen die Spalten und Zeilensummen (Randsummen).

Dann berechnen wir $k_{ij}^e = \frac{k_{i+} \cdot k_{+j}}{n}$

$$k_{1,1}^e = \frac{50 \cdot 30}{140} = 10.714$$

$$k_{2,1}^e = \frac{90 \cdot 30}{140} = 19.286$$

$$k_{1,2}^e = \frac{50 \cdot 110}{140} = 39.286$$

$$k_{2,2}^e = \frac{90 \cdot 110}{140} = 70.714$$

Zudem berechnen wir noch die Residuen. Wir erhalten also die zusammenfassende Tabelle 24.3.13: und die standardisierten Pearson-Residuen, die man natürlich auch in der obigen Tabelle angeben könnte:

$$\begin{array}{cc} 3.991574 & -3.991574 \\ -3.991574 & 3.991574 \end{array}$$

			Variable 1		Summen
			Ausprägung 1	Ausprägung 2	
Variable 2	Ausprägung 1	$k_{1,j}$	20	30	50
		$k_{1,j}^e$	10.71428571	39.28571429	50
		Residuum	9.285714286	-9.285714286	0
	Ausprägung 2	$k_{2,j}$	10	80	90
		$k_{2,j}^e$	19.28571429	70.71428571	90
		Residuum	-9.285714286	9.285714286	0
	Summen	$k_{+,j}$	30	110	140
		$k_{+,j}^e$	30	110	

Tabelle 24.3.13: Häufigkeitstabelle mit erwarteten Häufigkeiten und Residuen

Für die zweite Tabelle verfahren wir analog und erhalten die Tabelle [24.3.14](#):

Variable 2		Variable 1				Summen
		A 1	A 2	A 3	A 4	
A 1	$k_{1,j}$	10	50	30	8	98
	$k_{1,j}^e$	10.858	24.222	27.284	35.636	98
	Residuum	-0.858	25.778	2.716	-27.636	0
A 2	$k_{2,j}$	11	22	40	20	93
	$k_{2,j}^e$	10.304	22.986	25.892	33.818	93
	Residuum	0.696	-0.986	14.108	-13.818	0
A 3	$k_{3,j}$	14	7	8	40	69
	$k_{3,j}^e$	7.645	17.054	19.210	25.091	69
	Residuum	6.355	-10.054	-11.210	14.909	0
A 4	$k_{4,j}$	4	8	20	60	92
	$k_{4,j}^e$	10.193	22.739	25.614	33.455	92
	Residuum	-6.193	-14.739	-5.614	26.545	0
Summen		$k_{+,j}$	39	87	98	128
		$k_{+,j}^e$	39	87	98	128

Tabelle 24.3.14: Häufigkeitstabelle mit erwarteten Häufigkeiten und Residuen

und die standardisierten Pearson-Residuen:

-0.3250453	7.1065309	0.7205588	-6.831814
0.2680659	-0.2762656	3.8050187	-3.472518
2.7184108	-3.1293315	-3.3580005	4.161198
-2.3935514	-4.1448436	-1.5193149	6.694137

2. Da eine 2×2 -Tabelle vorliegt, verwenden wir den Fisher-Test:

H_0 : Frauen und Männer unterscheiden sich nicht.

H_A : Das Verhalten der Frauen unterscheidet sich von dem der Männer. ($\implies$ zweiseitiger Test; $\alpha = 0.05$)

$$P(X \geq 12) = \sum_{i=12}^{17} \frac{\binom{38}{i} \binom{80-38}{17-i}}{\binom{80}{17}} = 0.0299$$

$$P(X \leq 12) = \sum_{i=0}^{12} \frac{\binom{38}{i} \binom{80-38}{17-i}}{\binom{80}{17}} = 0.99279$$

p-Wert: $2 \cdot \min\{0.0299, 0.99279\} = 0.0598 > 0.05$ (knapp nicht signifikant).

Excel: $2 \cdot \min\{\text{hypgeom.vert}(12; 38; 17; 80; 1), 1 - \text{hypgeom.vert}(11; 38; 17; 80; 1)\}$

Wir berechnen übungshalber den Chi-Quadrat-Test. Die erwarteten Zellenhäufigkeiten sind (s. Tabelle 24.3.15):

			Internetnutzung		Summen
			Internutzer	Nicht Internetnutzer	
Geschlecht	Männer	$k_{1,j}$	12	26	38
		$k_{1,j}^e$	8.075	29.925	38
		Residuen	3.925	-3.925	0
	Frauen	$k_{2,j}$	5	37	42
		$k_{2,j}^e$	8.925	33.075	42
		Residuen	-3.925	3.925	0
Summen	$k_{+,j}$	17	63	80	
	$k_{+,j}^e$	17	63	80	

Tabelle 24.3.15: Faktische und erwartete Häufigkeiten sowie Residuen für die Variablen Einstellung und Geschlecht

Wir erhalten als Wert der Teststatistik (Testwert):

$$\frac{(12 - 8.075)^2}{8.075} + \frac{(26 - 29.925)^2}{29.925} + \frac{(5 - 8.925)^2}{8.925} + \frac{(37 - 33.075)^2}{33.075} = 4.6145$$

Wir berechnen den Wert der Verteilungsfunktion: $F_{\chi^2_1}(4.6145) = \text{chiqu.vert}(4.6145; 1; 1) = 0.968297264$. Der p -Wert ist somit $1 - 0.968297264 = 0.03170274 < 0.05$. Der Unterschied ist auf dem festgelegten Niveau bei diesem Test signifikant - wir würden allerdings den Fisher-Test verwenden. Bei der Interpretation mit Hilfe der Residuen stellen wir fest, dass im Falle einer 2×2 -Kreuztabelle die Abweichungen immer symmetrisch sind. Die standardisierten Pearson-Residuen sind:

2.148144	-2.148144
-2.148144	2.148144

Mit Statistikprogrammen ausser bei R müssen wir die Kreuztabelle zurück in eine Urliste von Daten verwandeln. Diese Vorbereitung machen wir am besten mit Hilfe von Excel oder mit einem Software-Routine (s. `math.logik.ch`, Statistik, `Kreuztabelle_in_Urliste.zip`).

- Da von der Firma keine Hypothesen aufgestellt wurden, sind die erhobenen Daten nur als Hypothesenlieferant für eine spätere Testphase zu verwenden. Da wir 4 Variablen haben, sind im Prinzip $\binom{4}{2} = 6$ bivariate Tests durchführbar. Da es um die Beeinflussung der Einstellung zum Produkt geht führen wir jedoch nur 3 Tests durch:

a) Da eine 2×2 -Kreuztabelle vorliegt, berechnen wir den Fisher-Test (zweiseitiger Test; $\alpha = 0.05$):

$$P(X \leq 10) = \sum_{i=0}^{10} \frac{\binom{18}{i} \binom{80-18}{43-i}}{\binom{80}{43}} = 0.66958$$

$$P(X \geq 10) = \sum_{i=10}^{18} \frac{\binom{18}{i} \binom{80-18}{43-i}}{\binom{80}{43}} = 0.53881$$

p -Wert: $2 \cdot \min\{0.66958, 0.53881\} = 1.0776$ wird aus W-theoretischen Gründen auf $1 > 0.05$ gesetzt. Keine signifikante Abweichung.

Wir berechnen noch den entsprechenden Chi-Quadrat-Test:

H_0 : Die Einstellung zum Produkt ist von der Variable Geschlecht unabhängig.

H_A : Die Einstellung zum Produkt ist von der Variable Geschlecht abhängig (d.h. es gibt Felder, wo die faktischen stark von den erwarteten Häufigkeiten abweichen; rechtsseitiger Test; $\alpha = 0.05$).

Wir erhalten (s. Tabelle 24.3.16):

			Geschlecht		Summen
			Männer	Frauen	
Einstellung	positiv	Anzahl	10	8	18
		Erwartete Anzahl	9.675	8.325	18
		Residuen	0.325	-0.325	0
	negativ	Anzahl	33	29	62
		Erwartete Anzahl	33.325	28.675	62
		Residuen	-0.325	0.325	0
Summe	Anzahl	43	37	80	
	Erwartete Anzahl	43	37	80	

Tabelle 24.3.16: Faktische und erwartete Häufigkeiten sowie Residuen für die Variablen Einstellung und Geschlecht

Die standardisierten Pearson-Residuen sind

0.1745224	-0.1745224
-0.1745224	0.1745224

b) H_0 : Die Einstellung zum Produkt ist von der Variable Alter unabhängig.

H_A : Die Einstellung zum Produkt ist von der Variable Alter abhängig (d.h. es gibt Felder, wo die faktischen stark von den erwarteten Häufigkeiten abweichen).

Wir erhalten (s. Tabelle 24.3.17):

Einstellung		Alter					Summen
		[0, 20]	]20, 35]	]35, 50]	]51, 65]	]66, 100]	
positiv	Anzahl	2	4	4	6	2	18
	$k_{1,j}^e$	2.475	3.6	4.275	3.375	4.275	18
	Residuen	-0.475	0.4	-0.275	2.625	-2.275	
negativ	Anzahl	9	12	15	9	17	62
	$k_{2,j}^e$	8.525	12.4	14.725	11.625	14.725	62
	Residuen	0.475	-0.4	0.275	-2.625	2.275	
Summen	Anzahl	11	16	19	15	19	80
	$k_{+,j}^e$	11	16	19	15	19	80

Tabelle 24.3.17: Faktische und erwartete Häufigkeiten sowie Residuen für die Variablen Einstellung und Alter

Testwert: 4.394; $df = 4$; p -Wert $0.355 > 0.05$. Die standardisierten Pearson-Residuen sind

-0.3692969	0.2677398	-0.173019	1.800652	-1.431339
0.3692969	-0.2677398	0.173019	-1.800652	1.431339

Die Hälfte der Zellen weisen einen Erwartungswert kleiner als 5 auf. Entsprechend dürfte der Test so nicht angewandt werden. Das Ergebnis ist zudem nicht signifikant, obwohl vermutlich die Altersgruppe 66- besonders negativ eingestellt ist. Wir können der Sache auf den Grund gehen, indem wir die übrigen Altersgruppen zusammenfassen und gegen die Altersgruppe 66- testen. Dazu schaffen wir eine neue Variable. Wir erhalten 24.3.18:

		Alter reklassiert		Summen
		[0, 60]	]60, 100]	
Einstellung	positiv	16	2	18
	negativ	45	17	62
Summen		61	19	80

Tabelle 24.3.18: Kreuztabelle mittels Reklassierung von Daten der Tabelle 24.3.17

und verwenden den Fisher-Test (zweiseitig; $\alpha = 0.05$):

$$P(X \leq 16) = \sum_{i=0}^{16} \frac{\binom{18}{i} \binom{80-18}{61-i}}{\binom{80}{61}} = 0.96759$$

$$P(X \geq 16) = \sum_{i=16}^{18} \frac{\binom{18}{i} \binom{80-18}{61-i}}{\binom{80}{61}} = 0.13004$$

p-Wert: $2 \cdot \min\{0.96749, 0.13004\} = 0.26008$

Wir berechnen noch den entsprechenden Chi-Quadrat-Test: Wir erhalten (s. Tabelle 24.3.19):

			Alter reklassiert		
			[0, 60]	]60, 100]	
Einstellung	positiv	Anzahl	16	2	18
		Erwartete Anzahl	13.725	4.275	18
		Residuen	2.275	-2.275	
	negativ	Anzahl	45	17	62
		Erwartete Anzahl	47.275	14.725	62
		Residuen	-2.275	2.275	
Gesamt	Anzahl	61	19	80	
	Erwartete Anzahl	61	19	80	

Tabelle 24.3.19: Faktische und erwartete Häufigkeiten sowie Residuen für die Variablen Alter reklassiert und Geschlecht

mit den standardisierten Pearson-Residuen:

1.431339	-1.431339
-1.431339	1.431339

Testwert: 2.049; df=1; p-Wert: 0.152 > 0.05. Nicht signifikante Abweichung. Entgegen unserer Vermutung liegt kein signifikanter Zusammenhang vor. Der p-Wert ist jedoch kleiner geworden. Nur mehr eine Zelle (25%) hat eine kleinere erwartete Häufigkeit als 5. Allerdings liegt der Wert nicht stark unter 5. Wir würden in der Praxis den Test trotzdem anwenden.

- c) H_0 : Die Einstellung zum Produkt ist von der Variable Wohnort unabhängig.
 H_A : Die Einstellung zum Produkt ist von der Variable Wohnort (d.h. es gibt Felder, wo die faktischen stark von den erwarteten Häufigkeiten abweichen).
 Wir erhalten (s. Tabelle 24.3.20):

			Wohnort			Summen
			Land	Stadt	Aglo	
Einstellung	positiv	Anzahl	4	8	6	18
		Erwartete Anzahl	5.175	6.975	5.85	18
		Residuen	-1.175	1.025	0.15	
	negativ	Anzahl	19	23	20	62
		Erwartete Anzahl	17.825	24.025	20.15	62
		Residuen	1.175	-1.025	-0.15	
Summen	Anzahl	23	31	26	80	
	Erwartete Anzahl	23	31	26	80	

Tabelle 24.3.20: Faktische und erwartete Häufigkeiten sowie Residuen für die Variablen Wohnort und Geschlecht

und den standardisierten Pearson-Residuen

-0.6950874	0.5633105	0.08574535
0.6950874	-0.5633105	-0.08574535

Es ergeben sich keine signifikante Zusammenhänge zwischen der Einstellung zum Produkt und den Variablen Alter, Wohnort und Geschlecht (und wenn wir welche gefunden hätten, würde dies ohne vorgängige, entsprechende Hypothesen noch wenig bedeuten!).

4. H_0 : Die faktischen Häufigkeiten weichen von den erwarteten nicht ab.
 H_1 : Die faktischen Häufigkeiten weichen insgesamt von den erwarteten ab. ($\alpha = 0.05$).
 Wir erhalten: (s. Tabelle 24.3.21):

Einstellung		Alter				Gesamt
		[0, 30]	]30, 40]	]40, 60]	]60, 100]	
dafür	Anzahl	50	40	60	50	200
	$k_{1,j}^e$	47.917	54.167	50	47.9167	200
	Residuen	2.083	-14.167	10	2.083	
eher dafür	Anzahl	30	30	40	40	140
	$k_{2,j}^e$	33.542	37.917	35	33.542	140
	Residuen	-3.542	-7.917	5	6.458	
neutral	Anzahl	20	30	10	5	65
	$k_{3,j}^e$	15.573	17.604	16.25	15.573	65
	Residuen	4.427	12.396	-6.25	-10.573	
eher dagegen	Anzahl	10	20	5	10	45
	$k_{4,j}^e$	10.781	12.188	11.25	10.781	45
	Residuen	-0.781	7.813	-6.25	-0.781	
dagegen	Anzahl	5	10	5	10	30
	$k_{5,j}^e$	7.188	8.125	7.5	7.188	30
	Residuen	-2.188	1.875	-2.5	2.813	
Gesamt	Anzahl	115	130	120	115	480
	$k_{+,j}^e$	115	130	120	115	480

Tabelle 24.3.21: Faktische und erwartete Häufigkeiten sowie Residuen für die Variablen Alter und Einstellung

mit den standardisierten Pearson-Residuen

0.2452762	-2.860388	1.9333475	0.7476672
-1.0004127	-1.688604	0.9856975	1.765513
1.2827887	3.841019	-2.0325202	-3.223340
-0.3656362	2.842676	-2.3416772	-0.1857594

Testwert: 41.066; df=12; p-Wert: 0.00004777558001918 < 0.05 Hochsignifikante Abweichungen mancher faktischer Häufigkeiten (von den erwarteten). Mit Hilfe der Residuen können wir überprüfen, wo Abweichungen vorliegen und in welche Richtung diese gehen. Beide Dimensionen sind ordinalskaliert. Somit braucht der Test nicht alle verfügbaren Informationen und man würde eher Gamma verwenden.

5. Die Lösung finden Sie unter `chiquadratbivariat.xls`.

24.4 Lernziele

- Situationen erkennen können, in denen der Fisher-Test und der χ^2 -Test angewandt werden kann.
- den Fisher- und χ^2 -Test durchführen können (Bedingungen überprüfen; Übungen von der obigen Art durchführen können).
- die häufigsten Fehler beim χ^2 -Test kennen und vermeiden können.
- Bemerkung 24.3.2 kennen und beachten!

Kapitel 25

Gamma

Ziel: Testen eines gerichteten Zusammenhangs in Kreuztabellen mit ordinalskalierten Variablen

Voraussetzungen: grosse Stichproben; Daten können als Werte unabhängiger und identisch verteilter Zufallsvariablen betrachtet werden.

(i) beide Variablen sind ordinalskaliert oder

(ii) eine der Variablen ist ordinalskaliert während die andere eine nominalskalierte Variable mit nur zwei Ausprägungen ist (dichotom skalierte Variable)

Zur Erinnerung: Gamma ist eine Kennzahl, welche den Zusammenhang von zwei ordinalskalierten Variablen ausdrückt. Zur Berechnung von Gamma siehe letztes Semester (bivariate beschreibende Statistik).

Bei positivem Gamma liegen die Daten konzentriert auf der Hauptdiagonalen (von oben links nach unten rechts) der Kreuztabelle. Bei negativem Gamma liegen die Daten konzentriert auf der Nebendiagonalen (von oben rechts nach unten links). Man sollte Gamma nur brauchen, wenn die Daten vorwiegend auf der Haupt- oder Nebendiagonalen liegen (also nicht bei Dreiecksformen, U-Formen oder Konzentrationen auf dem Rand einer Tabelle).

Es ergeben sich folgende mögliche Hypothesengruppen:

H_0 : Gamma ist nicht von 0 verschieden.

H_A : Gamma ist von 0 verschieden (zweiseitiger Test).

H_0 : Gamma ist kleiner-gleich 0.

H_A : Gamma ist positiv (rechtsseitiger Test)

H_0 : Gamma ist grösser-gleich 0.

H_A : Gamma ist negativ (linksseitiger Test)

Die standardisierte Verteilung von Gamma ist für genügend grosse Stichproben näherungsweise standardnormalverteilt, d.h. es gilt:

$\frac{\Gamma}{\frac{s}{\sqrt{n}}} \sim N(0, 1)$ (asymptotisch; für n genügend gross)
S^2 ist die Zufallsvariable, die den Wert $s^2 = \frac{16}{(\pi_c + \pi_d)^4} \cdot \sum_{i=1}^n \sum_{j=1}^m h_{ij} (\pi_d h_{ij}^c - \pi_c h_{ij}^d)^2$ annimmt:
$h_{ij} = \frac{k_{ij}}{n}$
k_{ij} = absolute Häufigkeit der Zelle i, j
n = Stichprobengrösse
$h_{ij}^c = \sum_{a < i} \sum_{b < j} h_{ab} + \sum_{a > i} \sum_{b > j} h_{ab}$
$h_{ij}^d = \sum_{a < i} \sum_{b > j} h_{ab} + \sum_{a > i} \sum_{b < j} h_{ab}$
$\pi_c = \sum_{i=1}^n \sum_{j=1}^m h_{ij} h_{ij}^c$
$\pi_d = \sum_{i=1}^n \sum_{j=1}^m h_{ij} h_{ij}^d$
Γ = Zufallsvariable, die konkrete γ als Werte annimmt.

Die Grenzen der β -Vertrauensintervalle für Gamma erhält man mit den eingeführten Methoden durch die Auflösung von

$$\frac{1 - \beta}{2} < \Phi\left(\frac{\sqrt{n}\Gamma}{S}\right) < 1 - \frac{1 - \beta}{2}$$

Man erhält die folgenden Grenzen γ_{g_i} des Vertrauensintervalls:

$$\Gamma_{g_i} = \Gamma \pm \Phi^{-1}\left(\beta + \frac{1 - \beta}{2}\right) \frac{S}{\sqrt{n}}$$

Beispiel 25.0.1. In einer Studie wird der Zusammenhang zwischen der Zufriedenheit mit der Arbeit und dem Einkommen analysiert. Es wird vermutet, dass gilt: Je mehr Einkommen, desto zufriedener mit der Arbeit. Aus den Daten wird die folgende Kreuztabelle gewonnen (s. Tabelle 25.0.1):

Einkommen	Zufriedenheit			
	sehr unzufrieden	etwas unzufrieden	ziemlich zufrieden	sehr zufrieden
[0, 6000]	20	24	80	82
]6000, 15'000]	22	38	104	125
]15'000, 25'000]	13	28	81	113
]25'000, ..[	7	18	54	92

Tabelle 25.0.1: Beispielkreuztabelle für das Testen von Gamma

Man erhält: $n = 901$

konkordante Paare: 109520

diskordante Paare: 84915

Gamma: 0.126546147 (Gamma weicht nicht stark von 0 ab!!!)

$$\frac{s}{\sqrt{n}} = 0.040801788$$

95%-Vertrauensintervall: die Hälfte des Intervalls ist $1.96 \cdot 0.040801788 = 0.079972$

Obere Grenze: $0.126546147 + 0.079972 = 0.206518$

Untere Grenze: $0.126546147 - 0.079972 = 0.046574$

Rechtsseitiger Test:

$$\text{Testwert: } \frac{0.126546147}{0.040801788} = 3.101485289$$

p -Wert (einseitig): $1 - \Phi(3.101485289) = 0.000962762 < 0.05$

Signifikante Abweichung Gammas von 0 - in die von der Vermutung angegebenen Richtung (allerdings kleine Abweichung - signifikant, weil grosse Stichprobe). $\diamond$

Bemerkung 25.0.2. Führt man auf die obige Kreuztabelle einen Chi-Quadrat-Test aus, erhält man einen Testwert von 11.989. Dies ergibt bei $df = 9$ einen p -Wert von $0.214 > 0.05$. Für Kreuztabellen mit einer Massierung der Häufigkeiten auf der Haupt- oder Nebendiagonalen ist der Gammatest bedeutend mächtiger als der entsprechende Chi-Quadrat-Test. Liegt eine solche Massierung nicht vor - entsprechend ist die Berechnung von Gamma nicht angemessen - so kann der Chi-Quadrat-Test mächtiger sein. $\diamond$

Bemerkung 25.0.3. Liegt eine $2 \times m$ -Kreuztabelle vor, kann man Gamma dazu verwenden, um eine Verschiebung der Häufigkeitsverteilung der beiden Zeilen gegeneinander zu messen. Statt einer Korrelation wird ein Unterschied zwischen zwei ordinalen Häufigkeitsverteilungen angegeben. Bei einer Kreuztabelle „Geschlecht“ mal „Einstellung zu einer Feriendestination“ (unzufrieden, mässig zufrieden, zufrieden, sehr zufrieden) drückt Gamma aus, inwiefern sich die Geschlechter unterscheiden. Liegen die Häufigkeiten der Frauen tendenzmässig eher rechts oder eher links der Häufigkeiten der Männer?). Entsprechend kann man Gamma verwenden, um mehrere ordinalskalierte Variablen mit identischen Ausprägungen miteinander zu vergleichen: man erstellt für jede der Variablen eine univariate Häufigkeitsverteilung und vergleicht dann paarweise alle Häufigkeitsverteilungen mit der Häufigkeitsverteilung einer als Referenz gewählten Variable. Man kann Graphiken mit den Gammas und deren Vertrauensintervallen erstellen. Das Verfahren kann eher fragwürdige Mittelwertsvergleiche von ordinalskalierte Variablen ersetzen - s. GammaVergleich-Befehl im R-Paket varia. $\diamond$

25.0.1 Übungen

1. Berechnen Sie eine geeignete Kennzahl für die folgende Kreuztabelle (s. Tabelle 25.0.2 und testen Sie die Hypothese, ob Raucher schlechtere Zähne haben als Nichtraucher ($\alpha = 0.05$; die Daten finden Sie auf <http://math.logik.ch> unter GammaUebungen.xls):

Zustand der Zähne	Rauchverhalten		
	raucht nicht	raucht mässig	raucht viel
gut	120	50	20
mässig	80	60	30
schlecht	30	65	50

Tabelle 25.0.2: Beispielkreuztabelle für das Testen von Gamma

2. Im Mittel- und Oberwallis wird eine Umfrage gemacht - Zufriedenheit mit einer spezifischen Staatsleistung (s. Tabelle 25.0.3):

Gegend	Zufriedenheit				
	unzufrieden	knapp zufrieden	mässig zufrieden	zufrieden	sehr zufrieden
Mittelwallis	11	16	23	30	18
Oberwallis	15	20	25	15	8

Tabelle 25.0.3: Beispielkreuztabelle für das Testen von Gamma

Berechnen Sie eine geeignete Kennzahl für diese Daten und testen Sie - die Daten finden Sie unter GammaUebungen.xls! ($\alpha = 0.05$)

3. Es wird die Frage „Sind Sie mit den Informationen des Tourismusbüros zufrieden“ mit den Antwortmöglichkeiten „unzufrieden“, „mässig zufrieden“, „zufrieden“ und „sehr zufrieden“ gestellt, um entsprechende Daten zu sammeln. Zudem wurde die Variable „Geschlecht“ erhoben. Es ergab sich die folgende Kreuztabelle

	unzufrieden	mässig zufrieden	zufrieden	sehr zufrieden
Männer	14	10	5	3
Frauen	2	3	11	13

Testen Sie, ob es einen geschlechtsspezifischen Unterschied zwischen den Geschlechtern gibt.

4. Man möchte testen, ob sich die Zufriedenheit mit dem Tourismusbüro von der Einschätzung der Freundlichkeit des Personals des Büros unterscheidet. Die Häufigkeitsverteilung der Variable Zufriedenheit mit dem Tourismusbüro ergab (16, 13, 16, 16) (s. Aufgabe 3). Die Häufigkeitsverteilung der Variable Einschätzung der Freundlichkeit des Personals mit denselben Kategorien ergab (10, 12, 15, 24). Gibt es einen Unterschied in den Beurteilungen?

25.0.2 Lösungen

1. Da beide Variablen ordinal skaliert sind, können wir Gamma berechnen:

$$\Pi_c = 120(60 + 30 + 65 + 50) + 50(30 + 50) + 80(65 + 50) + 60(50) = 40800$$

$$\Pi_d = 50(80 + 30) + 20(80 + 60 + 30 + 65) + 60(30) + 30(30 + 65) = 14850$$

$$\gamma = \frac{40800 - 14850}{40800 + 14850} = 0.46631$$

Die Vermutung ist: je mehr man raucht, desto schlechter sind die Zähne. Dies führt auf folgende Hypothesen:

H_0 : Es ist nicht der Fall, dass die Zähne um so schlechter sind, je mehr man raucht ($\gamma \leq 0$).

H_A : Wer mehr raucht, hat schlechtere Zähne ($\gamma > 0$) ($\alpha = 0.05$).

Die Übersetzung der Hypothesen in Aussagen über γ ergibt sich unter Berücksichtigung der Tabelle. Die Alternativhypothese trifft zu, wenn die Daten auf der Hauptdiagonale massiert sind.

$$\text{Test: } \frac{s}{\sqrt{n}} = 0.050288985.$$

$$\text{Testwert: } \frac{0.46631}{0.050288985} = 9.2726$$

$$\text{p-Wert: } 1 - \Phi(9.2726) \approx 0 < 0.05.$$

(hoch signifikanter proportionaler Zusammenhang zwischen Rauchverhalten und Zustand der Zähne).

2. Da eine der beiden Variablen ordinalskaliert ist und die andere nur zwei Ausprägungen aufweist (dichotome Variable), können wir Gamma verwenden. Da wir keine spezielle Richtung bezüglich Zufriedenheit vermuten, testen wir zweiseitig:

H_0 : Die Zufriedenheit unterscheidet sich nicht ($\gamma = 0$).

H_A : Die Zufriedenheit unterscheidet sich ($\gamma \neq 0$) ($\alpha = 0.05$).

$$\text{Gamma: } -0.294753086$$

$$\frac{s}{\sqrt{n}} = 0.099778505$$

$$\text{Testwert: } \frac{-0.294753086}{0.099778505} = -2.9541$$

$$\text{p-Wert (zweiseitig): } 2 \cdot \min\{1 - \Phi(-2.9541), \Phi(-2.9541)\} = 0.003136225 < 0.05$$

3. mit

```
geschl=matrix(c(14, 10, 5, 3, 2, 3, 11, 13), ncol=4, byrow=T)
```

```
geschl=as.table(geschl); library(varia); gamma.test(geschl)
```

```
erhält man: p-value = 5.618e-14
```

```
alternative hypothesis: two.sided; 95 percent confidence interval: 0.5620550 0.9585635
```

Gamma: 0.7603093. Das grosse positive Gamma gibt an, dass die Häufigkeiten auf der Hauptdiagonalen liegen und damit die Männer signifikant unzufriedener sind als die Frauen.


```
4. zufr=matrix(c(16,13,16,16,10,12,15,24), ncol=4,byrow=T); zufr=as.table(zufr)
   library(varia); gamma.test(zufr)
   ergibt: p-value = 0.08083; alternative hypothesis: two.sided
   95 percent confidence interval: -0.02790651 0.48313844
   sample estimates: Gamma: 0.227616.
```

Zwar ist die Häufigkeiten der zweiten Zeile gegenüber der ersten Zeile nach rechts verschoben, die Verschiebung ist aber nicht signifikant. (N.B. mit dem `GammaVergleich`-Befehl im `Varia`-Paket ist es nicht nötig, vorgängig univariate Häufigkeitesverteilung zu berechnen. Man kann die Tests unmittelbar auf Variablen der Urliste der Daten (dataframes) durchführen. Für Details s. Hilfe zum Befehl).

25.1 Lernziele

- Situationen erkennen kennen, für die der Gamma-Test angewendet werden kann.
- Gamma berechnen können (von Hand, mit einer Statistiksoftware und mit Excel - die Berechnung des Testwertes wird nicht verlangt)
- Mit Hilfe von Gamma und dessen Standardfehler den Gamma-Test durchführen können. Vertrauensintervalle von Gamma berechnen können.

Kapitel 26

Varianzanalyse und Kruskal-Wallis-Test

26.1 Einfaktorielle Varianzanalyse

Ziel: Die einfaktorielle Varianzanalyse untersucht den Zusammenhang einer nominalskalierten Variable und einer metrisch skalierten Variable. Es wird überprüft, ob sich die Mittelwerte der Daten bezüglich der Ausprägungen der nominalskalierten Variable unterscheiden.

Voraussetzungen: Die Daten können als Werte unabhängiger Zufallsvariablen betrachtet werden. Die Variable, deren Mittelwerte $\bar{x}_j$ verglichen werden, ist metrisch skaliert. Die Abweichungen der Daten von den jeweiligen Mittelwerten $\bar{x}_j$ sind normalverteilt mit $N(0, \sigma_j^2)$, wobei näherungsweise $\sigma_j^2 = \sigma_l^2$ für alle j und l . (Die Indizes j, l stehen für die Ausprägungen der nominalskalierten Variable). Die Varianzanalyse wird auch ANOVA genannt (Analysis of Variance).

Wir erklären zuerst die Problemlage mit Hilfe eines Beispiels:

Beispiel 26.1.1. *Wir möchten wissen, ob sich das durchschnittliche Einkommen von Bevölkerungsgruppen in einem Land unterscheidet. Wir nehmen an, wir hätten 4 Gruppen. Wir würden eine Zufallsstichprobe ziehen und jeweils das Einkommen in Erfahrung bringen und dabei die folgenden Resultate erhalten (s. Tabelle 26.1.1; die Daten finden Sie unter 19Anova.xls).*

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
50000	42000	41000	50100
40000	38000	39000	51000
45000	50100	36000	38000
43000	52000	43500	49000
60200	48000	38000	45000
55000	37000	37500	55000
47000	25000	28000	42000
58000	37500		48000
	41000		53000
			48000

Tabelle 26.1.1: Beispieldaten für eine Varianzanalyse

◇

Wir nehmen zuerst eine graphische Analyse vor (s. Abbildung 29.2.1 und Abbildung 26.1.2):

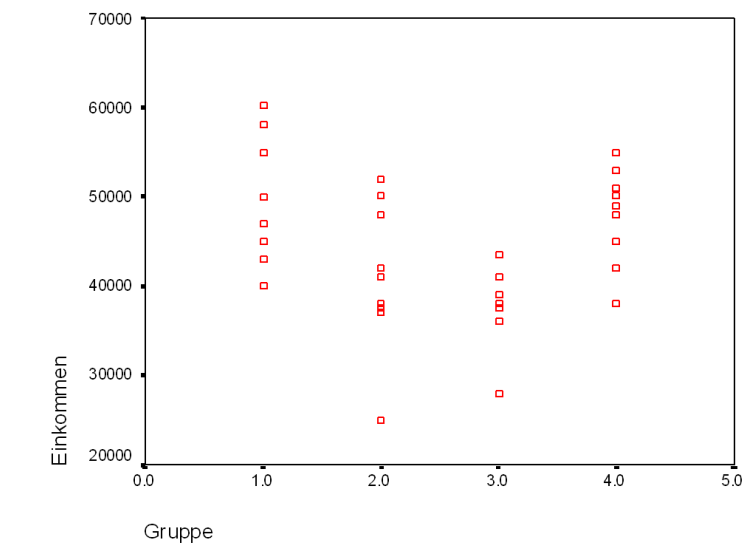


Abbildung 26.1.1: Graphische Darstellung der Daten der Tabelle 26.1.1 mittels Streudiagramm

Nützlich für Daten, die mit der Varianzanalyse untersucht werden können, sind auch die Box-plots (Herstellung und Interpretation der Graphik: siehe bivariate beschreibende Statistik und Abbildung 26.1.2):

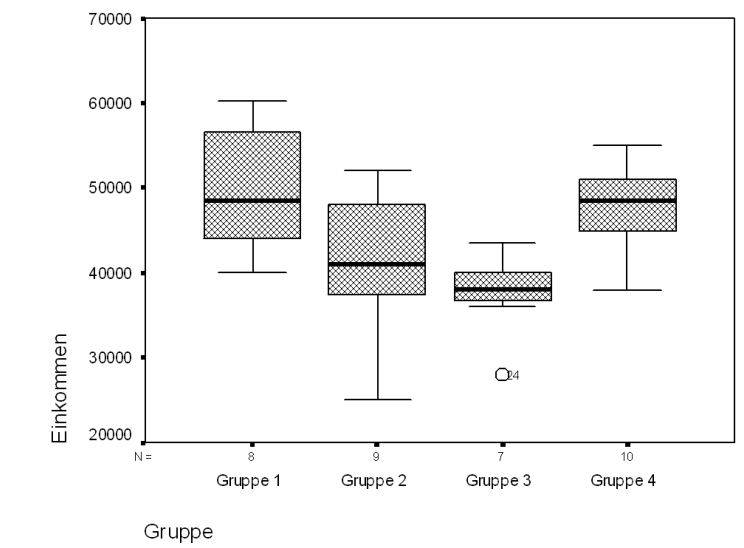


Abbildung 26.1.2: Graphische Darstellung der Daten der Tabelle 26.1.1 mittels eines Box-Plots

Wir können einen ersten Augenschein nehmen, ob sich die Streuungen der verschiedenen Gruppen wesentlich unterscheiden.

Die nominalskalierte Variable wird im Rahmen der Varianzanalyse „Faktor“ genannt. Die Menge der Objekte einer Ausprägung des Faktors wird „Gruppe“ genannt. In unserem Beispiel haben wir somit vier Gruppen (Gruppe 1, 2, 3 und 4).

Bei der Varianzanalyse verwenden wir folgenden Größen:

(1) SQM: Wir berechnen den Mittelwert des Einkommens für jede Gruppe. Für die vier Gruppen erhalten wir die Mittelwerte $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$. Die Mittelwerte unterscheiden sich u.a. nicht wesentlich, wenn die Differenz dieser Mittelwerte zum Gesamtmittelwert $\bar{x}_t$ zu vernachlässigen ist. Die Differenzen der Mittelwerte werden dabei durch den folgenden Ausdruck gemessen:

$$SQM = \sum_{j=1}^m k_j (\bar{x}_j - \bar{x}_t)^2.$$

(m Anzahl Gruppen, k_j Stichprobengrösse pro Gruppe j)

Wir bilden somit die quadrierte Abweichung des Mittelwertes $\bar{x}_j$ vom Gesamtmittelwert $\bar{x}_t$. Diesen Wert multiplizieren wir dann mit der Anzahl Daten k_j pro Gruppe j . Dann bilden wir die Summe dieser gewichteten, quadrierten Abweichungen. Diese Summe nennen wir künftig SQM (= Summe der gewichteten, quadrierten Abweichungen zwischen den Mittelwerten oder zwischen den Gruppen). Durch die Berechnungsart werden Abstände, die grösser sind, mehr gewichtet. Zudem erhalten wir nur positive Summanden.

(2) SQR: Um einen Testausdruck zu erhalten, werden wir zusätzlich die Summe der Summen der quadrierten Abweichungen der Daten x_{ij} der Gruppe j vom jeweiligen Mittelwert $\bar{x}_j$ verwenden.

$$SQR = \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j)^2$$

(SQR , R für „Residuum“, Summe der quadrierten Abweichungen innerhalb der Gruppen)

Definition 26.1.2. Das Residuum eines Datums x_{ij} der Gruppe j ist definiert als

$$r_{ij} := x_{ij} - \bar{x}_j$$

Der Ausdruck „Residuum“ (= Verbleibendes) ist wie folgt zu erklären. Man kann x_{ij} darstellen durch die folgende Summe:

$$x_{ij} = \bar{x}_t + (\bar{x}_j - \bar{x}_t) + r_{ij},$$

denn offensichtlich gilt:

$$\begin{aligned} \bar{x}_t + (\bar{x}_j - \bar{x}_t) + r_{ij} &= \bar{x}_t + (\bar{x}_j - \bar{x}_t) + x_{ij} - \bar{x}_j \\ &= x_{ij}. \end{aligned}$$

Während ein Teil der Abweichung der Daten der Gruppe j vom Gesamtmittelwert durch die Abweichung des Mittelwerts $\bar{x}_j$ vom Gesamtmittelwert *erklärt* wird (d.h. durch den Ausdruck $(\bar{x}_j - \bar{x}_t)$ in der Gleichung, wird die *restliche* oder *verbleibende* Abweichung r_{ij} durch den Ausdruck $(x_{ij} - \bar{x}_j)$ nicht erklärt.

Um die Berechnung der Testgrösse zu vereinfachen, führen wir neben SQM und SQR noch einen dritten Begriff ein:

(3) SQT: Die Gesamtsumme der quadratischen Abweichungen der Daten vom Gesamtmittelwert $\bar{x}_t$ (= SQT , T für „Total“) ist definiert als:

$$SQT = \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_t)^2.$$

Um SQT etwa mit Excel effizient zu berechnen, beachte man, dass gilt:

Satz 26.1.3.

$$SQT = (n-1)s_t^2,$$

wobei

$$s_t^2 := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_t)^2 = \frac{1}{n-1} \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_t)^2$$

die Varianz aller Daten ist und $n = \sum_{j=1}^m k_j$.

Weiterhin ist folgender Zusammenhang für die Berechnung nützlich:

Satz 26.1.4. $SQT = SQM + SQR$

Beweis.

$$\begin{aligned}
 \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_t)^2 &= \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j + \bar{x}_j - \bar{x}_t)^2 \\
 &= \sum_{j=1}^m \sum_{i=1}^{k_j} ((x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x}_t))^2 \\
 &= \sum_{j=1}^m \sum_{i=1}^{k_j} \left((x_{ij} - \bar{x}_j)^2 + 2((x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}_t)) + (\bar{x}_j - \bar{x}_t)^2 \right) \\
 &= \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j)^2 + \sum_{j=1}^m \sum_{i=1}^{k_j} 2((x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}_t)) + \sum_{j=1}^m \sum_{i=1}^{k_j} (\bar{x}_j - \bar{x}_t)^2
 \end{aligned}$$

Nun gilt:

$$\sum_{j=1}^m \sum_{i=1}^{k_j} (\bar{x}_j - \bar{x}_t)^2 = \sum_{j=1}^m k_j (\bar{x}_j - \bar{x}_t)^2$$

und

$$\begin{aligned}
 \sum_{j=1}^m \sum_{i=1}^{k_j} 2(x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}_t) &= \sum_{j=1}^m 2 \left((\bar{x}_j - \bar{x}_t) \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j) \right) \\
 &= 0
 \end{aligned}$$

(Denn $\sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j) = 0$ für jedes j).

Wir erhalten somit:

$\sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_t)^2 = \sum_{j=1}^m k_j (\bar{x}_j - \bar{x}_t)^2 + \sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j)^2$
$SQT = SQM + SQR$

□

Laut dem Satz können wir SQR mit Hilfe von SQT und SQM berechnen. SQT , SQM und SQR sind Streumasse, die wir künftig kurz „Streuung“ nennen. Die Streuungen sind mit den Varianzen nahe verwandt, wenn wir diese wie folgt definieren:

$$\text{Varianz} := \frac{\text{Summe quadrierter Abweichungen}}{\text{Freiheitsgrade}}$$

Die Freiheitsgrade sind die Anzahl der unabhängigen Zufallsvariablen, die in die Berechnung der Streuungen einfließen. Die Freiheitsgrade bei SQT betragen $n - 1$, da wir den Mittelwert aus den Daten berechnen und das letzte Datum bestimmt ist, wenn SQT und der Mittelwert bestimmt sind. Die Varianz ist damit $\frac{SQT}{n-1}$ und entspricht der in der beschreibenden Statistik eingeführten Varianz. SQM besteht aus m Zufallsvariablen. Der Wert der letzten ist wiederum bestimmt, sobald SQM und der Gesamtmittelwert gegeben sind. Damit betragen die Freiheitsgrade $m - 1$. Die Varianz der Mittelwerte ist damit gegeben durch $\frac{SQM}{m-1}$. Für SQR beträgt pro Gruppe die Anzahl

der Freiheitsgrade $k_j - 1$, da jeweils das letzte Residuum bestimmt ist, wenn des Gruppenmittel und die übrigen Residuen bestimmt sind. Insgesamt erhalten wir für SQR die Anzahl $\sum_{i=1}^m (k_j - 1) = \sum_{i=1}^m k_j - m = n - m$ von Freiheitsgraden. Damit beträgt die Varianz der Residuen $\frac{SQR}{n-m}$.

Satz 26.1.5. $df_{SQM} + df_{SQR} = df_{SQT}$ (df für Freiheitsgrade)

Beweis. Beweis: $m - 1 + n - m = n - 1$. □

Je grösser SQR , desto weniger sind die Unterschiede zwischen den Mittelwerten als bedeutsam zu betrachten, da die Daten dann stark streuen. Es hilft dann wenig zu wissen, zu welcher Gruppe ein Objekt gehört, um dessen vermutlichen Wert möglichst genau zu schätzen. Je grösser SQM , desto wichtiger sind die Unterschiede zwischen den Mittelwerten der Gruppen und desto nützlicher ist es zu wissen, zu welcher Gruppe ein Objekt gehört, um seinen Wert möglichst genau zu schätzen (siehe folgende Graphik. Bei den Daten links ist es wenig nützlich zu wissen, in welche Gruppe ein Objekt zu liegen kommt, bei den Daten rechts ist das nützlicher, obwohl die Mittelwerte der entsprechenden Gruppen identisch sind und damit die Abweichungen der Mittelwerte identisch sind (beachten Sie die unterschiedlichen Skalierungen der y-Achse!).

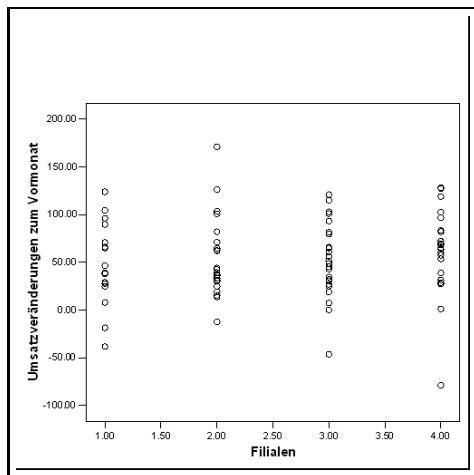


Abbildung 26.1.3: Beispiel für Datensätze mit identischen Mittelwerten und unterschiedlicher Varianz: in diesem Datensatz liegt eine grosse Varianz vor - man beachte die Skalierung

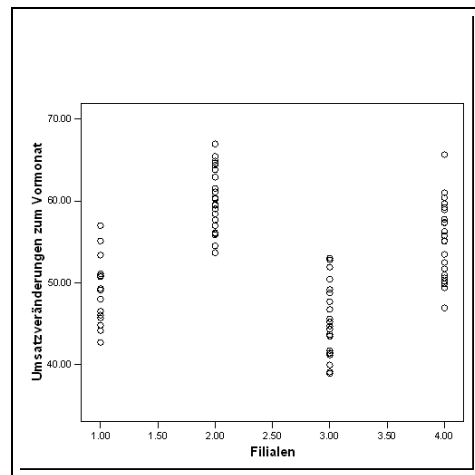


Abbildung 26.1.4: Beispiel für Datensätze mit identischen Mittelwerten und unterschiedlicher Varianz: in diesem Datensatz liegt eine kleinere Varianz vor - man beachte die Skalierung

Entsprechend kann man SQM an SQR messen, um ein Mass für die Bedeutsamkeit der Abweichungen der Mittelwerte der Gruppen zu erhalten. Statt $\frac{SQM}{SQR}$ werden wir jedoch als Teststatistik den Quotienten der entsprechenden Varianzen verwenden, da wir die Verteilung von $\frac{\frac{SQM}{m-1}}{\frac{SQR}{n-m}}$ kennen. Gemäss Voraussetzungen der Varianzanalyse sind die Residuen nämlich normalverteilt mit $\mu = 0$ und $\sigma := \sigma_1 = \dots = \sigma_m$. Damit sind die x_{ij} pro Gruppe j normalverteilt mit der Standardabweichung σ und dem geschätzten Erwartungswert $\bar{x}_j$ (s. Satz 14.2.4, Seite 309). Somit ist $\frac{x_{ij} - \bar{x}_j}{\sigma}$ standardnormalverteilt - wir nehmen ja eine Standardisierung der Zufallsvariable vor. Da die x_{ij} normalverteilt sind, ist $\bar{x}_j$ selbst für kleine Datensätze normalverteilt mit der Standardabweichung

$\frac{\sigma}{\sqrt{k_j}}$. Entsprechend ist auch $\frac{\bar{x}_j - \bar{x}_t}{\frac{\sigma}{\sqrt{k_j}}}$ standardnormalverteilt. Nun gilt durch Erweitern:

$$\frac{SQM}{SQR} = \frac{\sum_{j=1}^m k_j (\bar{x}_j - \bar{x}_t)^2}{\sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j)^2} = \frac{\sum_{j=1}^m \frac{k_j (\bar{x}_j - \bar{x}_t)^2}{\sigma^2}}{\sum_{j=1}^m \sum_{i=1}^{k_j} \frac{(x_{ij} - \bar{x}_j)^2}{\sigma^2}} = \frac{\sum_{j=1}^m \left(\frac{\bar{x}_j - \bar{x}_t}{\frac{\sigma}{\sqrt{k_j}}} \right)^2}{\sum_{j=1}^m \sum_{i=1}^{k_j} \left(\frac{x_{ij} - \bar{x}_j}{\sigma} \right)^2}$$

Gemäss Definition der $F_{m-1, n-m}$ -Verteilung, ist damit $\frac{\frac{SQM}{m-1}}{\frac{SQR}{n-m}} F_{m-1, n-m}$ verteilt.

$\frac{\frac{SQM}{m-1}}{\frac{SQR}{n-m}}$ ist $F_{m-1, n-m}$ -verteilt (sofern die Residuen gemäss $N(0, \sigma_i^2)$ verteilt sind)
m : Anzahl Gruppen
n : Anzahl Daten
SQM : $\sum_{j=1}^m k_j (\bar{x}_j - \bar{x}_t)^2$
SQR : $\sum_{j=1}^m \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j)^2$
k_j : Anzahl Daten der Gruppe j
$\bar{x}_j$: Mittelwert der Daten der Gruppe j
$\bar{x}_t$: Mittelwert aller Daten.
$s_i^2 = s_1^2 = \dots = s_m^2$ (die Varianzen der Residuen der Gruppen müssen identisch sein)

Bezüglich der Verteilung der Teststatistik wird jeweils rechtsseitig getestet, obwohl die Hypothesen bezüglich der Mittelwerte ungerichtet sind. Wie die Mittelwerte abweichen, muss separat durch den Vergleich der Mittelwerte überprüft werden.

Wir wenden uns nun der Berechnung des Beispiels zu:

Für beliebige einfaktorielle Varianzanalysen gelten die Hypothesen:

H_0 : Die Gruppenmittelwerte unterscheiden sich nicht.

H_A : Die Gruppenmittelwerte unterscheiden sich

Im obigen Beispiel berechnen wir ($\alpha = 0.05$):

$\bar{x}_t = 44438.23529$; $SQT := 2073080294$

	k_j	SQM_j
$\bar{x}_1 = 49775$	8	$8(49775 - 44438.23529)^2 = 2.2785 \times 10^8$
$\bar{x}_2 = 41177.77778$	9	$9(41177.77778 - 44438.23529)^2 = 9.5675 \times 10^7$
$\bar{x}_3 = 37571.42857$	7	$7(37571.42857 - 44438.23529)^2 = 3.3007 \times 10^8$
$\bar{x}_4 = 47910$	10	$10(47910 - 44438.23529)^2 = 1.2053 \times 10^8$
Summen	34	$SQM = 7.7413 \times 10^8$

Damit ist $SQR = SQT - SQM = 2073080294 - 7.7413 \cdot 10^8 = 1.2990 \times 10^9$. Wir können damit die folgende Tabelle konstruieren (mit „df“ für Freiheitsgrade):

Quadratsummen	df	Mittel der Quadrate (= Varianzen)	Testwert
$SQM = 7.7413 \times 10^8$	3	$\frac{SQM}{m-1} = \frac{7.7413 \times 10^8}{3} = 2.5804 \times 10^8$	$\frac{2.5804 \times 10^8}{4.33 \times 10^7} = 5.9594$
$SQR = 1.2990 \times 10^9$	30	$\frac{SQR}{n-m} = \frac{1.2990 \times 10^9}{30} = 4.33 \times 10^7$	
$SQT = 2073080294$	33		

Für den Testwert berechnen wir noch den p-Wert: $= 1 - f.vert(5.9594; 3; 30; 1) = 0.002589515 < 0.05$

Somit ist der Unterschied der Mittelwerte signifikant, sofern die Abweichungen der Daten von

den Gruppenmittelwerten normalverteilt sind und die Varianzen der verschiedenen Gruppen nicht signifikant verschieden sind.

Da sich die Summen der quadrierten Abweichungen additiv verhalten ($SQT = SQM + SQR$) können wir SQM in Prozenten von SQT ausdrücken. Es handelt sich um die erklärte Streuung in %. Im Beispiel ist $\frac{7.7413 \times 10^8}{1.2990 \times 10^9 + 7.7413 \times 10^8} \cdot 100 = 37.341\%$. Durch die Unterschiede zwischen den Mittelwerten der Gruppen werden 37.341% der Gesamtstreuung „erklärt“. $100 - 37.341\% = 62.659\%$ bleiben durch diese Unterschiede „unerklärt“.

Die zwei Voraussetzungen der Anwendungen des Testes müssen noch geprüft werden:

Normalverteilung der Residuen: Wir berechnen die Differenzen der Daten einer Gruppe j vom Mittelwerte der Gruppe j ; es gibt n Abweichungen für die konkrete Berechnung siehe „19Anova.xls“) und erstellen dann einen P-P-Plot Normal (s. Abbildung 26.1.5).

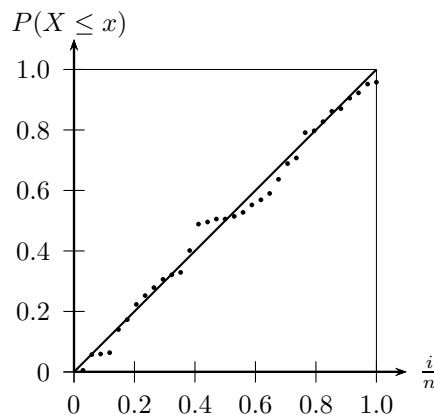


Abbildung 26.1.5: P-P-Plot auf Normalverteilung der Residuen der Daten von Tabelle 26.1.1

Es scheint nicht systematische Abweichungen von der $y=x$ -Achse zu geben. Wir betrachten somit die Abweichungen als normalverteilt (die Varianzanalyse ist relativ robust gegenüber Verletzungen der Normalverteilungsannahme).

Überprüfung der Varianzgleichheit: Wir führen dazu die Variante des Levene-Testes ein, der in diesem Zusammenhang z.B. vom Statistikprogramm SPSS angeboten wird. Dieser Levene-Test ist eine Varianzanalyse, die auf die *Beträge* der Abweichungen von den Gruppenmittelwerten ausgeführt wird. Wenn die Varianz der Residuen überall genügend ähnlich ist, so werden sich die Mittelwerte der Beträge nicht signifikant unterscheiden: durch die Berechnung der Beträge werden die negativen Residuen in den positiven Bereich reflektiert und die Beträge liegen bei ähnlicher Varianz alle in einem Band über der x-Achse (für die Berechnung, siehe „19Anova.xls“. Statt einer Varianzanalyse kann man auch einen Kruskal-Wallis-Test auf den Betrag der Residuen durchführen - s. nächstes Unterkapitel). Dies ergibt (s. Tabelle 26.1.2):

	Quadratsummen	df	Varianzen	Testwert	p-Wert
SQM	55833455.88	3	18611151.96	1.206009952	0.324468234
SQR	462960158.8	30	15432005.29		
SQT	518793614.7	33			

Tabelle 26.1.2: Berechnung der ANOVA zwecks Überprüfung der Varianzgleichheit

Die Varianzen unterscheiden sich nicht signifikant.

Bemerkung 26.1.6. • *Einzelvergleiche: Stellen wir ein signifikantes Resultat fest, so wäre es interessant zu untersuchen, bei welchen Gruppen die Unterschiede entscheidend sind. Dabei möchte man auch hier wieder die Signifikanz überprüfen. Es ergibt sich das Problem, dass dadurch die Gefahr steigt, Opfer eines Fehlers erster Ordnung zu werden. Um dieser*

Gefahr zu entgehen, gibt es verschiedene Verfahren, deren Darlegung jedoch zu weit ginge. Die Verfahren bestehen darin, das Signifikanzniveau mit der Anzahl der Einzelvergleiche zu verkleinern.

- Ist der Faktor ordinalskaliert, so werden auch Trenduntersuchungen vorgenommen. Ein linearer Trend liegt vor, wenn der Mittelwert von Gruppe zu Gruppe um eine ungefähr gleichbleibende Konstante ansteigt. Sinnvoll von einem linearen Trend kann man aber nur sprechen, wenn man gleiche Abstände zwischen den Ausprägungen der ordinalskalierten Variable postulieren kann. Damit wird sie zu einer metrisch skalierten Variable deklariert.

◇

26.2 Kruskal-Wallis-Test

Es handelt sich um einen Rangtest, der für die selbe Situation wie die einfaktorielle ANOVA angewendet wird. Er ist der ANOVA i.A. vorzuziehen, da er robuster ist (Dafür ist er etwas weniger mächtig). Zudem kann er für ordinalskalierte Daten verwendet werden. Er wird gewöhnlich im Levene-Test bei der Überprüfung der Varianzhomogenität statt der ANOVA verwendet.

Ziel: Der Kruskal-Wallis-Test untersucht den Zusammenhang einer nominalskalierten Variable und einer metrisch oder ordinal skalierten Variable. Es wird überprüft, ob sich die Verteilung der Daten bezüglich der Ausprägungen der nominalskalierten Variable unterscheiden.

Voraussetzungen: Die Variable, deren Verteilungen verglichen werden, ist metrisch oder ordinal skaliert. Es gibt nicht zu viele Bindungen.

Beim Kruskal-Wallis-Test werden die Daten der Grösse nach geordnet. Ihnen werden dann in bekannter Manier (siehe Wilcoxon-Test) Ränge zugeordnet. Man kann die folgende χ^2 -verteilte Teststatistik entwickeln:

$K = \frac{12}{n(n+1)} \sum_{j=1}^m k_j \left(\bar{R}_j - \frac{n+1}{2} \right)^2$ ist asymptotisch χ_{m-1}^2 -verteilt.
$\bar{R}_j = \frac{1}{k_j} \sum_{i=1}^{k_j} r_{ij}$ (Durchschnittlicher Rang der Gruppe j)
r_{ij} = Rang des Datums x_{ij}
n : Anzahl der Daten
k_j : Anzahl der Daten der Gruppe j
m : Anzahl der Gruppen

$\frac{n+1}{2}$ ist der Mittelwert aller Ränge. Die Summe der Ränge ist nämlich $\frac{n(n+1)}{2}$. Entsprechend ist deren Mittelwert $\frac{n(n+1)}{n \cdot 2} = \frac{n+1}{2}$. Wir berechnen also ähnlich wie bei der ANOVA (SQM) die Summe der gewichteten und quadrierten Abweichungen der Rangmittel $\bar{R}_j$ pro Gruppe j vom Mittel aller Ränge.

Man kann für die Varianz $s_{\bar{R}_j}^2$ der Rangmittel zeigen:

$$s_{\bar{R}_j}^2 = \frac{\frac{n(n+1)}{12}}{k_j} = \frac{1}{k_j} \frac{n(n+1)}{12}$$

Die Standardisierung der Rangmittel ist somit:

$$\frac{\bar{R}_j - \frac{n+1}{2}}{\sqrt{\frac{1}{k_j} \frac{n(n+1)}{12}}} \sim N(0, 1) \text{ (asymptotisch).}$$

Entsprechend ist dann

$$\sum_{j=1}^m \left(\frac{\bar{R}_j - \frac{n+1}{2}}{\sqrt{\frac{1}{k_j} \frac{n(n+1)}{12}}} \right)^2 \text{ ist asymptotisch } \chi_{m-1}^2 - \text{verteilt.}$$

(laut Definition der Chi-Quadrat-Verteilung)

Wobei

$$\begin{aligned}\sum_{j=1}^m \left(\frac{\bar{R}_j - \frac{n+1}{2}}{\sqrt{\frac{1}{k_j} \frac{n(n+1)}{12}}} \right)^2 &= \frac{1}{\frac{n(n+1)}{12}} \sum_{j=1}^m k_j \left(\bar{R}_j - \frac{n+1}{2} \right)^2 \\ &= \frac{12}{n(n+1)} \sum_{j=1}^m k_j \left(\bar{R}_j - \frac{n+1}{2} \right)^2\end{aligned}$$

Beispiel 26.2.1. Wir führen den Kruskal-Wallis-Test auf das erste Beispiel aus. Für die detaillierte Lösung mit Hilfe von Excel siehe 19Anova.xls.

Wir erhalten die Rangmittel:

$$\bar{R}_1 = 23.625$$

$$\bar{R}_2 = 13.66666667$$

$$\bar{R}_3 = 8.285714286$$

$$\bar{R}_4 = 22.5$$

$$\frac{12}{n(n+1)} \sum_{i=1}^k \left(\bar{R}_j - \frac{n+1}{2} \right)^2 = \frac{12}{34(35)} (8 \left(23.625 - \frac{35}{2} \right)^2 +$$

$$9 \left(13.66666667 - \frac{35}{2} \right)^2 + 7 \left(8.285714286 - \frac{35}{2} \right)^2 + 10 \left(22.5 - \frac{35}{2} \right)^2) = 12.874$$

$$F_{\chi^2_3}(12.874) = 1 - 0.004917164; p\text{-Wert: } 0.004917164$$

Somit sind die Verteilungen signifikant verschieden. $\diamond$

Bemerkung 26.2.2. Gewöhnlich wird eine Bindungskorrektur vorgenommen: bei vielen Bindungen (mehr als $\frac{1}{4}$ der Ränge) empfiehlt es sich, den Chi-Quadrat-Wert zu korrigieren. Die Bindungskorrektur wird berechnet durch die Multiplikation des Chi-Quadrat-Wertes mit der folgenden Zahl:

$$\frac{1}{1 - \frac{1}{n^3 - n} \sum_{i=1}^g (t_j^3 - t_j)}$$

wobei g die Anzahl der Gruppen von verbundenen Ränge ist; t_j Anzahl der Daten pro Gruppe j verbundener Daten. Fürs Beispiel erhalten wir 6 Gruppen von verbundenen Rängen mit 2 Daten sowie 2 Gruppen von verbundenen Rängen mit 3 Daten. Somit wird

$$\sum_{j=1}^g (t_j^3 - t_j) = 6(2^3 - 2) + 2(3^3 - 3) = 84$$

und

$$\frac{1}{1 - \frac{1}{34^3 - 34} 84} = 1.002143623$$

Wir multiplizieren $12.874 * 1.002143623 = 12.902$. Im Beispiel sehen wir jedoch, dass die Unterschiede bei der doch recht hohen Anzahl von Bindungen klein sind. $F_{\chi^2_3}(12.902) = 1 - 0.004853401$; $p\text{-Wert: } 0.004853401$ $\diamond$

26.2.1 Übungen

1. Es wird vermutet, dass je nach Einkommen unterschiedlich viel für Schmuck ausgegeben wird. Eine Marktstudie im Auftrag der Schmuckindustrie macht eine Erhebung, die zu folgenden Ergebnissen führt (Excel-Datei: 19Anova.xls). Nehmen Sie mit einer Statistiksoftware eine graphische Analyse des Materials vor. Machen Sie dann eine Varianzanalyse mit Hilfe von

Excel und mit einer Statistiksoftware. Überprüfen Sie die Voraussetzungen. Wenden Sie auf die Daten auch einen Kruskal-Wallis-Test an (mit Excel und mit einer Statistiksoftware berechnet).

- Denken Sie sich ein betriebswirtschaftliches Problem aus, das mit Hilfe der obigen Tests zu lösen ist.

26.2.2 Lösungen

- Wir nehmen zuerst eine graphische Analyse vor:
Streudiagramm (s. Abbildung 26.2.6):

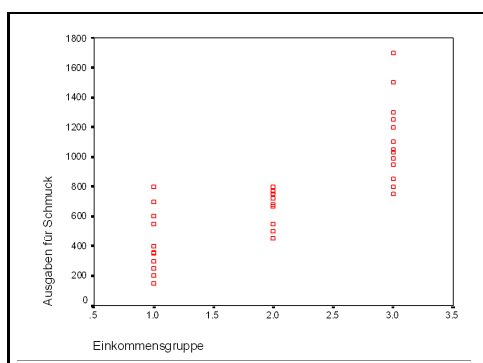


Abbildung 26.2.6: Streudiagramm der Daten der Übung

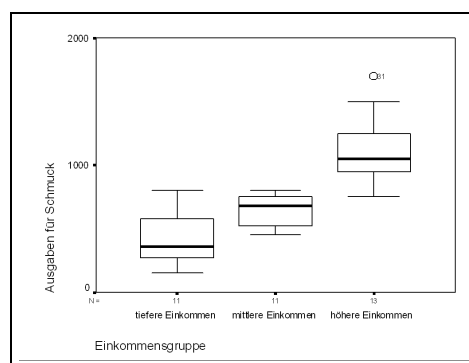


Abbildung 26.2.7: Boxplot der Daten der Übung

Auf Grund des Boxplots scheinen sich die erste und die dritte, sowie die zweite und die dritte Gruppe signifikant zu unterscheiden.

H_0 : Die Mittelwerte unterscheiden sich nicht.

H_A : Die Mittelwerte unterscheiden sich ($\alpha = 0.01$)

Berechnung mit Excel:

Wir berechnen die Mittelwerte für jede Gruppe:

$$\bar{x}_1 = 423.6363636$$

$$\bar{x}_2 = 649.0909091$$

$$\bar{x}_3 = 1113.076923$$

$$\bar{x}_g = 750.5714286$$

Dann füllen wir die folgende Tabelle schrittweise auf:

Quadratsummen	df	Varianzen	F
$SQM =$	$m - 1$	$\frac{SQM}{m-1}$	$\frac{SQM}{m-1} \cdot \frac{SQR}{m-r}$
$SQR =$	$n - m$	$\frac{SQR}{n-m}$	
$SQT =$	$n - 1$		

indem wir zuerst SQM berechnen:

$$SQM = 11(423.6363636 - 750.5714286)^2 + 11(649.0909091 - 750.5714286)^2 + 13(1113.076923 - 750.5714286)^2 = 2.9974 \times 10^6$$

Dann berechnen wir SQT : 4508188.571.

$$\text{Somit ist dann } SQR = SQT - SQM = 4508188.571 - 2.9974 \times 10^6 = 1.5108 \times 10^6$$

Quadratsummen	df	Varianzen	F
$SQM = 2.9974 \times 10^6$	$3 - 1$	$\frac{2.9974 \times 10^6}{2} = 1.4987 \times 10^6$	$\frac{1.4987 \times 10^6}{47213} = 31.743$
$SQR = 1510822.378$	$35 - 3$	$\frac{1510822.378}{32} = 47213.$	
$SQT = 4508188.571$	$35 - 1$		

$F_{F_{2,32}}(31.743) = 1 - f.vert(31.743; 2; 32; 1) = 2.53142E-08$; p-Wert = $2.53142E-08 < 0.01$

Überprüfung der Voraussetzungen:

Für die Abweichungen erhalten wir den folgenden P-P-Plot (s. 19Anova.xls und Abbildung 26.2.8)

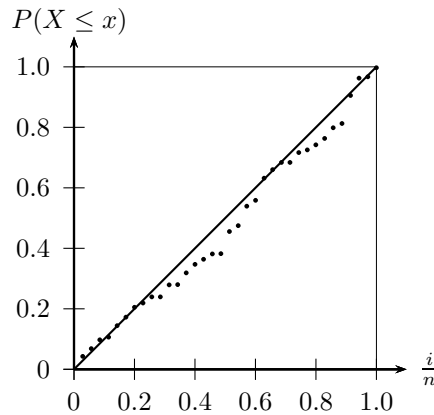


Abbildung 26.2.8: P-P-Plot auf Normalverteilung der Residuen der Übung

Die Abweichungen können als normalverteilt betrachtet werden.

Varianzhomogenität: Wir berechnen eine einfaktorielle ANOVA auf die Beträge der Abweichungen der Daten von den Gruppenmitteln und erhalten (s. Tabelle 26.2.3):

Quadratsummen	df	Varianzen	Testwert	p-Wert
SQM: 65738.85611	2	32869.42806	2.288366215	0.117793457
SQR: 459638.7112	32	14363.70973		
SQT: 525377.5673	34			

Tabelle 26.2.3: Berechnung der ANOVA auf den Betrag der Residuen zwecks Überprüfung auf Varianzgleichheit der Gruppen

Da der p-Wert grösser als 0.05 ist, kann die These von der Varianzgleichheit gehalten werden.

Übungshalber berechnen wir noch einen Test auf Verschiedenheit der Gruppenrangmittel mit dem Kruskal-Wallis-Test (mit Excel s. 19Anova.xls). Wir erhalten als mittlere Ränge:

$\bar{R}_1 = 8.318181818$;

$\bar{R}_2 = 15.13636364$;

$\bar{R}_3 = 28.61538462$

Dies ergibt dann:

$$\begin{aligned}
 K &= \frac{12}{n(n+1)} \sum_{j=1}^m k_j \left(\bar{R}_j - \frac{n+1}{2} \right)^2 \\
 &= \frac{12}{35(36)} \left(11 \left(8.318181818 - \frac{36}{2} \right)^2 \right. \\
 &\quad \left. + 11 \left(15.13636364 - \frac{36}{2} \right)^2 + 13 \left(28.61538462 - \frac{36}{2} \right)^2 \right) \\
 &= 24.631
 \end{aligned}$$

$F_{\chi^2_2}(24.631) = 1 - 4.48174E-06$; p-Wert: $4.48174E-06 < 0.01$;

Das Resultat ist hochsignifikant.

26.3 Lernziele

- Die Varianzanalyse vornehmen können (Graphiken, Berechnung der Teststatistik; Überprüfung der Anwendungsbedingungen)
- Den Kruskal-Wallis-Test vornehmen können (samt Bindungskorrektur)
- Sich je nach Eigenschaften der Daten für einen der beiden Tests bei geeigneter Datenlage entscheiden können.

Kapitel 27

Tests für Korrelationskoeffizienten

27.1 Test für den Pearson-Korrelationskoeffizienten

Ziel: Es soll der lineare Zusammenhang zwischen zwei metrisch skalierten Variablen X und Y untersucht werden.

Voraussetzung: Die Daten können als Ausprägungen unabhängiger, identisch normalverteilter Zufallsvariablen betrachtet werden. Metrisch skaliert. Die Daten liegen in etwa auf einer Geraden und liegen ungefähr punktsymmetrisch zum Punkt $(\bar{x}, \bar{y})$. (Quelle: (Hartung, 1999, S. 546 ff.))

Es kann gezeigt werden, dass für *normalverteilte* Zufallsvariablen gilt:

$\frac{R_{xy}\sqrt{n-2}}{\sqrt{1-R_{xy}^2}}$ ist t_{n-2} -verteilt.
R_{xy} : Zufallsvariable mit konkreten Pearson-Korrelationskoeffizienten r_{xy} als Wert
n : Stichprobengrösse

Für die Daten des Beispiels 3 (Korrelation.xls; s. auch bivariate beschreibende Statistik vom letzten Jahr):

H_0 : Es gibt keinen umgekehrt proportionalen Zusammenhang zwischen Alter und der Höhe der Jahresprämien für die Autoversicherung.

H_A : Je höher das Alter, desto weniger muss man Autoversicherung zahlen ($\implies$ linksseitiger Test, $\alpha = 0.05$)

Wir überprüfen mit einem P-P-Plot (und zusätzlich vielleicht mit einem χ^2 -Anpassungstest) die Normalverteilungsannahme (s. Abbildungen 27.1.1 und 27.1.2):

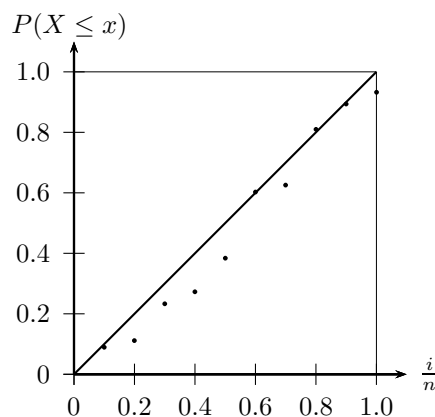


Abbildung 27.1.1: P-P-Plot für die Variable Alter

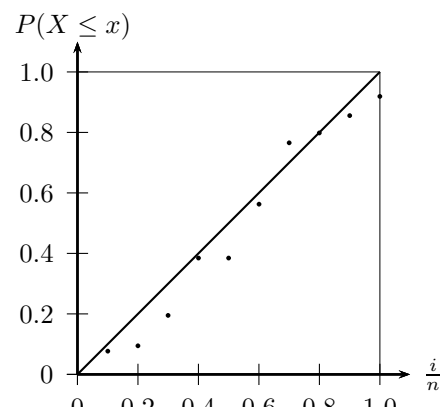


Abbildung 27.1.2: P-P-Plot für die Variable Prämienhöhe

Die Daten können in beide Dimensionen als normalverteilt betrachtet werden, wobei zu beachten ist, dass die Datenmenge sehr klein ist.

Wir können somit den obigen t-Test anwenden. Wir berechnen mit Excel oder mit einer Statistiksoftware den Pearson-Korrelationskoeffizienten: Excel: -0.992204036 . Wir berechnen den Wert der Teststatistik

$$\frac{R_{xy}\sqrt{n-2}}{\sqrt{1-R_{xy}^2}} : \frac{-0.992204036 \cdot \sqrt{10-2}}{\sqrt{1-(-0.992204036)^2}} = -22.519$$

$F_{t_8}(-22.519) = 1 - F_{t_8}(22.519) = 1 - (1 - 8.00423 \cdot E-09) \approx 0 < 0.05$. Die Nullhypothese kann verworfen werden.

27.2 Test für die Spearman-Rangkorrelation

Ziel: Es soll der Zusammenhang zwischen zwei metrisch oder ordinal skalierten Variablen X und Y untersucht werden.

Voraussetzung: Die Daten können als Werte identisch verteilter, unabhängiger Zufallsvariablen betrachtet werden. Sie sind metrisch oder ordinal skaliert. Es liegen nicht zu viele Bindungen vor (Quelle: (Hartung, 1999, S. 553 ff.)).

In der Teststatistik wird nicht R_S (Spearman-Rangkorrelation) verwendet, sondern die "Hotelling-Pabst-Statistik":

$$D = \sum_{i=1}^n [R(x_i) - R(y_i)]^2$$

mit $R(x_i)$ als Rang von x_i und $R(y_i)$ als Rang von y_i . Diese Zahl wird um so kleiner, je grösser die Korrelation ist und um so grösser, je kleiner die Korrelation ist. Beim Testen der Hypothese, dass die Korrelation grösser als 0 ist, liegt damit ein linksseitiger Test vor, beim Testen der Hypothese, dass die Korrelation kleiner als 0 ist, liegt ein rechtsseitiger Test vor. Beim Testen, dass die Korrelation von 0 verschieden ist, liegt ein zweiseitiger Test vor.

Für die Verteilung des Testwertes gelten ($n \leq 30$) die kritischen Werte der Tabelle 27.2.1, wobei jeweils nur die kritischen Werte $g_{u,\alpha}^k$ für den linksseitigen Test zum Signifikanzniveau α angegeben sind. Wegen der Symmetrie können damit auch rechtsseitige Tests berechnet werden, da für die entsprechende obere kritische Grenze $g_{o,\alpha}^k$ gilt:

$$g_{o,\alpha}^k = \frac{1}{3}n(n^2 - 1) - g_{u,\alpha}^k.$$

Die Nullhypothese ist bei zweiseitigem Testen auf dem α -Niveau zu verwerfen, wenn $D \leq g_{u,\frac{\alpha}{2}}^k$ oder $D \geq g_{o,\frac{\alpha}{2}}^k$.

Für $n > 30$ gilt näherungsweise:

$$\frac{D - E(D)}{\sqrt{Var(D)}} \sim N(0, 1).$$

mit: D wie oben.

n = Anzahl Daten pro Variable

$R(x_i)$ = Rang von x_i

$$E(D) = \frac{1}{6}(n^3 - n) - \frac{1}{12} \sum_{j=1}^m (d_j^3 - d_j) - \frac{1}{12} \sum_{j=1}^p (h_j^3 - h_j)$$

m = Anzahl der unterschiedlichen Werte x_i in der Messreihe X.

p = Anzahl der unterschiedlichen Werte y_i in der Messreihe Y.

d_j = Anzahl der Beobachtungen, die den Rang $R(x_i)$ aufweisen.

h_j = Anzahl der Beobachtungen, die den Rang $R(y_i)$ aufweisen.

n	$\alpha =$			
	0.01	0.025	0.05	0.1
4			2	2
5	2	2	4	6
6	4	6	8	14
7	8	14	18	26
8	16	24	32	42
9	28	38	50	64
10	44	60	74	92
11	66	86	104	128
12	94	120	144	172
13	130	162	190	226
14	172	212	246	290
15	224	270	312	364
16	284	340	390	450
17	356	420	480	550
18	438	512	582	664
19	532	618	696	790
20	638	738	826	934
21	758	870	972	1092
22	892	1020	1134	1270
23	1042	1184	1312	1464
24	1208	1366	1510	1678
25	1390	1566	1726	1912
26	1590	1786	1960	2168
27	1808	2024	2216	2444
28	2046	2284	2494	2744
29	2306	2564	2796	3068
30	2584	2868	3120	3416

Tabelle 27.2.1: (kritischer Werte der Hotelling-Pabst-Statistik; Nullhypothese wird verworfen, wenn Testwert kleiner-gleich kritischer Wert

Die Ausdrücke hinter $\frac{1}{6}(n^3 - n)$ stellen Bindungskorrekturen dar.
Für die Berechnung von

$$-\frac{1}{12} \sum_{j=1}^m (d_j^3 - d_j) - \frac{1}{12} \sum_{j=1}^p (h_j^3 - h_j)$$

genügt es deshalb, die Ausprägungen mit gleichen Rängen zu betrachten, da sonst

$$(d_j^3 - d_j) = 1^3 - 1 = 0 = (h_j^3 - h_j).$$

Liegen keine Bindungen - d.h. gleiche Ränge - vor, gilt somit:

$$E(D) = \frac{1}{6}(n^3 - n).$$

$$Var(D) = \frac{(n-1)(n+1)^2 n^2}{36} \left(1 - \frac{\sum_{j=1}^m (d_j^3 - d_j)}{n^3 - n} \right) \left(1 - \frac{\sum_{j=1}^p (h_j^3 - h_j)}{n^3 - n} \right)$$

Die Ausdrücke hinter $\frac{(n-1)(n+1)^2 n^2}{36}$ stellen Bindungskorrekturen dar. Auch hier gilt entsprechend,

dass wir für $\sum_{j=1}^m (d_j^3 - d_j)$ und $\sum_{j=1}^p (h_j^3 - h_j)$ nur Daten mit gleichen Rängen berücksichtigen müssen.

Wir führen das Beispiel fort:

H_0 : Die Körpergrösse und das Gewicht sind unabhängig.

H_A : Die Körpergrösse und das Gewicht sind abhängig ($\implies$ zweiseitig, $\alpha = 0.05$)

Da $n \leq 30$, genügt es, den Testwert D zu berechnen. Dieser ergibt (s. Tabelle 27.2.2)

Ränge Körpergrösse	Ränge Gewicht	$(R(x_i) - R(y_i))^2$
1	1	0
2	2	0
3	3	0
4	4	0
5	6	1
6	5	1
7.5	7	0.25
7.5	10	6.25
9	8	1
10	9	1
11	11	0
	Summe	10.5

Tabelle 27.2.2: Berechnung der Teststatistik der Spearman-Korrelation

Wir schauen für $n = 11$ in der obigen Tabelle nach und erhalten den Wert 86. Da $10.5 \leq 86$, ist das Resultat hoch signifikant auf dem angegebenen Niveau.

27.3 Test für die partielle Korrelation

Ziel: Es soll der Zusammenhang zwischen zwei Variablen (unter Ausschluss des Einflusses einer dritten Variable) getestet werden.

Voraussetzungen: Alle drei Variablen sind metrisch skaliert (jeweils unabhängig und identisch normalverteilt). Die paarweisen zweidimensionalen Punktwolken liegen in etwa auf einer Geraden und ungefähr punktsymmetrisch zu den jeweiligen Punkten, welche die Mittelwerte der beiden Variablen als Komponenten enthalten (Quelle: (Hartung, 1999, S. 561 ff.)).

$\frac{R_{xy/z} \cdot \sqrt{n-3}}{\sqrt{1-R_{xy/z}^2}}$ ist t_{n-3} -verteilt.
$R_{xy/z}$: Zufallsvariable mit partiellen Korrelationen von X und Y unter Ausschluss des Einflusses von Z als Werte.
n : Stichprobengrösse.

Am Beispiel 5 (s. Korrelation.xls):

$$\frac{-0.20527 \sqrt{12-3}}{\sqrt{1 - (-0.20527)^2}} = -0.62921 \text{ ist } t_{12-3} - \text{verteilt.}$$

Für zweiseitig: p-Wert: 0.544849731

Für einseitig: p-Wert: 0.272424865

Somit ist das Resultat nicht signifikant, während die obige Korrelation zwischen Gewicht und Geschicklichkeit ein signifikantes Testresultat ergibt mit:

$$\frac{R_{xy}\sqrt{n-2}}{\sqrt{1-R_{xy}^2}} = \frac{-0.755509151\sqrt{12-2}}{\sqrt{1-(-0.755509151)^2}} = -3.6468$$

p-Wert einseitig: 0.002672027

p-Wert zweiseitig: 0.005344054

Überprüfung der Normalverteilungsvoraussetzung: (s. Abbildungen 27.3.3 - 27.3.5):

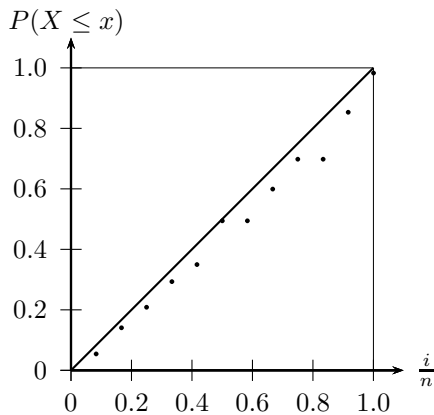


Abbildung 27.3.3: P-P-Plot für die Variable Schnelligkeit

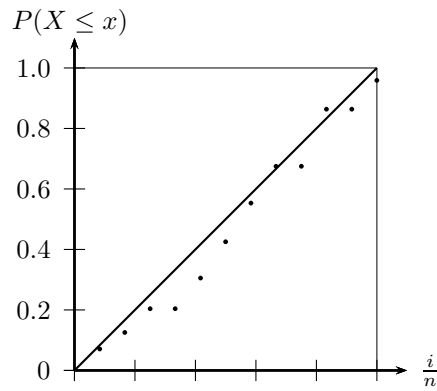


Abbildung 27.3.4: P-P-Plot für die Variable Körpergröße

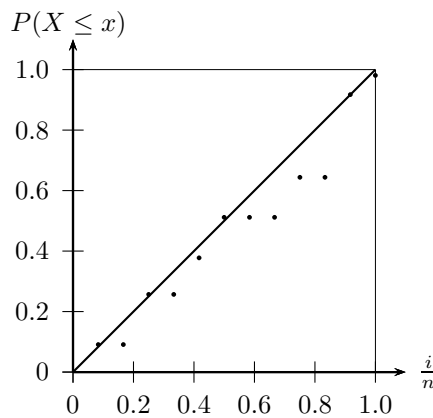


Abbildung 27.3.5: P-P-Plot für die Variable Alter

Die Normalverteilungsannahme ist wohl knapp haltbar (bei Körpergröße gibt es eine etwas systematische Abweichung; es hat eigentlich zu wenig Daten).

27.3.1 Übungen

1. Wählen Sie für die folgenden Daten einen geeigneten Korrelationskoeffizienten aus und berechnen Sie anschließend eine geeignete Teststatistik (von Hand, zweiseitig, $\alpha = 0.05$). der Daten in Korrelation.xls, Übung 1.
2. Berechnen Sie mit Excel und einer Statistiksoftware einen geeigneten Korrelationskoeffizienten der Daten in Korrelation.xls, Übung 2.

Quelle: Hängige Verfahren, EUR-OP, 4/00, Zahlen des Jahres 1998.

Wir vermuten, dass die Macht von Staaten bei sonst gleichen Bedingungen mit der Bevölkerungsgröße

zusammenhängt und dass Macht sich proportional zur Nichtbeachtung von internationalen Regeln verhält. Entsprechend schlagen wir bezüglich der Daten die folgenden Hypothesen vor:

H_0 : Die Bevölkerungsgrösse wirkt sich nicht auf Anzahl hängiger Verfahren aus

H_A : Je grösser die Bevölkerungszahlen, desto grösser die Anzahl hängiger Verfahren ($\Rightarrow$ einseitig, $\alpha = 0.05$)

Testen Sie!

3. In einem Betrieb wird das Verhalten der Mitarbeiter von zwei Vorgesetzten unabhängig beurteilt (auf einer Skala von 1 bis 10). Dabei ergeben sich die folgenden Resultate (Korrelation.xls, Übung 3).
Sind die Beurteilungen der Vorgesetzten vergleichbar? (führen Sie einen geeigneten Test durch).
4. Berechnen Sie die partielle Korrelation $r_{xy/z}$ für die Daten von Korrelation.xls (Übung 4) und testen Sie. (X = Lebenserwartung; Y = Alkoholkonsum; Z = Zigarettenkonsum)
5. Führen Sie die Übung 5 von Korrelation.xls durch.

27.3.2 Lösungen

1. Wir überprüfen zuerst graphisch, ob überhaupt eine lineare Tendenz vorliegt. Ausreisser liegen keine vor. Die Daten sind symmetrisch verteilt (s. Abbildung 27.3.5):

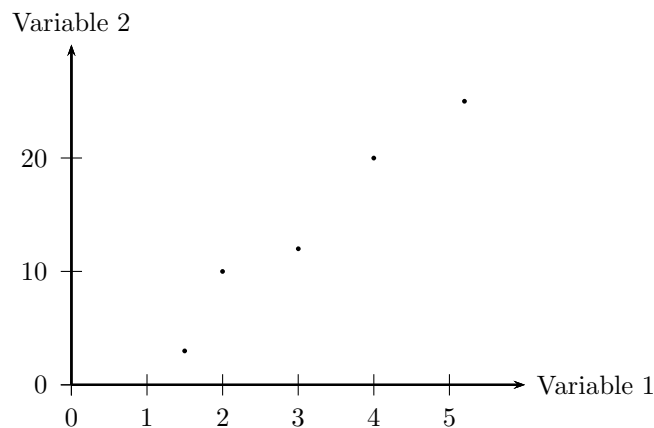


Abbildung 27.3.6: Zweidimensionales Streudiagramm der Daten der Übung

Allerdings liegen (um das Beispiel von Hand - mit vertretbarem Aufwand - ausrechenbar zu machen) wenig Daten vor.

a) Mit Excel erhalten wir: 0.979426112

Wir überprüfen diese Annahme mit P-P-Plots (s. Abbildungen 27.3.6 und 27.3.7):

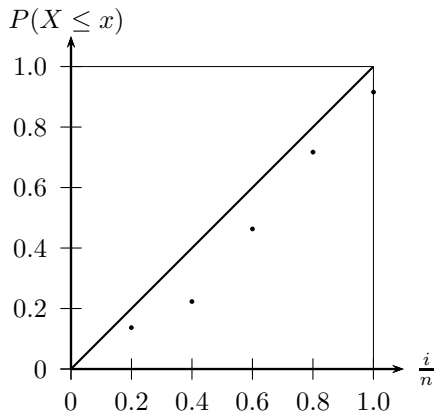


Abbildung 27.3.7: P-P-Plot für die Variable X

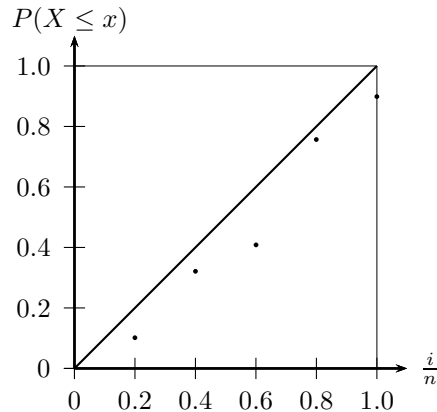


Abbildung 27.3.8: P-P-Plot für die Variable Y

Es liegen wenig Daten vor, so dass graphisch eine Aussage über deren Normalverteiltheit schwierig ist. Für den t-Test:

$$\frac{R_{xy}\sqrt{n-2}}{\sqrt{1-R_{xy}^2}} : \frac{0.97945 \cdot \sqrt{5-2}}{\sqrt{1-0.97945^2}} = 8.4113$$

$$F_{t_3}(8.4113) = 1 - 0.003525437 \Rightarrow$$

p-Wert bei zweiseitigem Testen: $2 \cdot 0.003525437 < 0.05$

Signifikanter Zusammenhang.

Es würde sich lohnen, sicherheitshalber noch den Test für die Spearman-Korrelation durchzuführen.

- Wir überprüfen zuerst graphisch, ob ein linearer, symmetrischer Zusammenhang vorliegt (s. Abbildung 27.3.9):

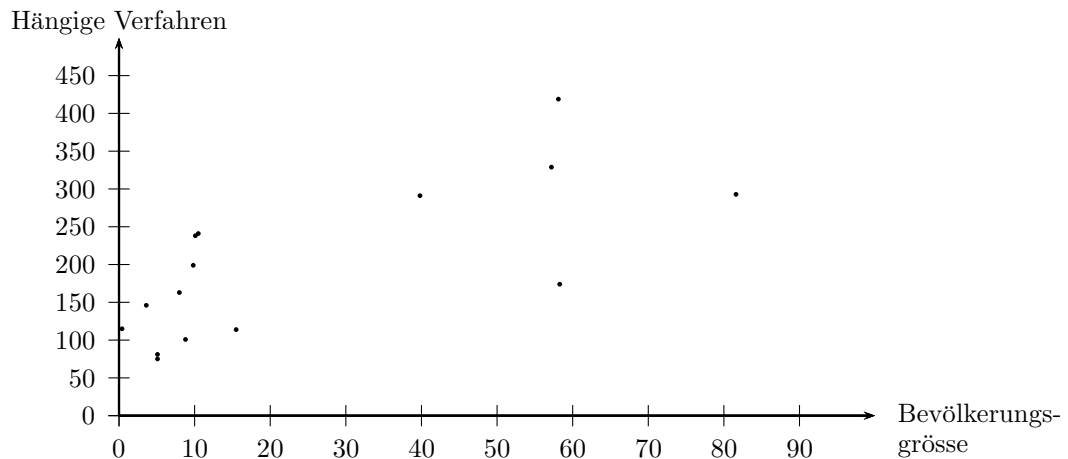


Abbildung 27.3.9: Zweidimensionales Streudiagramm der Daten der Übung

Die Symmetriebedingung ist kaum erfüllt. Deshalb sollte hier die Spearman-Korrelation verwendet werden (oder eine geeignete Transformation von Daten, z.B. logarithmieren einer oder beider Variablen). Aus der Verletzung der Symmetriebedingung folgt auch, dass die Daten nicht normalverteilt sind. Dies zeigen denn auch die entsprechenden P-P-Plots (s. Abbildungen 27.3.10 und 27.3.11):

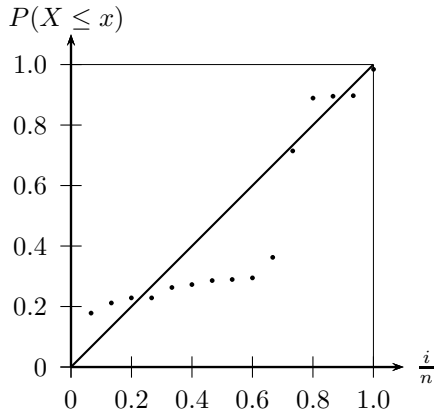


Abbildung 27.3.10: P-P-Plot für die Variable Bevölkerungsgröße

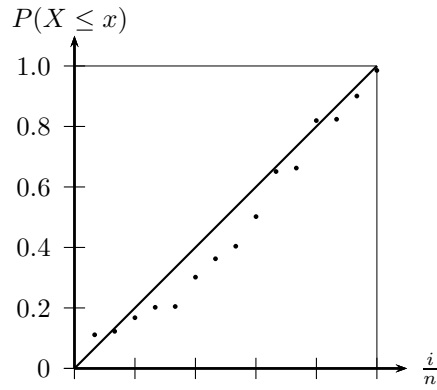


Abbildung 27.3.11: P-P-Plot für die Variable Hängige Verfahren

Während die Zahl der hängigen Verfahren als normalverteilt betrachtet werden können, ist die Bevölkerungsgröße der EU-Mitgliedstaaten nicht normalverteilt. Wir berechnen (auch) deshalb den Spearman-Korrelationskoeffizienten und erhalten durch die Pearson-Korrelation auf die Ränge: $r^{(Sp)} = 0.731010122$. Die Hypothese, dass es eine positive Korrelation zwischen Staatsgröße und hängigen Verfahren gibt führt im Falle der Spearman-Korrelation bezüglich der Statistik D zu einem linksseitigen Test.

Wir testen: $n < 30$, somit berechnen wir $D = 150.5$ und schauen in der Tabelle nach: für $n = 15$ erhalten wir auf dem 0.05–Niveau einseitig: $D = 150.5 < 312$. Somit ist der Zusammenhang signifikant.

Um das Verfahren zu illustrieren, berechnen wir noch den approximativen Testwert:

Es liegt nur eine Bindung mit zwei Ausprägungen bei der Variable X vor. Entsprechend gilt:

$$E(D) = \frac{1}{6}(n^3 - n) - \frac{1}{12} \sum_{j=1}^m (d_j^3 - d_j) - \frac{1}{12} \sum_{j=1}^p (h_j^3 - h_j) =$$

$$\frac{1}{6}(15^3 - 15) - \frac{1}{12}(2^3 - 2) = 559.5$$

$$Var(D) = \frac{(n-1)(n+1)^2 n^2}{36} \left(1 - \frac{\sum_{j=1}^m (d_j^3 - d_j)}{n^3 - n} \right) \left(1 - \frac{\sum_{j=1}^p (h_j^3 - h_j)}{n^3 - n} \right) =$$

$$\frac{(15-1)(15+1)^2 15^2}{36} \left(1 - \frac{2^3 - 2}{15^3 - 15} \right) (1) = 22360.0$$

Wir erhalten somit: $\frac{150.5 - 559.5}{\sqrt{22360.0}} = -2.7352$

$\Phi(-2.7352) = 0.003117173 < 0.05$

- Die Daten sind ordinalskaliert. Wir verwenden den Median, um die Gesamtbeurteilung zu vergleichen: Wir erhalten: 7.25 und 6.5. Der erste Vorgesetzte qualifiziert wohlwollender als der zweite. Die Vorgesetzten beurteilen vergleichbar, wenn ihre Qualifikationen positiv korreliert sind. Wir verwenden die Spearman-Korrelation:
 H_0 : Die Qualifikationen sind nicht positiv korreliert.
 H_A : Die Qualifikationen sind positiv korreliert ($\implies$ einseitig, $\alpha = 0.05$)
 Als Testwert: $D = 90$. Wir benutzen die Tabelle und erhalten für $\alpha = 0.05$ und $n = 12$:
 $g_{u,0.05} = 144 > 90 = D$
 Somit rangieren die beiden vergleichbar (positive Korrelation, die schlecht durch Zufall erklärbar ist).
- Die Streudiagramme können mit Statistiksoftwarepaketen gewöhnlich alle auf einmal erstellt werden. Wir erhalten (s. Abbildungen 27.3.12 - 27.3.14):

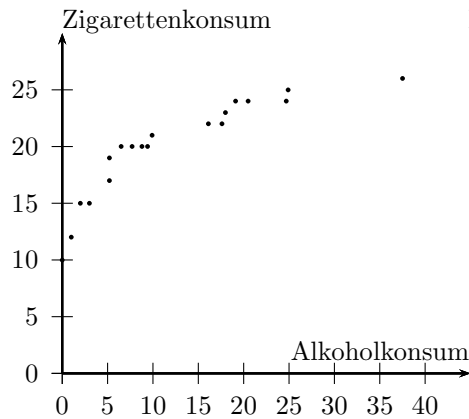


Abbildung 27.3.12: Streudiagramm Alkoholkonsum X Zigarettenkonsum

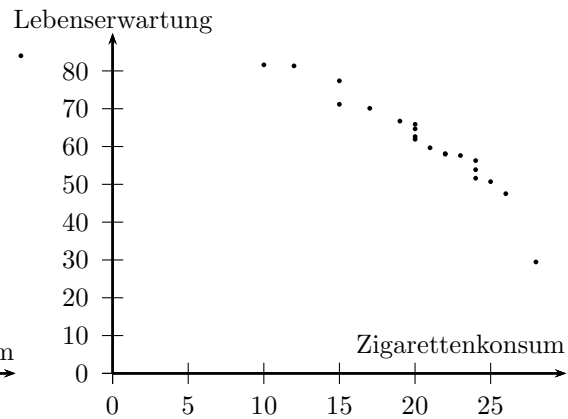


Abbildung 27.3.13: Streudiagramm Zigarettenkonsum X Lebenserwartung

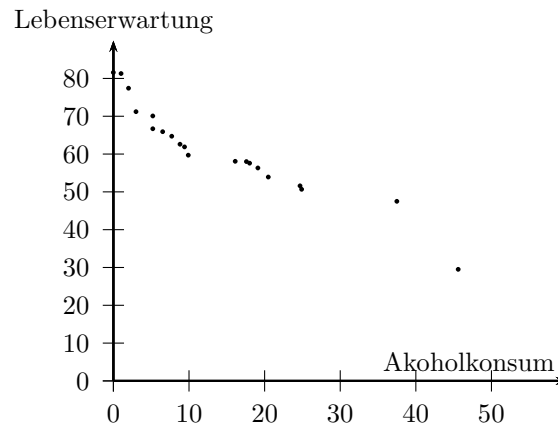


Abbildung 27.3.14: Zweidimensionales Streudiagramm der Daten der Übung

Ein Variablenpaar (Alkoholkonsum - Zigarettenkonsum) erfüllen die Voraussetzung bezüglich der Linearität nicht.

Übungshalber führen wir den Test zu Ende:

Normalitätsüberprüfung [27.3.3](#) - [27.3.5](#)):

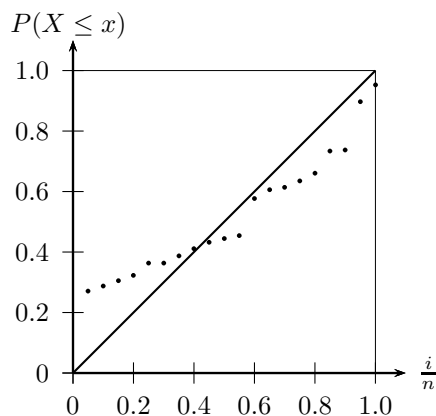


Abbildung 27.3.15: P-P-Plot für die Variable Alkoholkonsum

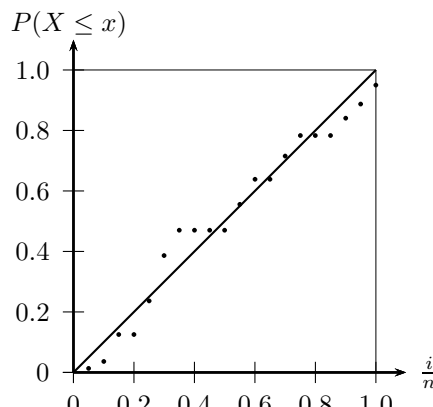


Abbildung 27.3.16: P-P-Plot für die Variable Zigarettenkonsum

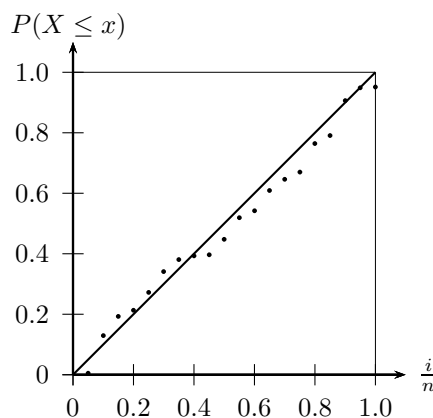


Abbildung 27.3.17: P-P-Plot für die Variable Lebenserwartung

Die Variable Alkoholkonsum ist nicht normalverteilt.

Übungshalber fahren wir fort:

Für $r_{xy/z}$ erhalten wir $-.7595$:

Testwert: $\frac{-.7595 \cdot \sqrt{20-3}}{\sqrt{1-(-.7595)^2}} = -4.8139$

Wir testen zweiseitig, da sonst nichts spezifiziert ist:

$F_{t_{20-3}}(-4.8139; 17) = 2 \cdot 8.1017 \times 10^{-5} < 0.05$.

Wäre - bei Erfüllung der Voraussetzungen hoch signifikant.

27.4 Lernziele

- Die eingeführten Tests für Korrelationskoeffizienten durchführen können.
- Übungen der obigen Art machen können.

Kapitel 28

Regression

28.1 Schätztheorie

Sind unabhängige Zufallsvariablen Y_i gegeben, so dass die Erwartungswert von Y_i von x_i abhängt, so lässt sich ein allenfalls vorliegender affiner Zusammenhang zwischen $\mathbf{x} := (x_1, \dots, x_n)$ und $Y = (Y_1, \dots, Y_n)$ durch das Modell

$$Y_i = b + ax_i + \mathcal{E}_i$$

ausdrücken - wobei für die unabhängigen und identisch verteilten Zufallsvariablen „Fehler“ $\mathcal{E}_i$ gilt:

$$E(\mathcal{E}_i) = 0. \quad (28.1.1)$$

Die $\mathcal{E}_i$ geben die zufälligen Abweichungen der y_i -Daten (= Werte der ZV Y_i) von der Regressionsgeraden an. Mit (28.1.1 und Satz 9.5.13 auf S. 218) erhält man (b , a und x_i sind konstant!)

$$E(Y_i) = E(b + ax_i + \mathcal{E}_i) = b + ax_i + E(\mathcal{E}_i) = b + ax_i.$$

b und a sind unbekannt und müssen geschätzt werden. Sind die Fehler $\mathcal{E}_i$ i.i.d. normalverteilt, so kann man zeigen, dass es vernünftig ist, a und b so zu schätzen, dass die Summe der quadrierten Abweichungen der Werte y_i der Zufallsvariablen Y_i von der geschätzten Geraden $\hat{f}(x_i) = \hat{a}x_i + \hat{b}$ minimal wird: wir berechnen also die Regressionsgerade durch die Punkte (x_i, y_i) . Die geschätzten Werte $\hat{a}$ und $\hat{b}$ sind dabei als Werte von Zufallsvariablen A und B zu betrachten (Statistiken). A und B werden definiert durch

$$A := \frac{\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

und

$$B := \bar{Y} - A\bar{x}$$

wobei n Paare (x_i, Y_i) gegeben sind. Man erhält diese Formeln, indem man in Satz 5.3.14 auf Seite 114 die entsprechenden Zufallsvariablen einsetzt.

Man kann zeigen, dass unter diesen Bedingungen $E(A) = a$ und $E(B) = b$ (d.h. A ist ein erwartungstreuer Schätzer von a und B ist ein erwartungstreuer Schätzer für b). Einen Beweis finden Sie am Schluss des Kapitels. Siehe für den Begriff der Erwartungstreue Seite 224 des Unterkapitels „Statistiken“ des Kapitels „Zufallsvariablen“ oder auch im Kapitel „Schätztheorie“, S. 539 ff.

Zudem kann man zeigen, dass A und B konsistente Schätzer von a beziehungsweise b sind. Eine Schätzstatistik A ist konsistent, wenn mit steigender Anzahl Zufallsvariablen, die in die Berechnung der Schätzstatistik einfließen, der Wert der Schätzstatistik sich dem Wert der zu schätzenden Grösse nähert. Präziser und für den vorliegenden Fall ist A eine konsistente Schätzstatistik für a genau dann, wenn $P(\lim_{n \rightarrow \infty} (A_n - a) = 0) = 1$, wobei n die Anzahl der Zufallsvariablen ausdrückt, die in die Berechnung der Schätzstatistik A einfließen (s. auch S. 539 ff).

Damit ist es sinnvoll, a und b im konkreten Fall durch $\hat{a}$ und $\hat{b}$ zu schätzen.

28.2 Testtheorie für die lineare Regression

28.2.1 Varianzanalyse (ANOVA)

Ziel: Es soll überprüft werden, ob die Regressionsgerade zu einer Reduktion der Varianz beim Schätzen führt.

Voraussetzungen: Die y -Daten können als Realisierungen unabhängiger Zufallsvariablen $(Y_1, \dots, Y_n)$ betrachtet werden. Die Residuen sind Werte von i.i.d. normalverteilten Zufallsvariablen mit Erwartungswert 0. Da die Residuen i.i.d. sind, müssen sie die selbe Varianz haben. Es darf keine Ausreisser haben und die Punkte sollten um eine gedachte Gerade herum liegen.

Wir können den Ausdruck $\frac{SQE}{SQR}$ als Mass für die „Erklärungskraft“ der Regressionsgerade betrachten (s. Seite 120, Kapitel Bivariate beschreibende Statistik). Ist SQE relativ zu SQR gross, so erklärt die Regressionsgerade die Daten gut, sonst schlecht. Bei guter „Erklärung“ ist es nützlich, die Regression zum Schätzen von y -Werten zu verwenden, da wir dann präzisere Schätzungen von y erhalten. Dies führt uns zu einer Varianzanalyse bezüglich der Regression, indem wir wieder statt $\frac{SQE}{SQR}$ die entsprechenden Varianzen (Summenquadrate geteilt durch Anzahl Freiheitsgrade) verwenden (siehe einfaktorielle Varianzanalyse, S. 480). Ohne Beweis halten wir fest, dass die Freiheitsgrade von SQE 1 und die Freiheitsgrade von SQR $n - 2$ betragen. Man könnte beweisen, dass gilt:

$\frac{SQE}{\frac{1}{n-2}SQR}$ ist $F_{1,n-2}$ -verteilt.
$SQE = \sum_{i=1}^n (F(x_i) - \bar{Y})^2$ (Zufallsvariable „erklärte Streuung“)
$SQR = \sum_{i=1}^n (Y_i - F(x_i))^2$ (Zufallsvariable „restliche Streuung“)
$n = \text{Anzahl Datenpaare } (x_i, Y_i)$
$(x_1, \dots, x_n)$ Werte der erklärenden Variable.
$(Y_1, \dots, Y_n) = \text{Zufallsvektor, der die Werte } (y_1, \dots, y_n) \text{ annimmt}$
$F(x) = Ax + B$ ist Zufallsgrösse, welche die Regressionsfunktionen
$\hat{f}(x) = \hat{a}x + \hat{b}$ als Werte annimmt.
$\hat{a}$ ist Wert der ZV A und $\hat{b}$ der Zufallsvariable B .
$\bar{Y} = \text{Zufallsvariable „Mittelwert“ der Zufallsvariablen } Y_i$

H_0 : Die Regressionsgerade „erklärt“ keine Varianz (reduziert keine Varianz).

H_A : Die Regressionsgerade „erklärt“ Varianz (reduziert Varianz).

(ist ein einseitiger Test bezüglich der F-Verteilung; ein Wert der unwahrscheinlich nahe bei 0 liegt drückt eine Überanpassung der Daten an die Nullhypothese aus. Andererseits testet das Verfahren negative wie positive Abweichungen von der 0-Hypothese. In diesem Sinne wird manchmal auch von einem zweiseitigen Test gesprochen, wobei wir hier die „Seitigkeit“ mit der Verteilung der Teststatistik definieren und damit ein rechtsseitiger Test vorliegt)

Beispiel 28.2.1. Wir berechnen für die folgenden Beispieldaten der Tabelle 28.2.1 die ANOVA ($\alpha = 0.05$):

x - Einkommen in Tausend pro Monat	y - Ausgaben für Schmuck in Hundert pro Jahr
1	4
2	3
2.5	4
3	4.9
7	7.5
5	5
6	6

Tabelle 28.2.1: Beispieldaten für die Berechnung der linearen Regression; unabhängige Variable: Einkommen; abhängige Variable: Ausgaben für Schmuck

Wir überprüfen die Normalverteilungsvoraussetzung bezüglich der Residuen (s. Abbildung 28.2.1):

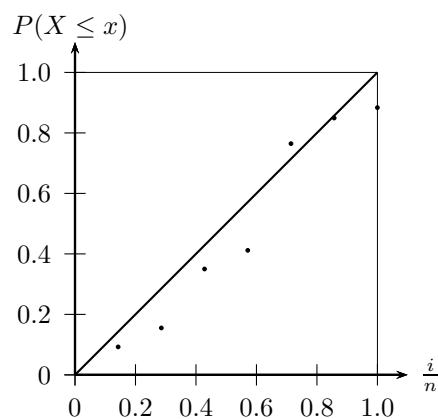


Abbildung 28.2.1: P-P-Plot der Residuen des Beispiels 28.2.1

Bei so wenig Daten ist eine Überprüfung mit Vorsicht zu genießen. Wir akzeptieren die Normalverteilungsannahme knapp. Es hat keine Ausreißer und die Punkte liegen ungefähr auf einer Geraden. Wir berechnen nun (s. Regression.xls oder die Tabelle 28.2.2):

x_i	y_i	$\hat{f}(x) = 0.60048x + 2.641$	$(y_i - \hat{f}(x_i))^2$	$(\hat{f}(x_i) - \bar{y})^2$
1	4	3.24148	0.57535259	2.798278958
2	3	3.84196	0.708896642	1.149882438
2.5	4	4.1422	0.02022084	0.59611635
3	4.9	4.44244	0.209361154	0.222638378
7	7.5	6.84436	0.42986381	3.725186748
5	5	5.6434	0.41396356	0.531607642
6	6	6.24388	0.059477454	1.767820965
Summen	$\bar{y} = 4.9143$		2.41713605 = SQR	10.79153148 =SQE

Tabelle 28.2.2: Berechnung der ANOVA für das Beispiel 28.2.1

Dies ergibt den folgenden Testwert der $F_{1,5}$ -verteilten Teststatistik

$$\frac{SQE}{\frac{1}{n-2}SQR} : \frac{10.7915315}{\frac{1}{5}2.41713605} = 22.32277211$$

und: $F_{F_{1,5}}(22.32277211) = 1 - 0.005220712$

p -Wert: $0.005220712 < 0.05$

Die Regressionsgerade „erklärt“ signifikant viel Varianz.

◇

Gewöhnlich wird die ANOVA in einer Tabelle der folgenden Art entwickelt und dargestellt (s. Tabelle 28.2.3:

	Quadratsummen	df	Varianzen	F	p-Wert
SQE	10.79153148	1	10.79153148	22.32297077	0.005220614
SQR	2.41713605	5	0.48342721		
SQT	13.20857143	6			

Tabelle 28.2.3: Übliche Berechnung und Darstellung der ANOVA; df = Freiheitsgrade; Varianzen = $\frac{\text{Quadratsummen}}{\text{Freiheitsgrade}}$; F = Testwert = $\frac{\text{VarianzE}}{\text{VarianzR}}$

28.2.2 Testtheorie für die Koeffizienten

Ziel: Es soll überprüft werden, ob die Steigung a und der Ordinatenabschnitt b der Regressionsgerade signifikant von 0 verschieden sind.

Voraussetzungen: Die Zufallsvariablen Y_i sind unabhängig. Die Residuen sind identisch normalverteilt und können als Realisierungen unabhängiger Zufallsvariablen mit Erwartungswert 0 betrachtet werden. Es gibt keine Ausreisser und die Punkte (x_i, y_i) liegen ungefähr um eine gedachte Gerade.

Ein weiteres informatives Testverfahren für die Regression liefert uns Vertrauensintervalle und Testresultate bezüglich der Koeffizienten. Sind die Y_i -Daten Werte von Zufallsvariablen, sind die konkret berechneten Koeffizienten $\hat{a}$ und $\hat{b}$ als Werte zweier Zufallsvariablen A und B zu betrachten (siehe Schätztheorie).

Wir wenden uns zuerst dem **Koeffizienten** a zu: Die Varianz von A ist

$$V(A) = \frac{V(\mathcal{E}_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

mit den Fehlern $\mathcal{E}_i = Y_i - b - ax_i$. Zudem ist A bei normalverteilten $\mathcal{E}_i$ normalverteilt: $A \sim N(a, V(A))$ (für einen Beweis s. im letzten Unterkapitel dieses Kapitels).

$$\frac{1}{n-2} \sum_{i=1}^n R_i^2$$

ist ein erwartungstreuer Schätzer für $V(\mathcal{E}_i)$ mit $R_i = Y_i - \hat{b} - \hat{a}x_i$. R_i sind Zufallsvariablen, welche die konkreten Residuen r_i als Werte annehmen. Beachten Sie, dass mit $\frac{1}{n-2}$ statt mit $\frac{1}{n-1}$ multipliziert wird. Man erhält damit, wie wir ohne Beweis festhalten:

$\frac{A}{\sqrt{V(A)}}$ ist t_{n-2} -verteilt.
A : Zufallsvariable mit der Steigung $\hat{a}$ als Wert
$V(A) = \frac{V(\mathcal{E}_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}$; $V(A)$ = Varianz der Zufallsvariable A
$V(\mathcal{E}_i)$ geschätzt durch $\frac{1}{n-2} \sum_{i=1}^n R_i^2$ (mit $R_i := Y_i - F(x_i) = Y_i - B - Ax_i$)
$\widehat{V(A)} = \frac{\frac{1}{n-2} \sum_{i=1}^n R_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$
B : Zufallsvariable mit der Steigung $\hat{b}$ als Wert
n : Anzahl Paare (x_i, Y_i)
$F(x) = Ax + B$ ist Zufallsgrösse, welche $\hat{f}(x) = \hat{a}x + \hat{b}$ annimmt.

Die Nullhypothese ist bei uns jeweils $a_0 = 0$. Die Alternativhypothese ist $a > 0$, $a < 0$ oder $a \neq 0$, je nach zu testendem Zusammenhang zwischen den Variablen (proportionaler Zusammenhang, umgekehrt proportionaler Zusammenhang, unspezifizierter Zusammenhang).

Beispiel 28.2.2. Wir berechnen den entsprechenden Test für das Beispiel 28.2.1 (Regression.xls): Wir überprüfen zuerst die Voraussetzungen: Die Normalverteilung der Residuen wurde bereits überprüft. Eine Überprüfung des bereits gelieferten Streudiagrammes zeigt, dass die Gleichheit der Varianzen der Residuen haltbar ist (die Residuen sind ja identisch verteilt! Entsprechend müssen die Varianzen identisch sein). Es hat keine Ausreisser (als Kriterium für Ausreisser kann man den Einfluss eines Datums auf die Koeffizienten a und b betrachten. Es gibt Verfahren, um diesen Einfluss zu begutachten. Wir behandeln diese Verfahren nicht) und die Punkte liegen ungefähr auf einer Geraden. Wir können somit zum Test übergehen:

$$H_0 : a_0 = 0$$

$$H_A : a > 0 \quad (\implies \text{einseitiger Test, } \alpha = 0.01)$$

$\frac{1}{n-2} \sum_{i=1}^n R_i^2$ nimmt den Wert $\frac{1}{n-2} \sum_{i=1}^n r_i^2 = 0.483427208$ an. Wir können dies mit Excel berechnen, indem wir auf die Residuen r_i die normale Varianz berechnen. Dann multiplizieren wir diese mit $\frac{n-1}{n-2}$:

$$\text{Wert von } \sum_{i=1}^n (x_i - \bar{x})^2 : 29.92857143$$

$$\text{Wert der Zufallsvariable } \widehat{V(A)} : \frac{0.483427208}{29.92857143} = 0.016152699$$

$$\text{Wert der Zufallsvariable } \sqrt{\widehat{V(A)}} : 0.127093269$$

$$\text{Teststatistik } \frac{A}{\sqrt{\widehat{V(A)}}} \text{ nimmt den Testwert } \frac{0.600477327}{0.127093269} = 4.724698097 \text{ an.}$$

$F_{t_5}(4.724698097) = 1 - 0.002610356$ ($n = 7 = \text{Anzahl Daten}$; $df = n - 2$)
 $p\text{-Wert}: 0.002610356 < 0.01$. ◇

Vertrauensintervalle können mit Hilfe der folgenden Formel berechnet werden (wird aus der oben gegebenen Formel für die Verteilung der Teststatistik hergeleitet):

$A_g = A \pm F_{t(n-2)}^{-1}(\beta + \frac{1-\beta}{2}) \sqrt{\widehat{V(A)}}$
$\beta = \text{Wahrscheinlichkeiten, dass faktisches } a \text{ im Vertrauensintervall ist}$

Beispiel 28.2.3. Im Beispiel 28.2.1 erhalten wir für ein 95%-Vertrauensintervall: $F_{t(5)}^{-1}(0.975) = 2.570577635$ (mit Excel: $=t.inv(0.975;5)$).

Wir erhalten somit: $\hat{a}_g = 0.600477327 \pm (2.570577635 \cdot 0.127093269) :]0.273\,72, 0.927\,23[$ ◇

Wir wenden uns nun dem **Koeffizienten** b zu.

$\frac{B}{\sqrt{\widehat{V(B)}}}$ ist t_{n-2} -verteilt.
B : Zufallsvariable mit dem Ordinatenabschnitt $\hat{b}$ als Wert
$V(B) = \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) V(\mathcal{E}_i)$ ($V(B)$ = Varianz der Zufallsvariable B)
$\hat{f}$: geschätzte Regressionsgerade
$V(\mathcal{E}_i)$ geschätzt durch $\frac{1}{n-2} \sum_{i=1}^n R_i^2$ (mit $R_i = Y_i - F(x_i) = Y_i - B - Ax_i$)
$\widehat{V(B)} = \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \frac{1}{n-2} \sum_{i=1}^n R_i^2$
n : Anzahl Paare (x_i, Y_i)
$F(x) = Ax + B$ ist Zufallsgrösse, welche $\hat{f}(x) = \hat{a}x + \hat{b}$ annimmt.

Gewöhnlich geht man von $H_0 : b_0 = 0$ aus. $H_A : b < 0, b > 0$ oder $b \neq 0$

Es handelt sich um einen links, rechts- oder zweiseitigen Test.

Beispiel 28.2.4. Wir betrachten erneut das Beispiel 28.2.1 (siehe für konkrete Berechnungen *Regression.xls*) mit

$H_0 : b_0 = 0$ und

$H_A : b > 0$ ($\implies$ rechtsseitiger Test, $\alpha = 0.01$)

$\frac{1}{n-2} \sum_{i=1}^n R_i^2$ nimmt den Wert 0.483427208 an.

Wert von $\bar{x}^2 : 3.785714286^2 = 14.33163266$

Wert von $\sum_{i=1}^n (x_i - \bar{x})^2 : 29.92857143$

Wert der Zufallsvariable $\widehat{V(B)} : \left(\frac{1}{7} + \frac{14.33163266}{29.92857143} \right) 0.483427208 = 0.300555579$

Dies ergibt den Wert der Zufallsvariable „Standardfehler“: $\sqrt{0.300\,56} = 0.548229495$

Die Teststatistik $\frac{B}{\sqrt{\widehat{V(B)}}}$ nimmt im Beispiel den Testwert $\frac{2.641050119}{0.548229495} = 4.817417056$ an.

$F_{t_5}(4.817417056) = 1 - 0.002404672$

$p\text{-Wert}: 0.002404672 < 0.01$

Die Schnittstelle der Ordinate der Regressionsgeraden unterscheidet sich signifikant von 0. ◇

Für das **Vertrauensintervall** gilt:

$B_g = B \pm F_{t(n-2)}^{-1}(\beta + \frac{1-\beta}{2}) \sqrt{\widehat{V(B)}}$
$\beta = \text{Wahrscheinlichkeit, dass faktisches } b \text{ im Vertrauensintervall ist}$

Beispiel 28.2.5. Im Beispiel 28.2.1 ergibt dies $2.641050119 \pm 2.570577635 \cdot 0.548229495 \cdot [1.2318, 4.0503]$ (für die Berechnung s. Regression.xls). $\diamond$

28.2.3 Schätzen von Werten und deren Vertrauensintervallen

Mit Hilfe der geschätzten Regressionsgeraden $\hat{f}$ können $\hat{f}(x)$ -Werte für beliebige x geschätzt werden. Dabei ist $\hat{f}(x)$ ein Schätzwert für $E(Y|x)$, dem Erwartungswert der erklärten Variable an der Stelle x . So kann man z.B. schätzen, wieviel die geschätzten mittleren Schmuckausgaben für ein spezifisches Einkommen betragen. Für solche Schätzungen können zudem Vertrauensintervalle angegeben werden. Ein entsprechendes β -Vertrauensintervall ist ein Intervall, in dem $E(Y|x)$ mit einer Wahrscheinlichkeit von β zu liegen kommt. Man kann die folgende Formel für Vertrauensintervalle herleiten:

$$\hat{f}(x) \pm \hat{\sigma}(x) F_{t_{(n-2)}} \left(\beta + \frac{1-\beta}{2} \right)$$

wobei

$$\hat{\sigma}(x) = \hat{\sigma} \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

und $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n r_i^2$ (geschätzte Varianz der Fehler $Y_i - f(x_i)$), d.h.

$$\hat{\sigma}^2 := \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{f}(x_i))^2.$$

(Achtung: $\hat{\sigma}(x)$ ist der Standardfehler der Schätzwerte $\hat{f}(x)$ an der Stelle x , $\hat{\sigma}$ ist der Standardfehler der Residuen).

Für das obige Beispiel und für $x = 6.2$ erhält man den geschätzten Wert

$$\begin{aligned} \hat{f}(6.2) &= 0.600477327 \cdot 6.2 + 2.641050119 \\ &= 6.364 \end{aligned}$$

Damit ergibt sich mit

$$\frac{1}{n-2} \sum_{i=1}^n r_i^2 = 0.483427208$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 29.92857143 \text{ und}$$

$\bar{x} = 3.785714286$ bei $\beta = 0.95$ (und damit $F_{t_{(7-2)}}(0.975) = 2.570581835$) die

obere Grenze:

$$\hat{f}(x) \pm \hat{\sigma}(x) F_{t_{(n-2)}}^{-1} \left(\beta + \frac{1-\beta}{2} \right)$$

$$\begin{aligned} &6.364 + \sqrt{0.483427208} \sqrt{\frac{1}{7} + \frac{(6.2 - 3.785714286)^2}{29.92857143}} \cdot 2.570581835 \\ &= 7.4025 \end{aligned}$$

untere Grenze:

$$\begin{aligned} &6.364 - \sqrt{0.483427208} \sqrt{\frac{1}{7} + \frac{(6.2 - 3.785714286)^2}{29.92857143}} \cdot 2.570581835 \\ &= 5.3255 \end{aligned}$$

Mit einer Wahrscheinlichkeit von 95% befindet sich der tatsächliche Wert $E(Y|x)$ an der Stelle $x = 6.2$ zwischen 5.3255 und 7.4025. An der Formel sieht man, dass die Vertrauensintervalle von der Lage der x abhängen. Die Varianz des Schätzers von $E(Y|x)$ an der Stelle x hängt insbesondere von $(x - \bar{x})^2$ ab. Dieser Ausdruck wird grösser, je mehr x vom Mittelwert der x -Daten $\bar{x}$ abweicht. Man kann die Grenzen der Vertrauensintervalle verschiedener x verbinden, wodurch man auf beiden Seiten der Regressionsgeraden je eine Kurve erhält (s. Abbildung 28.2.2 für ein punktweises 95%-Vertrauensband im Beispiel).

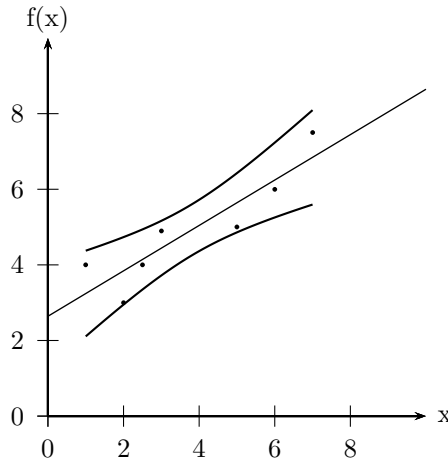


Abbildung 28.2.2: 0.95-Vertrauensintervalle der $E(Y|x)$ als Band um die Regressionsgerade

Neben diesem VI-Band für einzelne Erwartungswerte (Überdeckung des tatsächlichen $E(Y|x)$ durch VI an der Stelle x mit gegebener Wahrscheinlichkeit; Variante 1) gibt es auch die Möglichkeit, ein VI-Band für alle Erwartungswerte auf einmal zu berechnen (simultane Überdeckung von $E(Y|x)$ für alle x mit gegebener Wahrscheinlichkeit; Variante 2):

$$\hat{f}(x) \pm \sqrt{2F_{F_2; (n-2)}^{-1} \left(\beta + \frac{1-\beta}{2} \right) \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$$

Diese VI-Band ist breiter. Bei beiden Varianten liegen viele der Daten (x_i, y_i) im Streudiagramm ausserhalb der VI-Bänder. Möchte man ein VI-Band, das diese Daten mit gegebener Wahrscheinlichkeit enthält, kann man z.B. für eine der beiden obigen Varianten zu den oberen Grenzen den Wert

$$\sqrt{\frac{1}{n-2} \sum_{i=1}^n r_i^2 \cdot t_{1-\frac{\alpha}{2}}^{-1}}$$

hinzuzählen und von den unteren Grenzen diesen Wert abzuziehen ($\alpha = 1 - \beta$).

28.2.4 Prüfung der Voraussetzungen

Bezüglich Prüfung der Voraussetzungen werden in der Literatur viele Methoden diskutiert und etliche davon werden von Statistik-Programmen angeboten. Es folgt eine kurze Auswahl solcher Verfahren:

Die Prüfung auf Normalverteiltheit der Residuen ist einfach: man macht einen Q-Q- oder P-P-Plot auf die Residuen - diese werden in R mittels `modell$residuals` geliefert (für `modell` ist der Modellname zu setzen, unter dem das Ergebnis des `lm`-Befehls abgespeichert ist).

Um die Varianzgleichheit zu überprüfen (die Residuen sind als identisch, unabhängig normalverteilt vorausgesetzt, müssen also überall dieselbe Varianz aufweisen), kann man die Residuen gegen die geschätzten Werte $\hat{y}_i$ abtragen. Man sieht von Auge, ob die Streuung in etwa konstant ist (bis auf

wenige Ausreisser) oder ob sich tendenzartige Abweichungen von der Voraussetzung ergeben, z.B. eine Vergrößerung der Streuung von links nach rechts. Oft wird in dieses Diagramm eine Kurve an die Residuen angepasst - mittels Methoden, die hier nicht eingeführt werden können und die in den Bereich nicht-parametrischer Regression gehören (z.B. Spline-Regression). Das Diagramm mit dieser Kurve wird „Tukey-Anscombe-Diagramm“ genannt. Man sieht damit besser, ob die Residuen wirklich gleichmässig um die 0-Linie streuen oder nicht. Es kann sich auch ein Hinweis auf die fehlende Unabhängigkeit der Residuen ergeben, wenn die Abweichungen systematisch sind. Bei manchen Graphiken spielen die sogenannten Hebel (engl. leverage) eine wichtige Rolle. Die Hebelzahlen geben an, welchen Einfluss ein einzelnes x_i -Datum auf die Regression haben könnte - je weiter weg ein x_i Datum vom Mittelwert $\bar{x}$ liegt, desto grösser ist der potentielle Einfluss des entsprechenden y_i -Datums auf die Regressionsgerade. In der bivariaten linearen Regression sind die Hebel definiert als:

$$H_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

Sie streuen zwischen 0 und 1 und deren Summe ist 2 (in R mit `hatvalues(modell)`). Mit den Hebeln werden sogenannte standardisierte Residuen berechnet:

$$\tilde{r}_i = \frac{r_i}{\hat{\sigma}\sqrt{1-H_i}}$$

mit der geschätzten Varianz $\hat{\sigma}$ der Fehlerzufallsvariablen $\mathcal{E}_i$. Es werden dann in Diagrammen z.B. die mit $\frac{1}{2}$ potenzierten standardisierten Residuen gegen die die geschätzten Werte $\hat{y}_i$ abgetragen (mit nicht-parametrischer Regressionskurve versehen) oder die standardisierten Residuen gegen die Hebel. In letzterer Graphik werden von R noch Grenzen für die Distanzen von Cook (± 0.5 und ± 1) eingetragen: Die Distanz von Cook fasst die Veränderungen aller angepassten Werte $\hat{y}_i$ beim Weglassen der i -ten Beobachtung zu einer Zahl zusammen: Sie lässt sich schreiben als

$$d_i^{(C)} = \frac{R_i^2 H_i}{2\hat{\sigma}^2 (1-H_i)^2}$$

(R_i ist das i -te Residuum, $\hat{\sigma}$ die geschätzte Varianz der $\mathcal{E}_i$; in R mit `cooks.distance(modell)`; in der Literatur werden unterschiedliche Kriterien für einflussreiche Werte geliefert: z.B. Werte grösser als 1 oder Werte grösser als $4/n$ - n Anzahl Daten). In R werden die eben beschriebenen Graphiken mit `plot(modell)` geliefert. Will man alle Graphiken auf einmal erhalten, gibt man vor diesem Befehl `par(mfrow=c(2,2))` ein (dieser Befehl muss nachher wieder rückgängig gemacht werden mit `par(mfrow=c(1,1))`).

Interessant kann schliesslich der Befehl `dbeta(modell)` sein. Er berechnet die Differenz der Schätzer für die Regression beim Fehlen des i -te Datum zu den Schätzern im Modell mit allen Daten. Man sieht, welche Daten einflussreich sind (einen grossen Einfluss auf die Grösse der Koeffizienten haben).

Die Begutachtung und Bewertung der Graphiken ist nicht streng mathematisch zu begründen und erfordert Erfahrung und Augenmass. Im Allgemeinen wird es genügen, wenn man die ersten zwei der erwähnten Graphiken genauer anschaut. Sie sind zu Beginn intuitiv wohl am einsichtigsten.

28.2.5 Einige Beweise (fakultativ)

Satz 28.2.6. Für n Paare (x_i, Y_i) mit $E(Y_i) = ax_i + b$ gilt mit

$$A = \frac{\sum_{i=1}^n x_i Y_i - n\bar{x}\bar{Y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

$E(A) = a$ (Die Statistik A ist ein erwartungstreuer Schätzer für a).

Beweis. Mit Satz 5.3.16 auf Seite 117 ($a = r \frac{s_y}{s_x} = \frac{s_{xy}}{s_x^2}$) gilt (wir schreiben $\sum$ statt $\sum_{i=1}^n$)

$$A = \frac{\frac{1}{n-1} \sum (x_i - \bar{x}) (Y_i - \bar{Y})}{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x}) (Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$$

Mit $\tilde{x}_i := \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$ gilt (s. S. Bemerkung 4.1.9.4 auf Seite 47: $\sum (x_i - \bar{x}) = 0$)

$$\sum \tilde{x}_i = \sum \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = 0$$

und damit

$$\begin{aligned} A &= \sum \tilde{x}_i (Y_i - \bar{Y}) \\ &= \sum \tilde{x}_i Y_i - \bar{Y} \sum \tilde{x}_i \\ &= \sum \tilde{x}_i Y_i. \end{aligned}$$

Damit ist mit $E(Y_i) = b + ax_i$ und mit Satz 9.5.11 von Seite 217 ($E(\sum Y_i) = \sum E(Y_i)$)

$$\begin{aligned} E(A) &= E\left(\sum \tilde{x}_i Y_i\right) \\ &= \sum \tilde{x}_i E(Y_i) \\ &= \sum \tilde{x}_i (b + ax_i) \\ &= b \sum \tilde{x}_i + a \sum \tilde{x}_i x_i \\ &= a \sum \tilde{x}_i x_i \\ &= a, \end{aligned}$$

da mit $\sum \tilde{x}_i \bar{x} = \bar{x} \sum \tilde{x}_i = 0$ gilt

$$\sum \tilde{x}_i x_i = \sum \tilde{x}_i x_i - \sum \tilde{x}_i \bar{x} = \sum \tilde{x}_i (x_i - \bar{x})$$

und damit

$$\sum \tilde{x}_i x_i = \sum \frac{(x_i - \bar{x})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} = 1.$$

□

Satz 28.2.7. (i) $V(A) = \frac{V(\mathcal{E}_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

(ii) A ist normalverteilt: $A \sim N(a, V(A))$.

Beweis. Gemäss vorangegangenen Beweis (s. auch für Definition von $\tilde{x}_i$) und wegen der Unabhängigkeit der Y_i gilt mit Satz 9.5.21 auf S. 9.5.21 und mit Satz 9.5.35 auf S. 223:

$$V(A) = V\left(\sum \tilde{x}_i Y_i\right) = \sum \tilde{x}_i^2 V(Y_i)$$

Zudem gilt ja $V(Y_i) = V(a + bx_i + \mathcal{E}_i) = V(\mathcal{E}_i)$. Damit erhält man unter der Voraussetzung der

konstanten Varianz der Fehler

$$\begin{aligned}
 V(A) &= \sum \tilde{x}_i^2 V(Y_i) \\
 &= V(Y_i) \sum \tilde{x}_i^2 \\
 &= V(\mathcal{E}_i) \sum \left(\frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right)^2 \\
 &= V(\mathcal{E}_i) \frac{\sum (x_i - \bar{x})^2}{\left(\sum (x_i - \bar{x})^2 \right)^2} \\
 &= \frac{V(\mathcal{E}_i)}{\sum (x_i - \bar{x})^2}
 \end{aligned}$$

Da die Fehler gemäss Voraussetzung normalverteilt sind, sind es auch die Zufallsvariablen Y_i . Da $A = \sum \tilde{x}_i Y_i$, ist A eine Summe von normalverteilten Zufallsvariablen und entsprechend auch normalverteilt mit Erwartungswert a und Varianz $V(A)$. $\square$

Satz 28.2.8. Für n Paare (x_i, Y_i) mit $E(Y_i) = ax_i + b$ gilt mit

$$B = \bar{Y} - A\bar{x}$$

und A definiert wie oben:

(i) $E(B) = b$ (Die Statistik B ist ein erwartungstreuer Schätzer für b).

(ii) $V(B) = V(\mathcal{E}_i) \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right)$

(iii) B ist normalverteilt: $B \sim N(b, V(B))$

Beweis. (i) Wir setzen

$$\gamma := b + a\bar{x}$$

Damit ist

$$b = \gamma - a\bar{x}$$

und das Regressionsmodell wird zu

$$Y_i = b + ax_i + \mathcal{E}_i = \gamma - a\bar{x} + ax_i + \mathcal{E}_i = \gamma + a(x_i - \bar{x}) + \mathcal{E}_i$$

γ wird durch die Zufallsvariable $\Gamma := B + A\bar{x}$ geschätzt. Es gilt gemäss $B = \bar{Y} - A\bar{x}$: $B + A\bar{x} = \bar{Y} = \Gamma$. Damit gilt:

$$\begin{aligned}
 E(\Gamma) &= E(\bar{Y}) = \frac{1}{n} \sum E(Y_i) = \frac{1}{n} \sum E(\gamma + a(x_i - \bar{x}) + \mathcal{E}_i) \\
 &= \frac{1}{n} \sum (\gamma + a(x_i - \bar{x})) = \frac{1}{n} \left(n\gamma + a \sum (x_i - \bar{x}) \right) \\
 &= \gamma + a \frac{1}{n} \sum (x_i - \bar{x}) \\
 &= \gamma.
 \end{aligned}$$

Damit ist Γ eine erwartungstreuer Schätzer für γ und es gilt

$$\begin{aligned}
 E(B) &= E(\Gamma - A\bar{x}) = E(\Gamma) - \bar{x}E(A) \\
 &= \gamma - \bar{x}a \\
 &= b.
 \end{aligned}$$

(ii)

$$V(\Gamma) = V(\bar{Y}) = \frac{1}{n^2} \sum V(Y_i) = \frac{1}{n^2} n V(Y_i) = \frac{1}{n} V(\mathcal{E}_i).$$

Damit ist $\Gamma \sim N(\gamma, V(\Gamma))$, denn Γ ist eine Summe von normalverteilten Zufallsvariablen. Es gilt

$$Kov(\Gamma, A) = 0$$

denn

$$\begin{aligned} A &= \sum \tilde{x}_i Y_i = \sum \tilde{x}_i (b + ax_i + \mathcal{E}_i) \\ &= \sum \tilde{x}_i b + a \sum \tilde{x}_i x_i + \sum \tilde{x}_i \mathcal{E}_i \\ &= 0 + a + \sum \tilde{x}_i \mathcal{E}_i \end{aligned}$$

und

$$\begin{aligned} \Gamma = \bar{Y} &= \frac{1}{n} \sum Y_i = \frac{1}{n} \sum (\gamma + a(x_i - \bar{x}) + \mathcal{E}_i) \\ &= \frac{1}{n} n\gamma + a \frac{1}{n} \sum (x_i - \bar{x}) + \frac{1}{n} \sum \mathcal{E}_i \\ &= \gamma + 0 + \frac{1}{n} \sum \mathcal{E}_i \end{aligned}$$

Damit gilt

$$\begin{aligned} Kov(\Gamma, A) &= E((\Gamma - \gamma)(A - a)) \\ &= E\left(\frac{1}{n} \sum \mathcal{E}_i \cdot \sum \tilde{x}_i \mathcal{E}_i\right) \\ &= \frac{1}{n} E\left(\sum \mathcal{E}_i \cdot \sum \tilde{x}_i \mathcal{E}_i\right) \\ &= \frac{1}{n} E\left(\sum \tilde{x}_i \mathcal{E}_i^2 + \sum_{i=1}^n \tilde{x}_i \sum_{j \neq i} \mathcal{E}_i \mathcal{E}_j\right) \\ &= \frac{1}{n} \left(\sum \tilde{x}_i E(\mathcal{E}_i)^2 + \sum_{i=1}^n \tilde{x}_i \sum_{j \neq i} E(\mathcal{E}_i \mathcal{E}_j) \right) \\ &= 0 \end{aligned}$$

Da die $\mathcal{E}_j$ unabhängig sind, gilt nämlich $E(\mathcal{E}_i \mathcal{E}_j) = E(\mathcal{E}_i) E(\mathcal{E}_j) = 0 \cdot 0 = 0$ und wegen $\sum \tilde{x}_i = 0$ ist $\sum \tilde{x}_i E(\mathcal{E}_i)^2 = E(\mathcal{E}_i)^2 \sum \tilde{x}_i = 0$. Damit gilt dann

$$\begin{aligned} V(B) &= V(\Gamma - A\bar{x}) = V(\Gamma) + \bar{x}^2 V(A) - 2Kov(\Gamma, A) \\ &= V(\Gamma) + \bar{x}^2 V(A) \\ &= \frac{1}{n} V(\mathcal{E}_i) + \bar{x}^2 \frac{V(\mathcal{E}_i)}{\sum (x_i - \bar{x})^2} \\ &= V(\mathcal{E}_i) \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right) \end{aligned}$$

(iii) B ist normalverteilt: $B \sim N(b, V(B))$, denn B ist eine Summe von normalverteilten Zufallsvariablen. $\square$

28.2.6 Übungen

1. Berechnen Sie für das Alter-Autounfallprämien-Beispiel die Regressionsgerade (mit Excel von Hand, mit dem Excel-Spezial-Befehl und mit einer Statistiksoftware). Berechnen Sie die ANOVA für dieses Beispiel mit Excel und einer Statistik-Software. Berechnen Sie die

T-Teststatistik für die Koeffizienten und 95%-Vertrauensintervalle mit Excel und mit einer Statistiksoftware. (H_0 : Das Alter hat keinen Einfluss auf die Prämien; H_A : Das Alter hat einen umgekehrt proportionalen Einfluss auf die Prämienhöhe ($\Rightarrow$ linksseitiger Test, $\alpha = 0.05$))

Für b teste man die Hypothese, dass $b > 0$ ($\Rightarrow$ rechtsseitiger Test, $\alpha = 0.05$).

2. Berechnen Sie für die Daten 2 Regression.xls die Regressionsgerade, die ANOVA und die Teststatistik für die Koeffizienten sowie die Vertrauensintervalle (mit Excel und mit einer Statistiksoftware). (H_0 : Die Verkaufsfläche hat keinen Einfluss auf den Umsatz; H_A : Die Verkaufsfläche hat einen proportionalen Einfluss auf den Umsatz ($\Rightarrow$ rechtsseitiger Test, $\alpha = 0.05$))

Für b teste man die Hypothese, dass $b > 0$ ($\Rightarrow$ rechtsseitiger Test, $\alpha = 0.05$).

3. Lösen Sie die Übung 3 von Regression.xls.

28.2.7 Lösungen

1. Wir überprüfen die Voraussetzungen der Berechenbarkeit einer Regressionsgerade: Linearität der Punktwolke und Punktsymmetrie (d.h. identische Varianz und keine Ausreisser) (s. Abbildung 28.2.3):

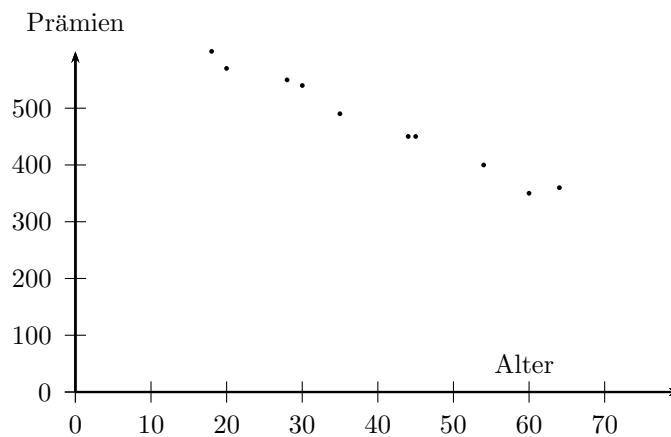


Abbildung 28.2.3: Zweidimensionales Streudiagramm der Daten der Übung

Die Bedingungen können als erfüllt betrachtet werden. Wir berechnen also die Koeffizienten der Gerade (s. Tabelle 28.2.4):

Objekt	Alter	Prämien	$x_i(y_i - \bar{y})$	$x_i(x_i - \bar{x})$
1	60	350	-7560	1212
2	30	540	1920	-294
3	20	570	1880	-396
4	35	490	490	-168
5	45	450	-1170	234
6	44	450	-1144	184.8
7	28	550	2072	-330.4
8	64	360	-7424	1548.8
9	18	600	2232	-392.4
10	54	400	-4104	766.8
Mittelwert	39.8	476		
Standardabweichung	16.21247805	88.46845012		
Summen			-12808	2365.6

Tabelle 28.2.4: Tabelle für die Berechnung der Koeffizienten der Regressionsgerade

$$\hat{a} = \frac{-12808}{2365.6} = -5.414271221$$

$$\hat{b} = 476 - (-5.414271221 \cdot 39.8) = 691.4879946$$

und damit die Gerade:

$$y = -5.414271221x + 691.4879946.$$

Die Bedingung für die Anwendung der Tests besteht in der Normalverteiltheit der Residuen und identischen Varianzen der Residuen. Wir überprüfen die Normalverteiltheit der Residuen mit Hilfe eines P-P-Plots (s. Abbildungen 28.2.4):

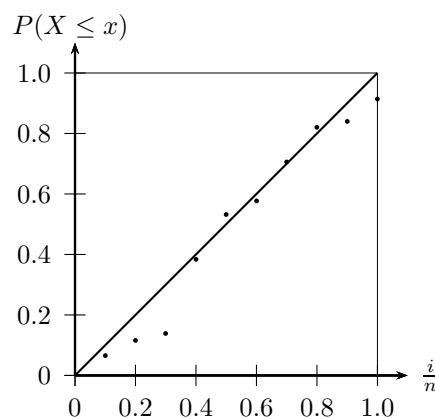


Abbildung 28.2.4: P-P-Plot der Residuen

Wir betrachten die (wenigen) Residuen als (knapp) normalverteilt. Wir überprüfen die Gleichheit der Varianzen mit Hilfe des Streudiagramms mit eingezeichneter Regressionsgerade (s. Abbildung 28.2.5):

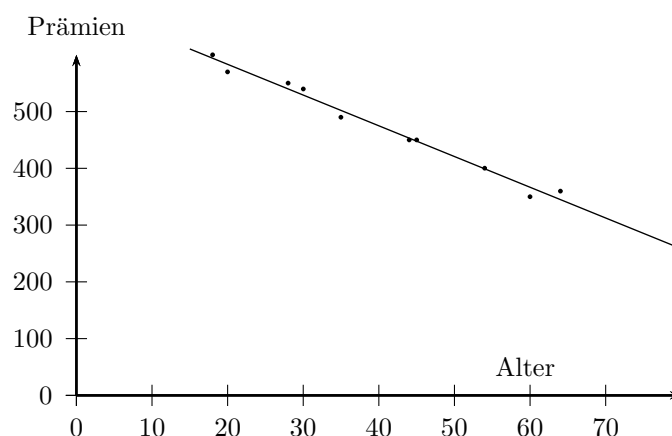


Abbildung 28.2.5: Zweidimensionales Streudiagramm der Daten der Übung mit eingezeichneter geschätzter Regressionsgerade

Die Gleichheit der Varianzen in verschiedenen x ist haltbar. Wir berechnen die ANOVA und dazu zuerst die Quadratsummen (s. Tabelle 28.2.5):

Objekt	Alter	Prämien	$\hat{f}(x)$	$(\hat{f}(x_i) - \bar{y})^2$	$(y_i - \hat{f}(x_i))^2$
1	60	350	366.6317213	11961.42038	276.6141547
2	30	540	529.059858	2815.348528	119.6867076
3	20	570	583.2025702	11492.39105	174.3078594
4	35	490	501.9885019	675.4022292	143.724177
5	45	450	447.8457897	792.6595602	4.640622211
6	44	450	453.2600609	517.1048314	10.62799692
7	28	550	539.8884004	4081.727707	102.2444462
8	64	360	344.9746365	17167.64589	225.7615496
9	18	600	594.0311126	13931.34355	35.62761653
10	54	400	399.1173487	5910.942076	0.779073377
Mittelwert	39.8	476			
Standardabweichung	16.21247805	88.46845012		SQE=	SQR=
Summen				69345.9858	1094.014204

Tabelle 28.2.5: Tabelle für die Berechnung der ANOVA

Mit Excel siehe regression.xls.

Testwert: $\frac{69345.9858}{\frac{1094.014204}{8}} = 507.0938609$

$F_{F_{1,8}}(507.0938609) = 1 - 1.60099 \cdot 10^{-8}$

p-Wert: $1.60099 \cdot 10^{-8} < 0.05$

Wert der Zufallsvariable $\frac{1}{n-2} \sum_{i=1}^N R_i^2$: 136.7517754

Wert der Zufallsvariable $V(A)$: $\frac{136.7517754}{2365.6} = 0.057808495$

Wert der Zufallsvariable Standardfehler von A: 0.240433972

Testwert: $\frac{-5.414271221}{0.240433972} = -22.519$

$F_{ts}(-22.519) = 8.00495E-09 < 0.05$

Wert der Zufallsvariable $V(B)$: $\left(\frac{1}{10} + \frac{39.8^2}{2365.6}\right) 136.7517754 = 105.2461457$

Wert der Zufallsvariable Standardfehler von B: 10.25895442

Testwert: $\frac{691.4879946}{10.25895442} = 67.40335968$

$F_{t_8}(67.40335968) = 1 - 1.30613E - 12$

p-Wert: $1.30613E - 12 < 0.05$.

Vertrauensintervalle: $t_8^{-1}(0.975) = 2.306005626$

Für a : $-5.414271221 \pm 0.240433972 \cdot 2.306005626 =] - 5.968713313, -4.859829129[$

Für b : $691.4879946 \pm 10.25895442 \cdot 2.306005626 =]667.830788, 715.1452012[$

2. Wir überprüfen zuerst die Voraussetzungen der Berechnung einer Regressionsgerade (Linearität; Punktsymmetrie; s. Abbildung 28.2.6):

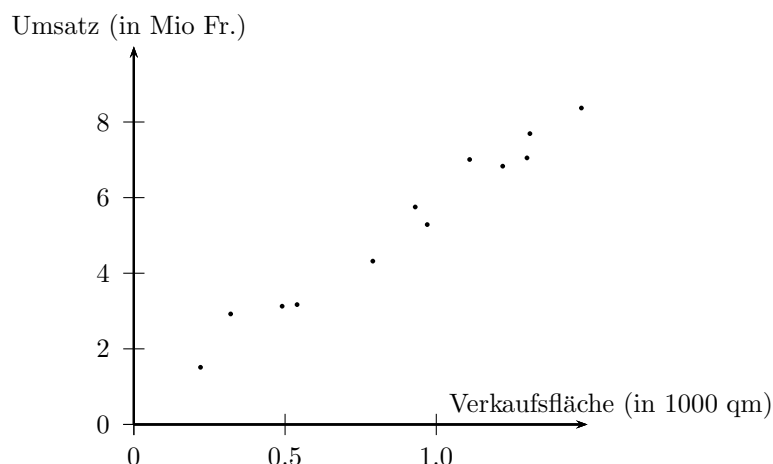


Abbildung 28.2.6: Streudiagramm der Daten

Die Voraussetzungen sind erfüllt. Wir berechnen die Koeffizienten (s. Excel-Tabelle) und überprüfen anschliessend die Voraussetzungen für die Tests - Normalverteilung der Residuen (s. Abbildung 28.2.7) und Gleichheit der Varianzen (s. Abbildung 28.2.8):

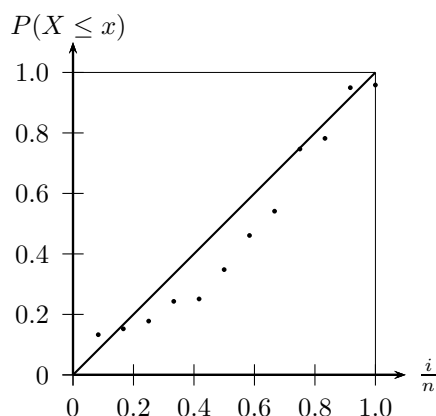


Abbildung 28.2.7: P-P-Plot der Residuen

Die Normalverteilungsannahme ist kaum haltbar - allerdings hat es zur Beurteilung zu wenig Daten. Übungshalber fahren wir mit der Überprüfung der Varianzgleichheit fort: s. Abbildung 28.2.8):

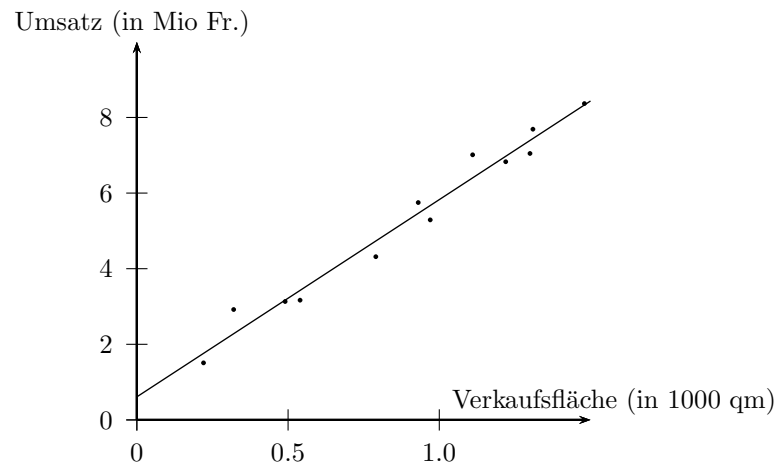


Abbildung 28.2.8: Streudiagramm der Daten mit eingezeichneter Regressionsgerade

Die Varianzgleichheit ist haltbar. Für den Rest der Lösung, s. Exceltabelle Regression.xls.

28.3 Lernziele

- Wissen, dass die Steigung und der Ordinatenabschnitt erwartungstreu und konsistent geschätzt werden können (wissen, was das bedeutet)
- Die ANOVA für die Regression berechnen können und diese korrekt interpretieren können.
- Die Regressionsgerade berechnen können. Die Tests für die Steigung und den Ordinatenabschnitt berechnen und interpretieren können.
- Vertrauensintervalle für die Koeffizienten der Regressionsgerade berechnen und interpretieren können.
- Aufgaben von der Art der Übungen lösen können

Teil VI

Multivariate schliessende Statistik

Bei der multivariaten schliessenden Statistik geht es darum, den „Einfluss“ von mehreren „unabhängigen“ Variable auf eine „abhängige“ Variable oder auf mehrere Variablen zu erfassen. Der Vorteil der multivariaten Methoden besteht darin, dass man weniger schnell Opfer von Scheinerklärungen wird. Dies kann ein Beispiel erläutern. Für das Schmuckkaufverhalten ist nicht nur das Einkommen wichtig, sondern wohl auch das Geschlecht, die Ausbildung, der kulturelle Hintergrund, die Schichtzugehörigkeit, etc. Manche dieser „erklärenden“ Variablen beeinflussen sich gegenseitig oder eine die andere. So hat Schichtzugehörigkeit etwas mit Ausbildung und mit Einkommen zu tun. Es stellt sich nun die Frage, ob die Schichtzugehörigkeit mehr Varianz beim Schmuckkauf wegerklärt als die Ausbildung oder das Einkommen oder ob andere Einflussfaktoren entscheidender sind. Bei der bivariaten Analyse können wir unter Umständen für jeden dieser Faktoren einen signifikanten „Einfluss“ feststellen. Berücksichtigen wir jedoch mehrere Variablen, so kann der „Einfluss“ einer Variablen, die bei bivariater Betrachtung noch einen signifikanten „Einfluss“ hatte, verschwinden (sehen Sie die ähnlich gelagerten Betrachtungen bezüglich der partiellen Korrelation).

In der Folge geht es nicht um eine detaillierte Behandlung der multivariaten Statistik, die ein eigenes Semester oder gar ein eigenes Studienjahr (à 2 Wochenstunden) verdienen würde. Es wird nur eine Methoden vorgestellt, die es erlauben sollen, sich eine Vorstellung von multivariater Statistik zu bilden.

Kapitel 29

Multiple Regression

29.1 Vorbemerkungen zu Matrizen und Vektoren (Wiederholung)

Eine $n \times m$ -Matrix ist umgangssprachlich eine Tabelle mit n Zeilen und m Spalten. Diese weist in den Zellen i, j (= Schnittstelle der i -ten Zeile und der j -ten Spalte) jeweils Zahlen auf, die wir mit a_{ij} bezeichnen. Mathematisch handelt es sich bei einer Matrix um eine Minifunktion $\mathbb{N}_n^* \times \mathbb{N}_m^* \rightarrow \mathbb{R}$

Beispiel 29.1.1. $\mathbf{A}_{3,4} = \{(1, 1, 3), (1, 2, 4), (1, 3, 6), (1, 4, 7), (2, 1, 8), (2, 2, 1), (2, 3, 8), (2, 4, 7), (3, 1, 11), (3, 2, 0.5), (3, 3, \sqrt{2}), (3, 4, \frac{3}{8})\}$

Diese Darstellung ist nicht üblich. In der üblichen Tabellenform:

$$\begin{pmatrix} 3 & 4 & 6 & 7 \\ 8 & 1 & 8 & 7 \\ 11 & 0.5 & \sqrt{2} & \frac{3}{8} \end{pmatrix}$$

◇

Man sieht, dass jede statistische Datensammlung in Tabellenform einer Matrix entspricht. Für die Matrizen verwenden wir als Variablen $\mathbf{A}, \mathbf{B}, \mathbf{C}$. Einspaltige Matrizen nennen wir „Vektoren“, da mit ihnen ein sogenannter Vektorraum konstruierbar ist (Es gibt auch Vektorräume, die andere Objekte wie Funktionen oder n -Tupel als Vektoren enthalten!). Wir verwenden kleine Buchstaben $\mathbf{b}, \mathbf{x}, \mathbf{y}$, für sie. $\mathbf{b}^T$ ist die Einzeilenmatrix, welche die Komponenten des Vektors $\mathbf{b}$ enthält. Auf Matrizen werden bestimmte Operationen definiert. Uns interessieren in diesem Zusammenhang nur drei Operationen: die Multiplikation, die Transponierte einer Matrix und die Inverse einer Matrix.

29.1.1 Die Multiplikation von Matrizen

Sei

$$\mathbf{A}_{mn} := (a_{ij})_{mn} := \{(i, j, x) : (i, j) \in \mathbb{N}_m \times \mathbb{N}_n \text{ und } x \in \mathbb{R}\}$$

eine Matrix mit m Zeilen und n Spalten. ($m, n \geq 1$). Sei

$$\mathbf{B}_{nl} := (b_{jk})_{nl}$$

eine Matrix mit n Zeilen und l Spalten. Dann wird $\mathbf{C} = \mathbf{AB}$ definiert mit

$$c_{ik} = \sum_{j=1}^n a_{ij} b_{jk}.$$

(Achtung: die Multiplikation ist nur definiert, wenn **B** gleichviele Zeilen hat wie **A** Spalten hat).

Beispiel 29.1.2. $\mathbf{A} = \begin{pmatrix} 3 & 4 & 6 & 7 \\ 8 & 1 & 8 & 7 \\ 11 & 0.5 & \sqrt{2} & \frac{3}{8} \end{pmatrix}$ ($= 3 \times 4$ -Matrix)

$\mathbf{B} = \begin{pmatrix} 3 & 4 \\ 8 & 1 \\ 11 & 0.5 \\ 3 & 8 \end{pmatrix}$ ($= 4 \times 2$ -Matrix). **C** ist damit eine 3×2 -Matrix.

Damit gilt dann $c_{11} = \sum_{j=1}^4 a_{1j}b_{j1} = 3 \cdot 3 + 4 \cdot 8 + 6 \cdot 11 + 7 \cdot 3 = 128$

$c_{12} = \sum_{j=1}^4 a_{1j}b_{j2} = 3 \cdot 4 + 4 \cdot 1 + 6 \cdot 0.5 + 7 \cdot 8 = 75$

$c_{21} = \sum_{j=1}^4 a_{2j}b_{j1} = 8 \cdot 3 + 1 \cdot 8 + 8 \cdot 11 + 7 \cdot 3 = 141$, etc.

Vervollständigen Sie übungshalber die Matrix **C**.

Lösung: $\mathbf{C} = \begin{pmatrix} 3 & 4 & 6 & 7 \\ 8 & 1 & 8 & 7 \\ 11 & 0.5 & \sqrt{2} & \frac{3}{8} \end{pmatrix} \cdot \begin{pmatrix} 3 & 4 \\ 8 & 1 \\ 11 & 0.5 \\ 3 & 8 \end{pmatrix} = \begin{pmatrix} 128 & 75 \\ 141 & 93 \\ 53.681 & 48.207 \end{pmatrix}$ $\diamond$

Cb mit dem Vektor **b** entspricht der Multiplikation von **C** mit der Einspaltenmatrix, welche die Komponenten von **b** enthält.

29.1.2 Die Transposition von Matrizen.

Diese Operation besteht einfach im Spiegeln einer Matrix an der Diagonalen (an den Zellen a_{ii}). Für die Transposition von **A** schreiben wir $\mathbf{A}^T$.

Beispiel 29.1.3. $\begin{pmatrix} 3 & 4 & 6 \\ 8 & 1 & 8 \\ 11 & 0.5 & \sqrt{2} \end{pmatrix}^T = \begin{pmatrix} 3 & 8 & 11 \\ 4 & 1 & 0.5 \\ 6 & 8 & \sqrt{2} \end{pmatrix}$

$\begin{pmatrix} 3 & 4 & 6 & 7 \\ 8 & 1 & 8 & 7 \\ 11 & 0.5 & \sqrt{2} & \frac{3}{8} \end{pmatrix}^T = \begin{pmatrix} 3 & 8 & 11 \\ 4 & 1 & 0.5 \\ 6 & 8 & \sqrt{2} \\ 7 & 7 & \frac{3}{8} \end{pmatrix}$ $\diamond$

Eine Matrix, die gleich viele Zeilen wie Spalten aufweist, wird „quadratische Matrix“ genannt. Eine quadratische Matrix, die in der Diagonalen nur 1 aufweist und sonst überall 0, nennen wir

„Einheitsmatrix“. Wir kürzen sie mit **E** ab. z.B. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

29.1.3 Die Inverse einer Matrix

Die Inverse $\mathbf{A}^{-1}$ einer Matrix **A** ist definiert als die Matrix, die mit **A** multipliziert eine Einheitsmatrix ergibt: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{E}$. (Achtung. Inversen sind nur für quadratische Matrizen definiert. Zudem müssen sie nicht immer existieren. Wann sie existieren, ist Gegenstand der Linearen Algebra. Die Beantwortung dieser Frage braucht etliches an Theoriebildung!). Die Berechnung ist mühsam und führt auf n lineare Gleichungssysteme mit n Gleichungen (für eine $n \times n$ -Matrix).

Beispiel 29.1.4. Zu berechnen sei: $\begin{pmatrix} 3 & 8 & 11 \\ 4 & 1 & 3 \\ 6 & 8 & 2 \end{pmatrix}^{-1} = \mathbf{A}^{-1} := \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$

Es gilt damit: $\begin{pmatrix} 3 & 8 & 11 \\ 4 & 1 & 3 \\ 6 & 8 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

Wir kürzen die Komponenten von $\mathbf{A}^{-1}$ mit x_{ij} ab (da noch unbekannt).

Wir erhalten die folgenden drei Gleichungssysteme mit je drei unbekannten:

$$3x_{11} + 8x_{21} + 11x_{31} = 1$$

$$4x_{11} + 1x_{21} + 3x_{31} = 0$$

$$6x_{11} + 8x_{21} + 2x_{31} = 0$$

$$3x_{12} + 8x_{22} + 11x_{32} = 0$$

$$4x_{12} + 1x_{22} + 3x_{32} = 1$$

$$6x_{12} + 8x_{22} + 2x_{32} = 0$$

$$3x_{13} + 8x_{23} + 11x_{33} = 0$$

$$4x_{13} + 1x_{23} + 3x_{33} = 0$$

$$6x_{13} + 8x_{23} + 2x_{33} = 1$$

Diese können wir auflösen und erhalten: $\begin{pmatrix} -\frac{11}{150} & \frac{6}{25} & \frac{13}{300} \\ \frac{1}{30} & -\frac{1}{5} & \frac{7}{60} \\ \frac{13}{150} & \frac{2}{25} & -\frac{29}{300} \end{pmatrix}$

◇

Nach diesen Ausführungen nun zur multiplen Regression.

29.2 Berechnung der Multiplen Regressions-Geraden (Deskriptive Statistik)

Ein Beispiel zur Veranschaulichung: Es soll untersucht werden, welche Bedeutung der Preis oder die Werbeausgaben für die Verkaufszahlen eines Produktes haben. Es liegen damit zwei unabhängige Variablen vor (Preis, Werbeausgaben) und eine Zielvariable (zu erklärende Variable; Verkaufszahlen). Es liegt damit der dreidimensionale Fall vor. Alle Variablen sind metrisch skaliert. Im dreidimensionalen Fall kann die Punktwolke graphisch dargestellt werden:

(Für die Daten „bivariat.xls“, Übung 8). Es geht nun darum, durch diese Punktwolke eine optimale Gerade zu berechnen.

29.2.1 Berechnung der Geraden (minimale Summe der quadrierten Abweichungen)

Gegeben seien n Datensätze (= Vektor von Werten von n Zufallsvektoren $(X_1, \dots, X_k, Y)$). Eine solche Menge von Datensätzen hat die Form einer $n \cdot (k+1)$ -Matrix. Es wird ein linearer Zusammenhang zwischen den Variablen $X_1, \dots, X_k$ sowie einer Konstante b_0 einerseits, und der Zielvariable Y andererseits vermutet. Mit anderen Worten gilt laut Vermutung:

$$y_i = \sum_{j=0}^k x_{ij} b_j + r_i$$

für $1 \leq i \leq n$,

$x_{i0} = 1$ (dies wird gesetzt, damit $x_{i0}b_0$ nicht durch Setzung 0 wird).

b_j , $0 \leq j \leq k$, sind die $k+1$ Koeffizienten, die wir bestimmen wollen.

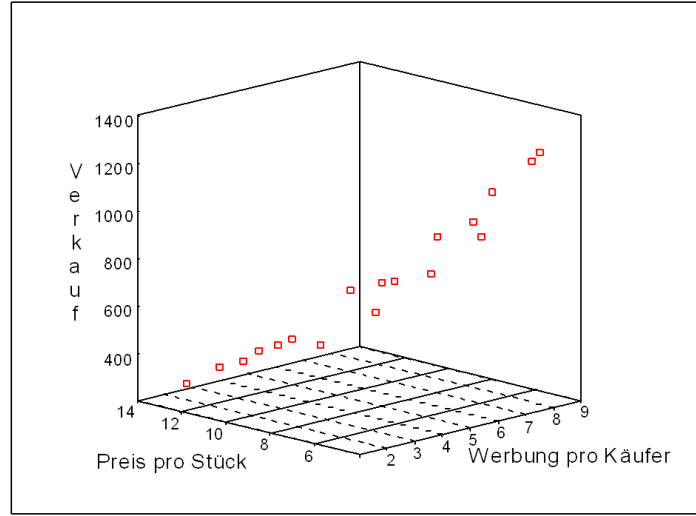


Abbildung 29.2.1: Graphische Darstellung der Daten mittels dreidimensionalem Streudiagramm

r_i ist das Residuum $y_i - \sum_{j=0}^k b_j x_{ij}$ des Datums y_i (Abweichung des Datums von der geschätzten Gerade).

Mit $\mathbf{X} = (x_{ij})$ (= die Matrix der Werte der Variablen $1, X_{i,1}, \dots, X_{i,k}; i \in \mathbb{N}_n, j \in \mathbb{N}_k$),

$\mathbf{b} = (b_0, \dots, b_k)$

$\mathbf{y} = (y_1, \dots, y_n)$ und

$\mathbf{r} = (r_1, \dots, r_n)$

kann man $y_i = \sum_{j=0}^k x_{ij} b_j + r_i$ für $0 \leq i \leq n$ auch in Matrixschreibweise ausdrücken als

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r}.$$

Auch hier geht es darum, die Summe der quadrierten Abstände zu minimieren, d.h. die b_j ($0 \leq j \leq k$) sind so zu bestimmen, dass $\sum_{i=1}^n (y_i - \sum_{j=0}^k b_j x_{ij})^2 =: \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$ minimal wird. Man kann beweisen

Satz 29.2.1. $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$ ist minimal genau dann, wenn

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Beweis. mit $\mathbf{b}_0 := (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ gilt:

$$\begin{aligned} \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 &= \|\mathbf{y}\|^2 - 2\mathbf{b}^T \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{b} \\ &= \|\mathbf{y}\|^2 - 2\mathbf{b}^T \left((\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \right) \mathbf{X}^T \mathbf{y} + \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{b} \\ &= \|\mathbf{y}\|^2 - 2\mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \right) + \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{b} \\ &= \|\mathbf{y}\|^2 - 2\mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0 + \mathbf{b}^T \mathbf{X}^T \mathbf{X} \mathbf{b}. \end{aligned} \quad (29.2.1)$$

Nun gilt

$$\begin{aligned} (\mathbf{b} - \mathbf{b}_0)^T (\mathbf{X}^T \mathbf{X}) (\mathbf{b} - \mathbf{b}_0) &= \mathbf{b}^T (\mathbf{X}^T \mathbf{X}) (\mathbf{b} - \mathbf{b}_0) - \mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) (\mathbf{b} - \mathbf{b}_0) \\ &= (\mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b} - \mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0 - \mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) \mathbf{b} + \mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0). \end{aligned}$$

$\mathbf{X}^T \mathbf{X}$ ist eine symmetrische Matrix, so dass $\mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0 = \mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}$ und man erhält:

$$(\mathbf{b} - \mathbf{b}_0)^T (\mathbf{X}^T \mathbf{X}) (\mathbf{b} - \mathbf{b}_0) = \mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b} - 2\mathbf{b}^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0 + \mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0.$$

Vergleicht man den Ausdruck rechts mit (29.2.1), erhält man

$$\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 = \|\mathbf{y}\|^2 - \mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0 + (\mathbf{b} - \mathbf{b}_0)^T (\mathbf{X}^T \mathbf{X}) (\mathbf{b} - \mathbf{b}_0).$$

Für symmetrische Matrizen, die invertierbar sind, gilt $(\mathbf{b} - \mathbf{b}_0)^T (\mathbf{X}^T \mathbf{X}) (\mathbf{b} - \mathbf{b}_0) > 0$ für $\mathbf{b} \neq \mathbf{b}_0$. $\|\mathbf{y}\|^2$ und $\mathbf{b}_0^T (\mathbf{X}^T \mathbf{X}) \mathbf{b}_0$ sind konstant. Damit wird $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$ minimal, genau dann, wenn $\mathbf{b} = \mathbf{b}_0$. $\square$

Man kann zeigen, dass $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ ein erwartungstreuer Schätzer des Parametervektors der Verteilung ist, aus der die Daten stammen. Zudem ist dieser Schätzer unter allen linearen Schätzern von minimaler Varianz. Für diesen letzten Beweis müssen wir voraussetzen, dass die Datensätze von $\mathbf{X}$ nicht als Realisationen von Zufallsvariablen zu deuten sind, sondern als feste Punkte, unter denen wir die Werte Y erheben.

29.2.2 Berechnungsbeispiel

Wir wählen ein Beispiel mit wenigen Zahlen, damit die Rechnung mit manuellen Mitteln nachvollziehbar ist. Wir nehmen an, wir hätten die Vermutung, dass es zwischen dem Nettoeinkommen, der Schmuckgeschäftsdichte und dem Geschlecht einerseits (statistische Variablen $X_1, \dots, X_3$) und den Ausgaben für Schmuck andererseits (Variable Y) einen linearen Zusammenhang gibt. Damit gilt für alle i : $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + r_i$. Zu schätzen sind die b_j ($0 \leq j \leq 3$).

Beispiel 29.2.2. Wir erheben die folgenden Daten, wobei „Einkommen“ das steuerbare Nettoeinkommen in Tausend SFr. misst, „Geschäftsdichte“ die Anzahl Verkaufsstellen pro 10000 Einwohner, „Geschlecht“ die Ausprägung 1 für „weiblich“ aufweist und „Ausgaben für Schmuck“ die Ausgaben in Franken pro Jahr sind (s. Tabelle 29.2.1).

Einkommen	Geschäftsdichte	Geschlecht	Ausgaben für Schmuck
5	4	1	160
3	3	1	140
2	1	0	130
1	2	0	100
3	2	1	110

Tabelle 29.2.1: Beispieldaten für eine multiple Regression

Damit erhalten wir die Matrix und den Vektor:

$$\mathbf{X} = \begin{pmatrix} 1 & 5 & 4 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 3 & 2 & 1 \end{pmatrix}; \mathbf{y} = \begin{pmatrix} 160 \\ 140 \\ 130 \\ 100 \\ 110 \end{pmatrix}$$

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 3 & 2 & 1 & 3 \\ 4 & 3 & 1 & 2 & 2 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 5 & 4 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 14 & 12 & 3 \\ 14 & 48 & 39 & 11 \\ 12 & 39 & 34 & 9 \\ 3 & 11 & 9 & 3 \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} 5 & 14 & 12 & 3 \\ 14 & 48 & 39 & 11 \\ 12 & 39 & 34 & 9 \\ 3 & 11 & 9 & 3 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{53}{34} & -\frac{9}{34} & -\frac{15}{34} & \frac{25}{34} \\ -\frac{9}{34} & \frac{15}{34} & -\frac{9}{34} & -\frac{19}{34} \\ -\frac{15}{34} & -\frac{9}{34} & \frac{19}{34} & -\frac{9}{34} \\ \frac{25}{34} & -\frac{19}{34} & -\frac{9}{34} & \frac{83}{34} \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{pmatrix} \frac{53}{34} & -\frac{9}{34} & -\frac{15}{34} & \frac{25}{34} \\ -\frac{9}{34} & \frac{15}{34} & -\frac{9}{34} & -\frac{19}{34} \\ -\frac{15}{34} & -\frac{9}{34} & \frac{19}{34} & -\frac{9}{34} \\ \frac{25}{34} & -\frac{19}{34} & -\frac{9}{34} & \frac{83}{34} \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 5 & 3 & 2 & 1 & 3 \\ 4 & 3 & 1 & 2 & 2 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -\frac{27}{34} & \frac{3}{17} & \frac{10}{17} & \frac{7}{17} & \frac{21}{34} \\ \frac{11}{34} & -\frac{5}{17} & \frac{6}{17} & -\frac{6}{17} & -\frac{1}{34} \\ \frac{7}{34} & \frac{3}{17} & -\frac{7}{17} & \frac{7}{17} & -\frac{13}{34} \\ -\frac{23}{34} & \frac{12}{17} & -\frac{11}{17} & -\frac{6}{17} & \frac{33}{34} \end{pmatrix}$$

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \begin{pmatrix} -\frac{27}{34} & \frac{3}{17} & \frac{10}{17} & \frac{7}{17} & \frac{21}{34} \\ \frac{11}{34} & -\frac{5}{17} & \frac{6}{17} & -\frac{6}{17} & -\frac{1}{34} \\ \frac{7}{34} & \frac{3}{17} & -\frac{7}{17} & \frac{7}{17} & -\frac{13}{34} \\ -\frac{23}{34} & \frac{12}{17} & -\frac{11}{17} & -\frac{6}{17} & \frac{33}{34} \end{pmatrix} \begin{pmatrix} 160 \\ 140 \\ 130 \\ 100 \\ 110 \end{pmatrix} = \begin{pmatrix} \frac{1415}{17} \\ \frac{305}{17} \\ \frac{55}{17} \\ -\frac{375}{17} \end{pmatrix} \approx \begin{pmatrix} 83.235 \\ 17.941 \\ 3.2353 \\ -22.059 \end{pmatrix} \approx \mathbf{b}$$

Damit erhalten wir für alle i jeweils die Gleichung:

$$y_i = \frac{1415}{17} + \frac{305}{17} x_{i1} + \frac{55}{17} x_{i2} - \frac{375}{17} x_{i3} + r_i,$$

wobei wir zur Schätzung von Werten

$$\hat{y}_i = \frac{1415}{17} + \frac{305}{17} x_{i1} + \frac{55}{17} x_{i2} - \frac{375}{17} x_{i3}$$

verwenden. Die Koeffizienten b_0, b_1 und b_2 sind positiv. Für b_0 bedeutet dies, dass b_0 an Schmuck gekauft wird, wenn alle Variablen 0 sind (kein Einkommen, kein Geschäft auf 10'000 Einwohner, männlich; b_0 ist der Ordinatenabschnitt). Ein positives b_1 bedeutet, dass mehr Schmuck gekauft wird, wenn das Einkommen steigt. Dabei gibt b_1 die Steigung des entsprechenden Zusammenhangs an. Analoges gilt für b_2 . b_3 ist negativ. Nimmt x_{i3} den Wert 0 an (= Mann), so wird nichts abgezählt, nimmt x_{i3} den Wert 1 an (= Frau), so wird $\frac{375}{17} = 22.059$ von den zu schätzenden Ausgaben für Schmuck abgezählt.

Die Gleichung können wir nun verwenden, um zu schätzen, wieviel jemand für Schmuck ausgibt, wenn wir wissen, wieviel er verdient, welches die Schmuckgeschäftsdichte seines Wohnortes ist und welches Geschlecht er hat. z.B. Jemand verdiene 3000 pro Monat, die Dichte sei 2 und das Geschlecht sei männlich. Dies ergibt in die Gleichung eingesetzt:

$$\hat{y}_i = \frac{1415}{17} + \frac{305}{17} \cdot 3 + \frac{55}{17} \cdot 2 - \frac{375}{17} \cdot 0 = 143.53 \text{ Fr.}$$

Diese Person wird also im Mittel 143.53 Franken für Schmuck ausgeben.
Wäre die Person eine Frau, würde sich ergeben:

$$\hat{y}_i = \frac{1415}{17} + \frac{305}{17} \cdot 3 + \frac{55}{17} \cdot 2 - \frac{375}{17} \cdot 1 = \frac{2065}{17} = 121.47 \text{ Fr.}$$

◇

29.2.3 Berechnung mit Excel

Sie finden die Excel-Berechnung für das obige Beispiel unter Beispiel 1, „multiple Regression.xls“ auf der Homepage.

- Berechnung der transponierten Matrix $\mathbf{X}^T$: $\mathbf{X}$ kopieren. Auf die Zelle gehen, von der weg nach links-unten die transponierte Matrix eingefügt werden soll. Anklicken. Dann „bearbeiten“, „Inhalte einfügen“, transponiert, ok.
- Als nächstes berechnen wir $\mathbf{X}^T \mathbf{X}$ mit Hilfe des Excelbefehles „mmult($\mathbf{X}^T$; $\mathbf{X}$)“: Dazu müssen wir vorgängig einen leeren Teil der Tabelle beleuchten (mit $(k+1) \times (k+1)$ Zellen, im Beispiel mit 4×4 Zellen schreibt dann „mmult($\mathbf{X}^T$; $\mathbf{X}$)“ mit den zu multiplizierenden Matrizen - in der richtigen Reihenfolge.
- Wir berechnen nun die inverse Matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ dieser Matrix: Dazu muss ein leerer Teil der Exceltabelle beleuchtet werden, der gleichviele Zeilen und Spalten hat wie die obige Datenmatrix (im Beispiel 4×4): Dann wird in „minv($\mathbf{X}^T \mathbf{X}$)“ eingegeben und bestätigt mit ctrl + shift + enter. Wir erhalten $(\mathbf{X}^T \mathbf{X})^{-1}$.
- Nun berechnen wir die Multiplikation (wie bereits beschrieben) dieser Matrix mit der $\mathbf{X}'$, indem wir in einem leeren Teil der Tabelle $(k+1) \times n$ Zellen beleuchten.
- Schliesslich müssen wir $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ berechnen, wobei $\mathbf{y} = (160, 140, 130, 100, 110)'$ (Multiplikation von $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}'$ mit $\mathbf{y}$ mit dem Befehl =mmult() und Bestätigung mit: ctrl, shift, enter.

Bemerkung 29.2.3. In der Praxis wird nicht die obige, theoretische korrekte Berechnungsmethode verwendet, da wegen Rundungsfehlern die Ergebnisse bei langen Datentabellen nicht zuverlässig sind. $\diamond$

29.2.4 Berechnung der Residuen

Setzen wir die erhobenen Daten in die Gleichung ein, können wir die Residuen berechnen:

$$\begin{aligned} \frac{1415}{17} + \frac{305}{17} \cdot 5 + \frac{55}{17} \cdot 4 - \frac{375}{17} \cdot 1 &= 163.8235294 \\ \frac{1415}{17} + \frac{305}{17} \cdot 3 + \frac{55}{17} \cdot 3 - \frac{375}{17} \cdot 1 &= 124.7058824 \\ \frac{1415}{17} + \frac{305}{17} \cdot 2 + \frac{55}{17} \cdot 1 - \frac{375}{17} \cdot 0 &= 122.3529412 \\ \frac{1415}{17} + \frac{305}{17} \cdot 1 + \frac{55}{17} \cdot 2 - \frac{375}{17} \cdot 0 &= 107.6470588 \\ \frac{1415}{17} + \frac{305}{17} \cdot 3 + \frac{55}{17} \cdot 2 - \frac{375}{17} \cdot 1 &= 121.4705882 \end{aligned}$$

oder in Matrizenform: $\mathbf{Xb} = \begin{pmatrix} 1 & 5 & 4 & 1 \\ 1 & 3 & 3 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 1 & 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1415}{17} \\ \frac{305}{17} \\ \frac{55}{17} \\ -\frac{375}{17} \end{pmatrix} = \begin{pmatrix} 163.8235294 \\ 124.7058824 \\ 122.3529412 \\ 107.6470588 \\ 121.4705882 \end{pmatrix}$

$$\mathbf{r} = \mathbf{y} - \mathbf{Xb} = \begin{pmatrix} 160 \\ 140 \\ 130 \\ 100 \\ 110 \end{pmatrix} - \begin{pmatrix} 163.8235294 \\ 124.7058824 \\ 122.3529412 \\ 107.6470588 \\ 121.4705882 \end{pmatrix} = \begin{pmatrix} -3.82353 \\ 15.29412 \\ 7.64706 \\ -7.64706 \\ -11.47059 \end{pmatrix}$$

29.3 Testtheorie

29.3.1 ANOVA

Ziel: Man will überprüfen, ob die unabhängigen Variablen Varianz einer Zielvariable „erklären“ und ob die durch die Variablen wegerklärte Varianz bezüglich der verbleibenden Varianz bedeutsam ist.

Voraussetzungen: Die Zielvariable ist metrisch skaliert. Die erklärenden Variablen sind metrisch skaliert oder weisen nur zwei Ausprägungen auf. Nominalskalierte Variablen mit mehr als zwei Variablen werden in dichotome Dummyvariablen umgewandelt (Dies wird von R automatisch gemacht, wenn man die Variablen als factor eingibt). Die Residuen sind normalverteilt ($N(0, \sigma^2)$) und es liegen mehr als 40 Objekte vor. Die Varianzen der Residuen bezüglich der Regressionsgerade sind identisch. Die unabhängigen Variablen sind mit den Residuen nicht korreliert. Die zweidimensionalen Punktwolken (x_{ij}, y_i) ($j \in \mathbb{N}_k$) liegen in etwa auf einer Geraden und sind punktsymmetrisch bezüglich $(\bar{x}_j, \bar{y}_j)$. Die unabhängigen Variablen sind nicht linear abhängig.

Trotz dieser Ansammlung von Bedingungen, lässt sich die multiple Regression oft fruchtbringend anwenden.

Man kann zeigen, dass gilt:

$\frac{\frac{1}{k-1}SQE}{\frac{1}{n-k}SQR}$ ist $F_{k-1, n-k}$ -verteilt.
$SQE = \sum_{i=1}^n (f(x_{0,i}, \dots, x_{k,i}) - \bar{Y})^2$ (erklärte Streuung)
$SQR = \sum_{i=1}^n (Y_i - f(x_{0,i}, \dots, x_{k,i}))^2$ (restliche Streuung)
n = Anzahl der Untersuchungsobjekte
k = Anzahl erklärender Variablen in der Regressionsgleichung
$f((x_{0,i}, \dots, x_{k,i})) = b_0 + b_1x_{i,1} + \dots + b_kx_{i,k}$ (Regressionsgerade)
$\bar{Y}$ = Zufallsvariable Mittelwerte der Zufallsvariablen Y_i

Wir berechnen dies am obigen Beispiel. Wie oben berechnet gilt:

$$SQR = \begin{pmatrix} -3.82353 \\ 15.29412 \\ 7.64706 \\ -7.64706 \\ -11.47059 \end{pmatrix}^T \cdot \begin{pmatrix} -3.82353 \\ 15.29412 \\ 7.64706 \\ -7.64706 \\ -11.47059 \end{pmatrix} = 497.059$$

$\bar{Y}$ nimmt den Wert $\frac{160+140+130+100+110}{5} = 128$ an. Damit erhalten wir:

$$\left(\begin{pmatrix} 163.8235294 \\ 124.7058824 \\ 122.3529412 \\ 107.6470588 \\ 121.4705882 \end{pmatrix} - \begin{pmatrix} 128 \\ 128 \\ 128 \\ 128 \\ 128 \end{pmatrix} \right) = \begin{pmatrix} 35.82352941 \\ -3.294117647 \\ -5.647058824 \\ -20.35294118 \\ -6.529411765 \end{pmatrix}$$

und

$$SQE = \begin{pmatrix} 35.82352941 \\ -3.294117647 \\ -5.647058824 \\ -20.35294118 \\ -6.529411765 \end{pmatrix}^T \cdot \begin{pmatrix} 35.82352941 \\ -3.294117647 \\ -5.647058824 \\ -20.35294118 \\ -6.529411765 \end{pmatrix} = 1782.941176.$$

Es werden 4 Koeffizienten geschätzt. Damit sind die Zählerfreiheitsgrade 3.
 $n = 5$. Damit erhalten wir $n - k = 5 - 4 = 1$ Nennerfreiheitsgrade. Die Teststatistik

$$\frac{\frac{1}{k-1}SQE}{\frac{1}{n-k}SQR}$$

nimmt damit den Wert

$$\frac{\frac{1}{3} \cdot 1782.941176}{\frac{1}{1} \cdot 497.059} = 1.1957$$

an. p-Wert: $F_{F_{3,1}}(1.1957) = 0.57212402$

Zusammenfassend (s. Tabelle 29.3.2):

	Quadratsummen	df	Varianzen	Teststatistik	p-Wert
Regression	1782.941	3	594.3136667	1.196	0.572
Residuen	497.059	1	497.059		
Gesamt	2280	4	570		

Tabelle 29.3.2: Darstellung der Berechnung der ANOVA des Beispiels

29.3.2 Testen der Koeffizienten - Vertrauensintervalle für die Koeffizienten

Ziel: Es soll überprüft werden, ob die Koeffizienten der Gleichung $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r}$ signifikant von 0 verschieden sind.

Voraussetzungen: siehe oben.

Ein Zufallsvektor ist definiert als ein Vektor (= n-Tupel) von Zufallsvariablen. Sei $\mathbf{B}$ der Zufallsvektor $\mathbf{B} = (B_0, \dots, B_k)$, welcher die geschätzten Koeffizienten $\mathbf{b}$ als konkrete Werte annimmt. Dann gilt bei iid. normalverteilten Residuen:

$\frac{B_i}{\sqrt{V(B_i)}}$ ist t_{n-k} -verteilt (für $i > 0$).
B_i : Zufallsvariable mit dem Koeffizienten b_i als Wert
$V(B_i)$ geschätzt durch $\frac{c_{ii}}{n-k} \sum_{i=0}^n (Y_i - f(x_{i,0}, \dots, x_{i,k}))^2$ ($V(B_i) =$
Varianz der Zufallsvariable B_i);
$\sqrt{V(B_i)}$ = Standardfehler von B_i
c_{ii} die Werte der Diagonalen von $\mathbf{C} := (\mathbf{X}^T \mathbf{X})^{-1}$
k = Anzahl erklärender Variablen der Regressionsgleichung
n : Anzahl der untersuchten Objekte.

Wir können wieder rechts- links- oder zweiseitig testen. Wir testen fürs Beispiel zweiseitig:
 Fürs obige Beispiel erhalten wir:

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} \frac{53}{34} & -\frac{9}{34} & -\frac{15}{34} & \frac{25}{34} \\ -\frac{9}{34} & \frac{15}{34} & -\frac{9}{34} & -\frac{19}{34} \\ -\frac{15}{34} & -\frac{9}{34} & \frac{19}{34} & -\frac{9}{34} \\ \frac{25}{34} & -\frac{19}{34} & -\frac{9}{34} & \frac{83}{34} \end{pmatrix}$$

Damit gilt mit $\sum_{i=0}^n (Y_i - f(x_{i,0}, \dots, x_{i,k}))^2 = 497.059$ und $\mathbf{b} = \begin{pmatrix} \frac{1415}{17} \\ \frac{305}{17} \\ \frac{55}{17} \\ -\frac{375}{17} \end{pmatrix}$ (siehe oben)

(a) $V(B_1)$ wird geschätzt durch $\frac{15}{34} \cdot 497.059$.

Die Teststatistik $\frac{B_1}{\sqrt{V(B_1)}}$ nimmt den Wert damit den Wert $\frac{\frac{305}{17}}{\sqrt{\frac{15}{34} \cdot 497.059}} = 1.211548636506$

an. p-Wert: $F_{t_1}(1.211548636506) = 1 - 0.21964400501$

p-Wert (zweiseitig): $0.4392880106065 > 0.05$.

Die Nullhypothese kann nicht verworfen werden.

(b) $V(B_2)$ wird geschätzt durch $\frac{19}{34} \cdot 497.059$.

$\frac{B_2}{\sqrt{V(B_2)}}$ nimmt den Wert damit den Wert $\frac{\frac{55}{17}}{\sqrt{\frac{19}{34} \cdot 497.059}} = 0.1941210055828$

an. $F_{t_1}(0.1941210055828) = 1 - 0.438968427$

p-Wert (zweiseitig): $0.8779368531072 > 0.05$

Die Nullhypothese kann nicht verworfen werden.

(d) $V(B_3)$ wird geschätzt durch $\frac{83}{34} \cdot 497.059$.

$\frac{B_3}{\sqrt{V(B_3)}}$ nimmt den Wert damit den Wert $\frac{-\frac{375}{17}}{\sqrt{\frac{83}{34} \cdot 497.059}} = -0.6332553460944$

an. p-Wert: $F_{t_1}(-0.6332553460944) = 0.320309699$

p-Wert (zweiseitig): 0.6406193972566

Die Nullhypothese kann nicht verworfen werden.

(Man kann auch Vertrauensintervalle für die Konstante berechnen. Da diese inhaltlich gewöhnlich nicht besonders wichtig ist, verzichten wir darauf).

Wir erhalten bei zweiseitigen Tests (s. Tabelle 29.3.3):

	Koeffizienten	Testwert	p-Wert
Konstante	83.235	2.99	0.205
Einkommen	17.941	1.212	0.439
Schmuckgeschäftsichte	3.235	0.194	0.878
Geschlecht	-22.059	-0.633	0.641

Tabelle 29.3.3: Test- und p-Werte bei Testen der Koeffizienten

Die Parameter - abgesehen von der Konstanten geben - die Steigung der Geraden in die Richtung der jeweiligen Achse an. Damit kann man ablesen, um wieviel die abhängige Variable zunimmt, wenn die der Wert der unabhängigen Variable in eine bestimmte Dimension um eine Einheit zunimmt. Man kann die Steigungsparameter allerdings nicht unmittelbar vergleichen, da sie unterschiedliche Einheiten aufweisen. Um die Bedeutsamkeit der verschiedenen unabhängigen Variablen einzuschätzen werden darum die sogenannte standardisierten Koeffizienten berechnet. Diese erhält man, wenn man die Regression auf die standardisierten Daten berechnet oder indem man die Koeffizienten wie folgt standardisiert:

$$beta_j = b_j \cdot \frac{S_{X_j}}{S_Y}$$

(Koeffizient b_j ; S_{X_j} = Standardabweichung der Variable X_j ; S_Y = Standardabweichung der Variable Y). Mit Hilfe der standardisierten Koeffizienten kann man die Wichtigkeit des Einflusses der verschiedenen Variablen direkt vergleichen. Im Beispiel wäre das Einkommen, dann das Geschlecht und schliesslich die Schmuckgeschäftsdichte in dieser Reihenfolge die bestimmenden Faktoren (aber alles nicht signifikant!).

Nützlich sind noch folgende Berechnungen: Das R -Quadrat ($= R^2$) ist das lineare Bestimmtheitsmass $\frac{SQE}{SQT}$. Es ist gleich definiert wie bei der bivariaten Regression. R ist die Wurzel aus R^2 und wird „Quasi-Korrelation“ genannt. Sie drückt nicht mehr die Korrelation zwischen zwei Variablen aus. Das Bestimmtheitsmass erlaubt eine erste Überprüfung, ob die Regression (Gerade) zu den Daten passt. Das korrigierte R^2 berücksichtigt den Umstand, dass durch die Berücksichtigung von mehr Variablen R^2 steigt, obwohl die zusätzlichen Variablen u.U. nichts entscheidendes zur Erklärung (= Reduktion der Varianz) beitragen. Deshalb wird R^2 wie folgt korrigiert:

$$R_{kor}^2 = R^2 - \frac{k(1 - R^2)}{n - k}.$$

Während R^2 mit zusätzlich berücksichtigen erklärenden Variablen immer steigt, kann R_{kor}^2 sinken. Dies zeigt an, dass die zusätzlichen Variablen nichts entscheidendes zur Erklärungskraft beitragen. Mit „Standardfehler des Schätzers“ wird manchmal die Standardabweichung der Residuen bezeichnet. Es handelt sich um eine Masszahl, die ebenfalls eine Einschätzung der Anpassung der Geraden an die Daten erlaubt.

Bei der multiplen Regression stellt sich das Problem, ein möglichst einfaches Modell

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r}$$

zu finden, das die Daten möglichst gut erklärt. Entsprechend rechnet man die Regression für verschiedene Modelle durch, um das beste zu finden. Es gibt dabei verschiedene Methoden. In der Praxis bewährt hat sich die sogenannte „Rückwärtselimination“. Zuerst wird ein möglichst vollständiges Modell aufgestellt - auf Grund unseres ausserstatistischen Wissens bezüglich des Untersuchungsgebietes. Dann werden jeweils Variablen mit von 0 nicht signifikant verschiedenen Koeffizienten weggelassen, bis man zu einem Modell gelangt, das nur mehr Variablen mit von 0 signifikant verschiedene Koeffizienten aufweist (dies betrifft nicht die Konstante, die man drin lässt). Statistikprogramme erlauben dieses Vorgehen mit einem Schritt, indem man die entsprechende Methode anklickt. Zu beachten ist, dass man dabei - wegen mehrfachen Testens - bei konstantem Signifikanzniveau die Fehlerwahrscheinlichkeiten 1. Art vergrössert. Entsprechend wird pro Test das Signifikanzniveau oft um den Faktor $\frac{1}{k}$ verkleinert - mit k des ausführlichsten Modells (Bonferroni-Korrektur).

Beispiel 29.3.1. *Zum Einstieg in die Statistik betrachteten wir ein Beispiel von Immobiliendaten, die verwendet wurden, um den künftigen Preis von Immobilien zu schätzen. Wir erhalten:* $\diamond$

Wir können beobachten, dass die Quasi-Korrelation und R^2 durch das Weglassen von Variablen im Beispiel sinken, während die R_{kor}^2 steigt. Der Standardfehler des Schätzers sinkt durch das Weglassen von Variablen, d.h. dass die Schätzungen bei Voraussagen präziser werden.

Bei allen Modellen ist die ANOVA signifikant. Allerdings sehen wir dass der F-Wert steigt. Zudem sinken die Freiheitsgrade der Regression und die Freiheitsgrade der Residuen steigen. Damit steigt der p-Wert von $F_{k-1, n-k-1}$ bei den Modellen mit weniger Variablen. Damit wird die ANOVA immer signifikanter.

Im Modell 1 sind das Alter, die Anzahl Objekte und die Lage nicht signifikante Einflüsse. Den grössten p-Wert hat das Alter. Deshalb wird diese Variable im Modell 2 weggelassen. Anzahl Objekte und Lage sind weiterhin nicht signifikant. Da die Lage den höchsten p-Wert hat, wird diese Variable weggelassen. Wir erhalten Modell 3. Die Anzahl Objekte ist in diesem Modell nicht signifikant. Wir lassen deshalb diese Variable weg und erhalten das Modell 4. In diesem sind die drei verbleibenden Variablen (ausser der Konstanten) signifikant auf dem Niveau $\frac{0.05}{6} = 0.008333$.

Modellzusammenfassung ^e

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	.926 ^a	.857	.842	160.54614
2	.926 ^b	.857	.845	159.31335
3	.925 ^c	.856	.846	158.43754
4	.925 ^d	.855	.848	157.66952

- a. Einflußvariablen : (Konstante), Steuern, Alter, Anzahl Objekte, Ecklage, Lage, Wohnraum in sqf
b. Einflußvariablen : (Konstante), Steuern, Anzahl Objekte, Ecklage, Lage, Wohnraum in sqf
c. Einflußvariablen : (Konstante), Steuern, Anzahl Objekte, Ecklage, Wohnraum in sqf
d. Einflußvariablen : (Konstante), Steuern, Ecklage, Wohnraum in sqf
e. Abhängige Variable: Verkaufspreis

Abbildung 29.3.2: Berechnung der Kennzahlen für verschiedene Regressionsmodelle, wie sie etwa von Statistikpaketen angeboten wird

ANOVA ^e

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	9108035	6	1518005.9	58.894	.000 ^a
	Residuen	1520729	59	25775.062		
	Gesamt	1.1E+07	65			
2	Regression	9105919	5	1821183.8	71.755	.000 ^b
	Residuen	1522845	60	25380.745		
	Gesamt	1.1E+07	65			
3	Regression	9097514	4	2274378.5	90.604	.000 ^c
	Residuen	1531250	61	25102.455		
	Gesamt	1.1E+07	65			
4	Regression	9087464	3	3029154.6	121.850	.000 ^d
	Residuen	1541300	62	24859.678		
	Gesamt	1.1E+07	65			

- a. Einflußvariablen : (Konstante), Steuern, Alter, Anzahl Objekte, Ecklage, Lage, Wohnraum in sqf
b. Einflußvariablen : (Konstante), Steuern, Anzahl Objekte, Ecklage, Lage, Wohnraum in sqf
c. Einflußvariablen : (Konstante), Steuern, Anzahl Objekte, Ecklage, Wohnraum in sqf
d. Einflußvariablen : (Konstante), Steuern, Ecklage, Wohnraum in sqf
e. Abhängige Variable: Verkaufspreis

Abbildung 29.3.3: Berechnung ANOVA für verschiedene Regressionsmodelle, wie sie etwa von Statistikpaketen angeboten wird

Bemerkungen: (a) Kommen mehrere Variablen ins Spiel, so können auch Interaktionen zwischen den Variablen mit der Regression analysiert werden. Interaktionen liegen vor, wenn sich Variablen gegenseitig verstärken oder abschwächen. So kann die Wirksamkeit der Vertreter zunehmen, wenn man die Werbeausgaben erhöht. Bei metrisch skalierten unabhängigen Variablen kann man Interaktionen analysieren, indem man eine neue Variable schafft, welche aus dem Produkt der zu untersuchenden Variablen besteht. Diese Variable wird in die Berechnung der Regression einbezogen.

(b) Bei nominalskalierten Variablen mit mehr als zwei Ausprägungen kann wie folgt vorgegangen

Koeffizienten ^a					
Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	Signifikanz
		B	Standardfehler	Beta	
1	(Konstante)	53.453	99.190		.592
	Wohnraum in sqf	.334	.096	.423	.001
	Alter	-.580	2.023	-.018	.775
	Anzahl Objekte	10.994	18.212	.035	.548
	Lage	-20.899	47.711	-.025	.663
	Ecklage	172.486	54.261	.184	.002
	Steuern	.534	.158	.418	.001
2	(Konstante)	40.795	88.128		.645
	Wohnraum in sqf	.322	.085	.407	.000
	Anzahl Objekte	12.653	17.135	.040	.463
	Lage	-25.591	44.469	-.031	.567
	Ecklage	170.625	53.458	.182	.002
	Steuern	.560	.129	.438	.000
3	(Konstante)	46.745	87.039		.593
	Wohnraum in sqf	.313	.083	.396	.000
	Anzahl Objekte	10.530	16.641	.033	.529
	Ecklage	172.822	53.028	.185	.002
	Steuern	.562	.128	.440	.000
4	(Konstante)	74.799	74.537		.320
	Wohnraum in sqf	.321	.082	.406	.000
	Ecklage	178.463	52.020	.191	.001
	Steuern	.561	.128	.439	.000

a. Abhängige Variable: Verkaufspreis

Abbildung 29.3.4: Berechnung der Koeffizienten und der p-Werte für verschiedene Regressionsmodelle, wie sie etwa von Statistikpaketen angeboten wird

werden: Wir zerlegen die Variable in $k - 1$ Variablen (k = Anzahl Ausprägungen der Variable), indem wir dichotome Variablen („Dummy-Variablen“ genannt) mit den Ausprägungen „Ausprägung j, nicht Ausprägung j“ schaffen. z.B. Variable Nationalität mit den Ausprägungen „Schweizer“, „Italiener“, „Franzose“, „Deutscher“, „andere“ kann aufgeteilt werden in die Variablen „Schweizer (= 1) - Nicht-Schweizer (= 0)“, „Italiener (= 1), Nicht-Italiener (= 0)“, „Franzose (= 1), Nicht-Franzose (= 0)“, „Deutscher (= 1), Nicht-Deutscher (= 0)“. Der Ausprägung „andere“ muss keine neue Variable zugeordnet werden, da diese zutrifft, wenn allen anderen Dummy-Variablen 0 zugeordnet ist. Der Nachteil des Verfahrens besteht in einer starken Vermehrung von Variablen.

29.4 Überprüfung der Voraussetzungen

(a) Überprüfung auf Normalverteilung der Residuen durch einen P-P-Plot. Im Immobilienbeispiel erhalten wir (s. Abbildung 29.4.5):

Die Residuen scheinen nicht normalverteilt zu sein. Allerdings liegen mehr als 40 Daten vor, so dass die Abweichung von der Voraussetzung eher unproblematisch ist. Das Statistikpaket SPSS liefert z.B. auch die folgende Graphik zur Überprüfung der Normalverteilung der Residuen (s. Abbildung 29.4.6):

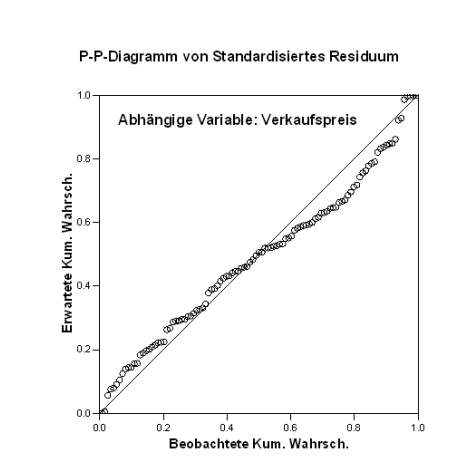


Abbildung 29.4.5: Überprüfung der Normalverteilungsannahme für die Residuen

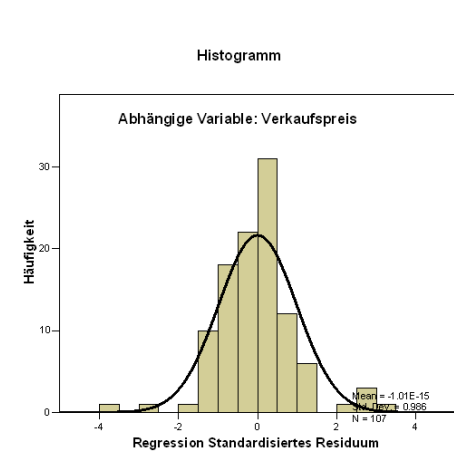


Abbildung 29.4.6: Überprüfung der Normalverteilungsannahme für die Residuen

(b) Die Varianzgleichheit (Homoskedastizität) kann getestet werden. So können wir z.B. die Menge der Residuen auf der y-Achse halbieren und dann die Varianz links mit der Varianz rechts mit einem F-Test vergleichen (dazu müssen die Residuen normalverteilt sein). Nützlich ist auch der folgende Plot: Wir berechnen die nicht standardisierten vorausgesagten Werte und die nicht standardisierten Residuen. Es kann dann ein Streudiagramm erstellt werden (s. Abbildung 29.4.7):

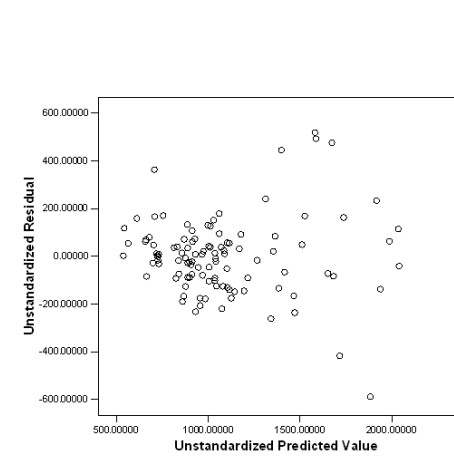


Abbildung 29.4.7: Testen der Varianzgleichheit

Man sieht, dass die Residuen eine ähnliche Streuung aufweisen - wenn wir zwei, drei Ausreisser weglassen. Die Varianzgleichheit ist verletzt, wenn die Punktwolke z.B. die Form eines Dreiecks aufweist (wenn man die Ausreisser nicht weglässt, hätte die obige Punktwolke die Form eines Dreiecks). Im Falle mit den y-Werten wachsender Varianz hilft oft das Logarithmieren der y-Werte. Im Beispiel würden wir erhalten (s. [Abbildung 29.4.8](#)):

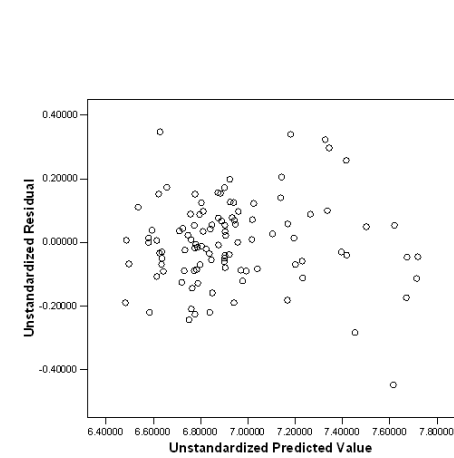


Abbildung 29.4.8: Testen der Varianzgleichheit

Die Varianzgleichheit ist nun bedeutend besser gewährleistet. Für die Koeffizienten erhalten wir in diesem Fall (s. [Tabelle 29.4.4](#)):

	Koeffizienten	Testwerte	p-Werte
Konstante	6.147	132.979	0
Wohnraum in sqf	0	3.602	0
Ecklage	0.98	2.683	0.008
Steuern	0.001	7.058	0

Tabelle 29.4.4: Koeffizienten mit logarithmierten Daten

(Dabei gilt beim Wohnraum genauer $B = 0.0001791$). Beim Schätzen von y -Werten müssen wir bei diesem Modell die y -Werte zurückrechnen: Es gilt nun $\ln(y) = \sum_{i=0}^k b_i x_i$. Um y auf Grund von $(x_1, \dots, x_k)$ zu berechnen, berechnen wir mit Hilfe der Regressionsgleichung $\ln(y)$ und dann $\exp(\ln(y)) = y$.

c) paarweise Linearität: Statistikprogramme liefern oft unter einem Befehl alle partiellen Diagramme (s. Abbildungen 29.4.9, 29.4.9 und 29.4.11):

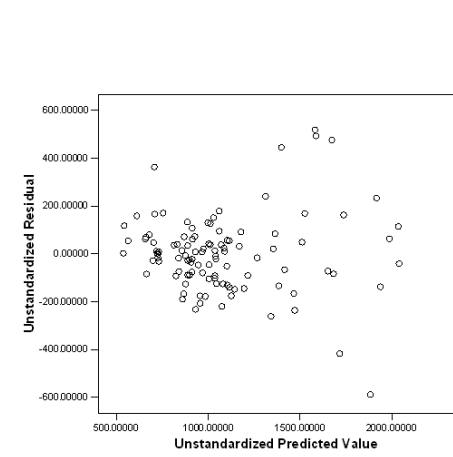


Abbildung 29.4.9: Testen der Varianzgleichheit

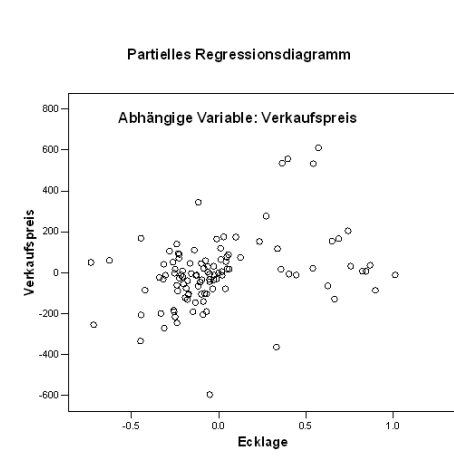


Abbildung 29.4.10: Testen der Varianzgleichheit

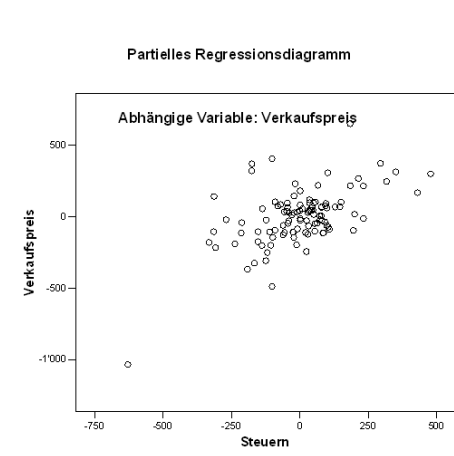


Abbildung 29.4.11: Testen der Varianzgleichheit

Die Linearität als solche kann akzeptiert werden. Punktsymmetrie ist problematischer. Es kommen Ausreisser vor. Es ist möglich, dass diese Ausreisser das Resultat beeinflussen. Man sollte u.U. die Regression ohne diese Ausreisser nachberechnen - Statistikprogramm liefern gewöhnlich Tools, um Ausreisser zu finden und kaltzustellen. Als Ausreisser gelten dabei Daten, die um mehr als 3 Standardabweichungen vom geschätzten Wert der Regressionsgeraden liegen.

Man sieht, dass die Residuen - wie gefordert - einen Mittelwert von 0 haben.

(d) Zwecks Prüfung der linearen Unabhängigkeit der unabhängigen Variablen kann man die Korrelationen zwischen den Variablen untersuchen. Im Beispiel korrelieren die verschiedenen Variablen, was z.B. bezüglich Steuern und Wohnfläche ja nicht erstaunt. Die Voraussetzung der linearen Unabhängigkeit wird aber erst bei Korrelationen nahe bei 1 als missachtet angeschaut. Liegt keine paarweise Unabhängigkeit vor, kann immer noch eine insgesamt lineare Abhängigkeit vorliegen (Multikollinearität). Diese kann man entdecken, indem man jeweils eine unabhängige Variable als abhängige wählt und mit den restlichen unabhängigen Variablen eine Regression berechnet.

(e) Die unabhängigen Variablen dürfen mit den Residuen nicht korrelieren. Um dies zu prüfen, führen wir Korrelationstests zwischen den unabhängigen Variablen und den Residuen durch. Wir sehen, dass die unabhängigen Variablen mit den Residuen nicht korrelieren (alle Korrelationskoeffizienten sehr nahe bei 0).

29.4.1 Übungen

1. Berechnen Sie Beispiel 1 mit Excel und mit R oder SAS
2. (a) Berechnen Sie Beispiel 2 mit R oder SAS.
 (b) Kritisieren Sie das ausführliche Modell des Beispiel 2 bezüglich Skalenniveaus der Variablen.
 (c) Berechnen Sie den zu erwartenden Verkaufspreis für folgende Immobilie:
 Wohnraum in sqf: 1850; Ecklage: 1; Steuern: 500
 (d) Wie exakt ist die Schätzung (Geben Sie ein Vertrauensintervall für die Schätzung an).
3. Berechnen Sie eine Regressionsgleichung für die Daten der Tabelle 3
 (Aus (Backhaus, Erichson, Plinke & Weiber, 2003)). Im Beispiel soll untersucht werden, durch welche Faktoren der Margarineumsatz beeinflusst wird. Ins Auge gefasst werden der Preis, die Ausgaben für die Verkaufsförderung sowie die Zahl der Vertreterbesuche. Dazu wurden Daten von 37 Verkaufsgebieten gewählt.

29.4.2 Lösungen

1. Siehe im Text
2. (b) Die Variable „Anzahl Objekt“ ist nicht metrisch skaliert, da die untersuchten Objekte kaum alle denselben Wert bezüglich Standard aufweisen. Diese Variable fällt aus dem Regressionsmodell aber dann heraus.
 (c) $74.799 + 1850 \cdot 0.321 + 1 \cdot 178.463 + 500 \cdot 0.561 = 1127.6$
 Da in Hundert gemessen wird: 112760 Dollars.
 (d) Die Residuen sind (oder sollten) normalverteilt sein. Der Standardfehler der Residuen $R_i =$ (Standardfehler des Schätzers $= \frac{1}{n-k} \sum_{i=1}^n R_i$; $k =$ Anzahl erklärender Variablen) beträgt für das Modell 4: 157.6695211571. Der Mittelwert der Residuen ist 0. Damit gilt: Schätzwert $\sim N(0, 157.6695211571)$. $F^{-1}(0.975) = 309.0263942$ mit F Verteilungsfunktion von $N(0, 157.6695211571)$ erhalten wir die Grenzen des Vertrauensintervalls $1127.6 + 309.0263942 = 1436.6$ und $1127.6 - 309.0263942 = 818.57$ sowie das 95%-Vertrauensintervall: $]818.57, 1436.6[$ oder in \$: $]81857, 143660[$
3. Siehe Exceltabelle

29.5 Lernziele

- Situationen erkennen, wo man die multiple lineare Regression anwenden kann.
- die multiple lineare Regression mit einem Statistikpaket berechnen können (Rückwärtsverfahren) und die Resultate interpretieren können. Die oben diskutierten Prüfverfahren bezüglich Voraussetzungen durchführen und interpretieren können.

Teil VII

Schliessende Statistik - Schätztheorie und Stichprobentheorie

Kapitel 30

Ein paar Bemerkungen zur Schätztheorie

In der Schätztheorie geht es darum, die Verteilung zu bestimmen, laut der *i.i.d.* Zufallsvariablen X_i , als deren Werte wir die Daten x_i betrachten, verteilt sind. Dies erfolgt oft durch das Schätzen von Parametern von Verteilungen aus spezifischen Verteilungsfamilien. Schätztheorie wurde bisher an mehreren Stellen des Skripts behandelt:

- Bei den Statistiken zeigten wir, dass die Statistik „Mittelwert“ ein erwartungstreuer Schätzer des Erwartungswertes der *i.i.d.* Zufallsvariablen X_i ist (s. Seite 216). Zudem zeigten wir, dass die Statistik „Mittelwert“ von *i.i.d.* Zufallsvariablen, die nur die Werte 0 oder 1 annehmen, die Wahrscheinlichkeit schätzt, dass der Wert 1 angenommen wird (s. Seite 228).
- Bei der Familie der Poissonverteilungen schätzten wir eine Poisson-Verteilung, laut der Zählraten verteilt sind, aus diesen Daten (s. Seite 245).
- Schliesslich schätzten wir mit Hilfe von empirischen Verteilungsfunktionen metrisch skalierten Daten stetige Wahrscheinlichkeitsverteilungen, gemäss der die *i.i.d.* Zufallsvariablen, als deren Werte wir die Daten betrachten, verteilt sind (s. Seite 295).

In der Folge geht es darum, Schätztheorie etwas systematischer zu betreiben und die bisherigen Ergebnisse in einen grösseren Zusammenhang einzuordnen.

Wir nennen künftig die Verteilung P der einzelnen Stichprobenzufallsvariablen X_i die „theoretische Verteilung von X_i “. Zu beachten ist, dass die theoretische Verteilung der X_i nicht identisch ist mit der Verteilung der Daten einer Grundgesamtheit. In der Schätztheorie betrachten wir die Verteilung der Daten der Grundgesamtheit genau so wie die Verteilung der Stichprobendaten - aus dieser Grundgesamtheit - als Realisierung von laut einer unbekannten theoretischen Verteilung verteilten Zufallsvariablen X_i . Da allerdings die empirische Verteilung einer grossen Datenmenge, die wir als Realisation von ebenso vielen *i.i.d.* nach P verteilten Zufallsvariablen betrachten, mit hoher Wahrscheinlichkeit sehr nahe bei der theoretischen Verteilung P liegt (s. Kapitel 13, Schätzen von stetigen Wahrscheinlichkeitsverteilungen durch Daten), unterscheidet sich die Verteilung der Daten der Grundgesamtheit kaum von der theoretischen Verteilung.

Beim Schätzen von Verteilungen gibt es im wesentlichen zwei Verfahren:

1. Vermuten wir - etwa auf Grund der Situation - das Vorliegen einer spezifischen Verteilungsfamilie, können wir die Verteilung schätzen, indem wir die Parameter der Verteilung schätzen. Es liegt im Rahmen einer Verteilungsfamilie nämlich eine bijektive Beziehung zwischen Verteilung und Parameter vor. Dieses Verfahren nennt man „parametrisches Schätzen“.
2. Haben wir keine Vermutung bezüglich des Vorliegens einer Verteilungsfamilie, müssen wir zuerst eine Verteilungsfamilie bestimmen. Wir können dabei den Umstand nutzen, dass

es z.B. zwischen Verteilungsfunktionen und Verteilungen einen bijektiven Zusammenhang gibt. Wir können entsprechend die Verteilungsfamilie anhand der graphischen Form der empirischen Verteilungsfunktion bestimmen. Dieses Verfahren wird „nicht-parametrisches Schätzen“ genannt. Haben wir die Verteilungsfamilie bestimmt, können wir in einem zweiten Schritt die Methoden des parametrischen Schätzens anwenden.

Haben wir eine Verteilung bestimmt, können wir in einem weiteren Schritt mit Hilfe der Testtheorie überprüfen, ob die geschätzte Verteilung mit den Daten verträglich ist. In der Folge werden die beiden Verfahren und die jeweiligen Tests ausführlicher behandelt.

30.1 Parametrisches Schätzen

Wir setzen zuerst voraus, dass die unbekannte theoretische Verteilung, laut der die *i.i.d* Zufallsvariablen X_i verteilt sind, zu einer spezifischen Verteilungsfamilie gehört. In diesem Fall ist die Verteilung durch einen Vektor von Parametern genau bestimmt ist. Der Vektor kann nur einen Parameter enthalten wie bei der Exponentialverteilung und der Poissonverteilung oder mehrere Parameter wie bei der Normalverteilung, der stetigen uniformen Verteilung, der Binomialverteilung oder der hypergeometrischen Verteilung. Umgekehrt ist durch die Verteilungsfamilie und die Verteilung genau ein Vektor von Parametern bestimmt, d.h. es gibt eine bijektive Funktion zwischen der Menge der Parametervektoren und der Menge der Verteilungen der Familie.

Beim Schätzen geht es darum, mit Hilfe einer Stichprobe $(x_1, \dots, x_n)$, die wir als Wert eines Zufallsvektors $(X_1, \dots, X_n)$ betrachten, die Parameter der Verteilung P (und damit P selber) zu schätzen (wobei gilt $X_i \sim P$). Dazu definieren wir auf $\times_{i=1}^n \text{Bild}(X_i)$ Zufallsvariablen, die wir in diesem Zusammenhang (statt Teststatistiken) „Schätzstatistiken“ nennen werden. Mit Hilfe dieser Statistiken schätzen wir die Parameter der Verteilung. Dieses Verfahren wurde bisher schon mehrmals angewendet. So haben wir bereits folgende Schätzstatistiken verwendet:

Die Statistik Mittelwert

$$\begin{aligned}\bar{X} : \times_{i=1}^n B(X_i) &\rightarrow \mathbb{R} \\ \bar{X}(x) &= \frac{1}{n} \sum_{i=1}^n X_i\end{aligned}$$

um den Erwartungswert von Verteilungen zu schätzen (Normalverteilung; Poissonverteilung). Die Statistik Mittelwert tauchte zudem in Statistiken auf, die wir zur Schätzung von Parametern weiterer Verteilungsfamilien verwendeten. So haben wir den Parameter λ der Exponentialverteilung durch $\frac{1}{\bar{X}}$ geschätzt. Den Parameter p der Binomialverteilung $B(n, p)$ schätzten wir durch $\bar{X}$.

Die Statistik Varianz

$$\begin{aligned}S^2 : \times_{i=1}^n B(X_i) &\rightarrow \mathbb{R}^+ \\ S^2(x) &:= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\end{aligned}$$

verwendeten wir, um den Parameter σ^2 der Normalverteilung zu schätzen.

Die Statistik $\max\{X_1, \dots, X_n\}$ verwendeten wir, um bei der stetigen uniformen Verteilung mit den Parametern $[a, b]$ den Parameter b zu schätzen. Den Parameter a schätzten wir mittels des Nullpunktes der Geraden durch $(x_1, \frac{1}{n})$ und $(x_n, 1)$, wobei der Datensatz der Grösse nach geordnet wurde. Ist n genügend gross, können wir a auch durch die Statistik $\min\{X_1, \dots, X_n\}$ schätzen. Es geht in der Folge darum, dieses Verfahren im Nachhinein zu rechtfertigen. Dazu diskutieren wir ein paar Gütekriterien für Schätzstatistiken.

30.1.1 Erwartungstreue

Eine gute Schätzstatistik sollte *im Mittel* den zu schätzenden Wert treffen. Mathematischer formuliert: Der Erwartungswert der Schätzstatistik sollte mit dem zu schätzenden Wert zusammen-

treffen. Dies führt uns zum ersten Gütekriterium für Schätzstatistiken, der Erwartungstreue.

Definition 30.1.1. Eine Schätzstatistik T heisst erwartungstreu genau dann, wenn $E_{P^n}(T) = \theta_P$ (= grosses Teta, griechisches T ; Abkürzung für einen beliebigen Parameter einer Verteilung P ; $P^n := P_{(X_1, \dots, X_n)}$ = Produktmass auf kartesisches Produkt der Bilder der identisch verteilten und unabhängigen Stichprobenzufallsvariablen X_i).

Beispiel 30.1.2. Die Schätzstatistik $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ist ein erwartungstreuer Schätzer von $E(X_i)$. Denn $E_{P^n}(\bar{X}) = E(X_i)$ (i.i.d. X_i), wie wir im Unterkapitel „Statistiken“ gezeigt haben (s. Seite 218). $\diamond$

Da $\bar{X}$ ein erwartungstreuer Schätzer von $E(X_i)$ ist, können wir das arithmetische Mittel der Stichprobendaten verwenden, um $E(X_i)$ zu schätzen (im „Mittel“ treffen wir ja dann diesen Wert. Wissen wir nämlich nichts über die Verteilung einer Grösse, wo wählen wir am besten jeweils das Mittel. Dann liegen wir im Durchschnitt am wenigsten daneben).

Satz 30.1.3. Die Schätzstatistik $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ($S^2 : \mathfrak{X}_i^n \rightarrow \mathbb{R}$) ist ein erwartungstreuer Schätzer von $V(X_i)$. Denn $E_{P^n}(S^2) = V(X_i)$

Beweis. Es gilt:

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2 \sum_{i=1}^n X_i\bar{X} + \sum_{i=1}^n \bar{X}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2\frac{n}{n} \sum_{i=1}^n X_i\bar{X} + n\bar{X}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2n\bar{X}\bar{X} + n\bar{X}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \end{aligned}$$

Im folgenden setzen wir $E_P(X_i) = 0$, da die Varianz von der Lage der Verteilung unabhängig ist. Zudem verwenden wir Theoreme aus dem Unterkapitel „Statistiken“:

$$\begin{aligned} E_{P^n}(S^2) &= E_{P^n} \left(\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) \right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E_P(X_i^2) - nE_{P^n}(\bar{X}^2) \right) \end{aligned}$$

Da $E_P(X_i) = 0$, gilt

$$E_P(X_i^2) = E_P \left[(X_i - E(X_i))^2 \right],$$

$$\begin{aligned} E_P \left[(X_i - E_P(X_i))^2 \right] &= E_P(X_i^2 - 2X_i E_P(X_i) + (E_P(X_i))^2) \\ &= \underset{\text{mit } E_P(X_i)=0}{=} E_P(X_i^2) \end{aligned}$$

Nun gilt:

$$E_P \left[(X_i - E(X_i))^2 \right] = V_P(X_i).$$

Somit ist

$$\sum_{i=1}^n E_P(X_i^2) = nV_P(X_i)$$

Da $E_{P^n}(\bar{X}) = E_P(X_i) = 0$, gilt:

$$E_{P^n}(\bar{X}^2) = E_P \left[(\bar{X} - E(\bar{X}))^2 \right] = V_{P^n}(\bar{X}) = \frac{V_P(X_i)}{n}$$

Somit ist dann:

$$nE_{P^n}(\bar{X}^2) = n \frac{V_P(X_i)}{n} = V_P(X_i)$$

Deshalb gilt:

$$\begin{aligned} \frac{1}{n-1} \left(\sum_{i=1}^n E_P(X_i^2) - nE_{P^n}(\bar{X}^2) \right) &= \frac{1}{n-1} (nV_P(X_i) - V_P(X_i)) \\ &= \frac{1}{n-1} (V_P(X_i)(n-1)) = V_P(X_i). \end{aligned}$$

□

Somit ist die Varianz einer Stichprobe ein erwartungstreuer Schätzer der Varianz der theoretischen Verteilung. Dieser Beweis stellt eine Rechtfertigung für die Definition der Varianz mit $\frac{1}{n-1}$ dar. Bei einer Definition mit $\frac{1}{n}$ wäre ein Nachweis der Erwartungstreue der Varianz nicht möglich. Gilt $E_{P^n}(T) \neq \theta_P$, so wird $E_{P^n}(T) - \theta_P$ „systematischer Fehler“ oder „Bias“ genannt. In der Praxis nennen wir auch den konkreten Wert, den die Schätzfunktion annimmt, einen Schätzer von θ_P .

Beispiel 30.1.4. Die Schätzstatistik „Median“

$MEDIAN : \mathfrak{X}_i^n \rightarrow \mathbb{R};$

$$MEDIAN(x) = \begin{cases} x'_k & \text{für } k = 0.5(n+1) \in \mathbb{N}^* \\ x'_{k+} \frac{x'_{k+1} - x'_k}{2} & k = \text{Zahl vor dem Komma von } 0.5(n+1) \end{cases}$$

für den geordneten Datensatz $(x'_1, \dots, x'_n)$ von $(x_1, \dots, x_n)$ ist im allgemeinen kein erwartungstreuer Schätzer für $MEDIAN(X_i) = x$ in $P_{X_i}(x) = 0.5$. ◇

30.1.2 Effizienz

Neben der Erwartungstreue sollte eine gute Schätzstatistik eine möglichst kleine Streuung aufweisen. Damit ist das Risiko, weit neben dem richtigen Wert zu liegen, kleiner. Dies führt uns zum zweiten Gütekriterium für gute Schätzstatistiken.

Definition 30.1.5. Gegeben seien zwei erwartungstreue Schätzstatistiken T_1 und T_2 eines Parameters θ_P . T_1 ist effizienter als T_2 genau dann, wenn $V_{P_n}(T_1) \leq V_{P_n}(T_2)$.

Dies bedeutet, dass eine effizientere Schätzfunktion eine kleinere oder gleich grosse Irrtumswahrscheinlichkeit aufweist! Gibt es unter den erwartungstreuen Schätzfunktionen eine, die am effizientesten ist, nennt man diese „effizient“. Effizienzbeweise setzen einen mathematischen Apparat voraus, den wir nicht zur Verfügung haben. Wir liefern einfach ein paar Beispiele für effiziente Schätzstatistiken.

Beispiel 30.1.6. Gilt $X_i \sim B(1, p)$, dann ist $\bar{X} = \frac{1}{n} \sum X_i$ eine effiziente Schätzstatistik für p (= Erwartungswert von X_i).

Gilt $X_i \sim B(1, p)$. Dann ist $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ eine effiziente Schätzstatistik für $p(1-p)$ (= Varianz von X_i).

Gilt $X_i \sim N(a, \sigma^2)$, dann ist $\bar{X} = \frac{1}{n} \sum X_i$ eine effiziente Schätzstatistik für a .

Gilt $X_i \sim N(a, \sigma^2)$, dann ist $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$ eine effiziente Schätzstatistik für σ^2 . $\diamond$

Zu beachten ist, dass man entsprechende Ergebnisse für den Erwartungswert und die Varianz nicht für beliebige Verteilungen herleiten kann.

30.1.3 Konsistenz

Eine Schätzstatistik kann zwar erwartungstreu sein. Wir wissen damit aber noch nicht, wie nahe der empirische Mittelwert einer Stichprobe dem zu schätzenden Parameter kommt. Eine gute Schätzstatistik hat die Eigenschaft, bei grösseren Stichproben mit hoher Wahrscheinlichkeit näher beim zu schätzenden Parameter zu liegen als bei kleineren Stichproben. Diese Eigenschaft nennt man Konsistenz:

Definition 30.1.7. Eine Schätzstatistik T_n , in deren Berechnung n Stichprobenzufallsvariablen X_i einfließen, ist konsistent genau dann, $P(\lim_{n \rightarrow \infty} (T_n - \theta_P) = 0) = 1$.

Beispiel 30.1.8. Die Schätzstatistik $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ (X_i i.i.d.) ist ein konsistenter Schätzer von $E_P(X_i)$. Dies folgt aus einem Ergebnis im Unterkapitel „Statistiken“. Dort haben wir gezeigt, dass $V_{P^n}(\bar{X}) = \frac{V_P(X_i)}{n}$ (s. Seite 223). Damit nimmt die Varianz der Schätzstatistik mit steigendem n ab, d.h. die Wahrscheinlichkeit mit $\bar{X}$ weit weg von $E_P(X_i)$ zu liegen, nimmt mit steigendem n ab.

Die Schätzstatistik $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ ist ein konsistenter Schätzer von $V(X_i)$. Ebenso ist

$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ ein konsistenter Schätzer von $V(X_i)$ (dieser Schätzer ist jedoch weder erwartungstreu noch ist er im Falle normalverteilter Zufallsvariablen effizient). $\diamond$

Damit kann ein Schätzer konsistent sein, ohne erwartungstreu zu sein. Ist ein Schätzer aber konsistent, so kann er bei genügend grossem n problemlos verwendet werden.

Die Konsistenz der Schätzstatistiken *Mittelwert* und *Varianz* erlauben es, eine Begründung dafür zu liefern, dass wir im Einstichproben- z -Test mit der Teststatistik $Z = \frac{\bar{X} - \bar{x}_0}{\frac{s_0}{\sqrt{n}}}$ den Erwartungswert $E(X_i) = \bar{x}_0$ durch den Mittelwert aus grossen Datensätzen schätzen können. Da die empirische Grundgesamtheit sehr gross ist, gilt auf Grund der Konsistenz, dass $\bar{x}_0 - E(X_i)$ sehr klein ist - mit sehr hoher Wahrscheinlichkeit. Analoge Überlegungen gelten für s_0^2 .

Ohne die jeweiligen Beweise erfolgen ein paar Bemerkungen zum Verhältnis der Schätzer „Median“ und „Mittelwert“. (i) Die Medianschätzfunktion unterscheidet sich von der Mittelwertschätzfunktion durch den Umstand, dass erstere bei normalverteilten Daten eine grössere Varianz hat (wir haben ja oben erwähnt, dass bei normalverteilten Daten der Mittelwert ein Schätzer mit Minimalvarianz ist). Der Median ist als Schätzer somit weniger effizient als das arithmetische Mittel, sofern die Daten normalverteilt sind. Der Median ist bei asymmetrischen Verteilungen nicht erwartungstreu, ist jedoch konsistent, d.h. je grösser der Datensatz, desto kleiner ist die Wahrscheinlichkeit, dass der Wert der Schätzstatistik „Median“ stark vom $MEDIAN_{X_i}$ abweicht. Der Mittelwert hat den

Nachteil, sehr sensitiv auf Ausreisser zu reagieren (nicht robust). Der Median ist ein sehr robuster Schätzer. Ist die Normalverteilungsannahme nicht gegeben, ist der Mittelwert oft nicht effizienter als der Median.

Es gibt verschiedene Methoden, um für Verteilungen oder andere Modelle Parameter zu schätzen. Wir werden im Rahmen der Regressionsgleichung die Methode der kleinsten Quadrate kennenlernen. Gebräuchlich ist zudem die Maximum-Likelihood-Methode, die im nächsten Unterabschnitt eingeführt wird.

30.1.4 Berechnung von Schätzstatistiken (fakultativ)

Bisher haben wir die Eigenschaften vorhandener Schätzer untersucht. Wir analysierten etwa die Eigenschaften der Schätzstatistik „Mittelwert“. Wie können wir aber einen Schätzer für einen Parameter θ finden, wenn wir keinen Hinweis darauf haben, wie dieser Schätzer aussehen könnte. Es gibt im wesentlichen zwei Methoden, um Schätzer zu berechnen: die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode. Die Methode der kleinsten Quadrate werden wir im Rahmen der Regression behandeln.

Die Maximum-Likelihood-Methode führt nicht immer zu erwartungstreuen Schätzern. Deshalb sollten durch sie gefundene Schätzer nur für genügend grosse Datensätze verwendet werden - ausser wir können im Nachhinein die Erwartungstreue beweisen.

Die Maximum-Likelihood-Methode (auf deutsch bedeutet dieser Ausdruck etwa: die Methode der maximalen Plausibilität) wird jeweils in bezug auf einen Datensatz angewendet, der als fixiert gilt. Man verwendet die Wahrscheinlichkeitsfunktion (diskreter Fall) oder die Dichtefunktion (stetiger Fall). Wir nennen in diesem *Unterkapitel 30.1.4* auch die Wahrscheinlichkeitsfunktion „Dichtefunktion“.

Sei f_i die Dichtefunktion der nach P verteilten Zufallsvariable X_i (Die zu bestimmenden Parameter von P sind unbekannt). Man kann zeigen, dass die gemeinsame Verteilung der n i.i.d. Zufallsvariablen X_i durch die Dichtefunktion $\prod_{i=1}^n f_i$ dargestellt werden kann. Setzen wir nun die

Daten in die Dichtefunktion $\prod_{i=1}^n f_i$ ein oder fixieren wir die Datenvariablen, so bleiben in diesem Ausdruck als Unbekannte einzig die Parameter der Dichtefunktion. Es ist nun plausibel, die Parameter so zu bestimmen, dass $\prod_{i=1}^n f_i$ maximal wird. Die Dichtefunktion sollte dort am meisten Fläche zwischen sich und der x-Achse aufweisen, wo Daten vorliegen: wo die Dichtefunktion ein Maximum annimmt, ist nämlich die Wahrscheinlichkeit des Eintretens von Daten maximal. Die Parameter, für welche die Dichtefunktion ihr Maximum annimmt, verwenden wir als Schätzer der Parameter der gesuchten Verteilung P .

Konkret sei a ein zu bestimmender Parameter. Dann berechnen wir das Maximum der Funktion $L(a) = \prod_{i=1}^n f_i(x_i)$ für feste x_i und festes n (a kommt in der Formel von f_i vor). Gewöhnlich wird

dieser Ausdruck logarithmiert, um die Berechnung zu vereinfachen: $\ln(L(a)) = \sum_{i=1}^n \ln(f_i(x_i))$. Man

kann zeigen, dass die logarithmierte Funktion an der selben Stelle Extrema annimmt wie die nicht logarithmierte Funktion (wenn Sie überhaupt im Inneren des betrachteten Definitionsbereichs von f_i Extrema annimmt!!). Wir müssen damit mit Hilfe der Differentialrechnung das Maximum von $\ln(L(a))$ bestimmen (erste und zweite Ableitung!).

Beispiel 30.1.9. Wir haben bisher das arithmetische Mittel der Stichprobendaten x_i verwendet, um den Parameter λ einer Poisson-Verteilung zu schätzen. Wir können dies auch mit Hilfe eines ML-Schätzers rechtfertigen. Wir haben n Stichprobendaten x_i mit $\sum_{i=1}^n x_i > 0$. Die Dichtefunktion (= Wahrscheinlichkeitsfunktion) der Poissonverteilung ist

$$f(x_i) = e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}.$$

Damit gilt es

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$$

zu maximieren (für feste x_i). Wir logarithmieren:

$$\begin{aligned} \ln L(\lambda) &= \sum_{i=1}^n \ln \left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!} \right) \\ &= \sum_{i=1}^n (-\lambda \ln e + x_i \ln \lambda - \ln x_i!) \\ &= \sum_{i=1}^n (-\lambda + x_i \ln \lambda - \ln x_i!). \end{aligned}$$

Wir differenzieren $\ln L(\lambda)$. Da $\ln(x_i!)$ eine Konstante ist, fällt diese beim Differenzieren in jedem Summanden weg. Mit

$$\begin{aligned} \frac{d}{d\lambda}(-\lambda) &= -1 \\ \frac{d}{d\lambda}(x_i \ln \lambda) &= \frac{1}{\lambda} x_i \end{aligned}$$

für alle x_i gilt damit:

$$\begin{aligned} \frac{d \ln(L(\lambda))}{d\lambda} &= \sum_{i=1}^n \left(\frac{1}{\lambda} x_i - 1 \right) \\ &= \frac{1}{\lambda} \sum_{i=1}^n x_i - \sum_{i=1}^n 1 \\ &= \frac{1}{\lambda} \sum_{i=1}^n x_i - n \end{aligned}$$

In einem Extremum muss dieser Ausdruck 0 sein:

$$\begin{aligned} \frac{1}{\lambda} \sum_{i=1}^n x_i - n &= 0 \\ \iff \frac{1}{\lambda} \sum_{i=1}^n x_i &= n \\ \iff \lambda &= \frac{1}{n} \sum_{i=1}^n x_i \end{aligned}$$

Wir überprüfen, ob dort ein Maximum oder ein Minimum vorliegt:

$$\frac{d}{d\lambda} \left(\frac{1}{\lambda} \sum_{i=1}^n x_i - n \right) = -\frac{1}{\lambda^2} \sum_{i=1}^n x_i$$

Wir setzen die Nullstelle der ersten Ableitung ein:

$$-\frac{1}{\left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2} \sum_{i=1}^n x_i = -\frac{1}{\frac{1}{n^2} \sum_{i=1}^n x_i} < 0,$$

da $\sum_{i=1}^n x_i > 0$ und $\frac{1}{n^2} > 0$.

Damit liegt bei $\frac{1}{n} \sum_{i=1}^n x_i$ tatsächlich ein Maximum der Maximum-Likelihood-Funktion $L(\lambda)$ vor. Deshalb ist laut der Maximum-Likelihood-Methode das arithmetische Mittel ein plausibler Schätzer von λ . $\diamond$

Die Maximum-Likelihood-Methode spielt eine wichtige Rolle in der multivariaten Statistik (bei den verallgemeinerten linearen Modellen: Probit-Regression, logistische Regression, etc.). Die einzelnen Berechnungen können algebraisch aufwendig sein.

30.2 Nichtparametrisches Schätzen von Verteilungen

30.2.1 Empirische Verteilungsfunktion als Schätzer

Bei der oben eingeführten Schätzmethode von theoretischen Verteilungen nahmen wir jeweils zuerst eine Verteilungsfamilie als gegeben an und bestimmten dann die Verteilung mit Hilfe einer Schätzung der Parameter. Zur Bestimmung der Verteilungsfamilie gelangten wir mit Hilfe einer graphischen Darstellung der empirischen Verteilungsfunktion der Daten (Punkteform der empirischen Verteilungsfunktion). Dieses Vorgehen ist jedoch nur gerechtfertigt, wenn die empirische Verteilungsfunktion als Schätzer der Verteilungsfamilie gerechtfertigt ist. Dies ist in der Tat der Fall, da die empirische Verteilungsfunktion (in der Form der Treppenfunktion) ein *konsistenter* Schätzer der theoretischen Verteilungsfunktion ist, d.h. je grösser die Stichprobe, die wir betrachten, desto besser passt die empirische Verteilungsfunktion (Treppenfunktion) zur theoretischen Verteilungsfunktion. Dies wurde bereits im Kapitel 13 (Schätzung von Wahrscheinlichkeitsverteilung durch Daten) gezeigt. Wir führen nun noch einen Test ein, der überprüft, ob die Verteilungsfunktion gut zur empirischen Verteilungsfunktion passt. Es handelt sich um den Kolmogorov-Smirnov-Test.

Anpassungstest für Verteilungsfunktionen - der Kolmogorov-Smirnov-Test

Ziel: Es soll überprüft werden, ob die Verteilungsfunktion zur empirischen Verteilungsfunktion der Daten passt.

Voraussetzungen: stetig, metrisch skalierte Daten. i.i.d.

Im Testwert verwenden wir unter Anderem den Betrag des maximalen Abstands der empirischen Verteilungsfunktion F_e^n (Treppenform) der geordneten Daten $(x_1, \dots, x_n)$

$$F_e^n = \begin{cases} \frac{i}{n} & \text{für } x \in [x_i, x_{i+1}[\text{ und } 1 \leq i \leq n-1 \\ 1 & \text{für } x \geq x_n \\ 0 & \text{für } x < x_1 \end{cases}$$

von der theoretischen Verteilungsfunktion.

Zur Wiederholung die folgende Abbildung 30.2.1:

Da wir die maximale Abweichung der empirischen Verteilungsfunktion vom Modell berechnen wollen, berechnen wir für jedes Datum x_i die Werte $|F(x_i) - \frac{i}{n}|$ und $|F(x_i) - \frac{i-1}{n}|$ und das Maximum

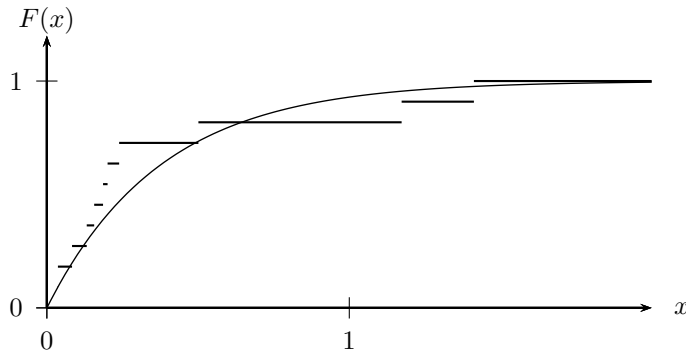


Abbildung 30.2.1: Empirische Verteilungsfunktion und Verteilungsfunktion der Verteilung, laut der die Daten verteilt sind

aller dieser $2n$ Wert. Wir verwenden als Teststatistik

$\sqrt{n}\Delta_n \sim G$
$n =$ Anzahl Daten
$\Delta_n := \max\{ F(x_i) - \frac{i}{n} , F(x_i) - \frac{i-1}{n} \} : x_i \text{ ist das } i\text{-te Datum des}$
geordneten Datensatzes $(x_1, \dots, x_n)$ (d.h. $x_i \leq x_{i+1}$)
$F =$ theoretische Verteilung, deren Anpassung überprüft wird.
Verteilungsfunktion von G : $F_G(t) = \begin{cases} \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 t^2) & \text{für } t > 0 \\ 0 & \text{sonst} \end{cases}$

Da die Summanden für sinkendes und steigendes k schnell sehr klein werden, genügt es, um die Werte der Verteilungsfunktion zu berechnen, mit Excel z.B.

$$F(t) \approx \sum_{k=-200}^{200} (-1)^k \exp(-2k^2 t^2)$$

zu berechnen (siehe Tabelle Kolmogorov-Smirnov-Berechnung.xls).

Für die obige Abbildung 30.2.1 (für die Daten s. Excel-Tabelle) erhalten wir als maximale Differenz 0.258411092 und als Testwert

$$\sqrt{n} \cdot \Delta_n = \sqrt{110.258411092} = 0.857\,05$$

sowie unter Verwendung der Tabelle Kolmogorov-Smirnov-Berechnung.xls den p-Wert: $0.454672311 > 0.05$. Die Daten können als zum Modell passend betrachtet werden.

Bemerkung 30.2.1. Es sollte nie auf Grund von Anpassungstests allein entschieden werden, ob Daten zu einem Modell passen oder nicht. Eine Graphik sollte immer zusätzlich begutachtet werden (Verteilungsfunktion, P-P-Plot, Q-Q-Plot). So kann man sehen, ob der Mangel an Anpassung vielleicht einem einzigen Ausreisser zuzuschreiben ist oder ob vielleicht doch systematische Abweichungen vorkommen, die allerdings im Einzeldatum nie die kritische Grenze überschreiten. Wird das Modell durch Schätzung aus den Daten berechnet, so kann man mit dem KS-Test nicht die Anpassung testen, da durch die Schätzung eine Überanpassung an die Daten resultiert. Passt das Modell nach Schätzung aber gemäss KS-Test nicht, so kann die Hypothese von der Angemessenheit des Modell verworfen werden. $\diamond$

30.2.2 Relative Häufigkeitsverteilung als Schätzer

Wir betrachten einen diskreten Zufallsvektor $(X_1, \dots, X_n)$ mit den Werten $x = (x_1, \dots, x_n)$. Sei

$W = \{x_i \mid x_i \text{ ist Komponente von } x\}$. Wir definieren die relative Häufigkeitsverteilung des Auftretens der x_i :

$$h : W \rightarrow [0, 1]$$

$$h(x_i) := \frac{|\{x_j \mid x_j = x_i, x_j \text{ ist Komponente von } (x_1, \dots, x_n)\}|}{n}$$

Wir ordnen also jeder möglichen Ausprägung die relative Häufigkeit ihres Auftretens zu.

Beispiel 30.2.2. Beim Ziehen aus einer Urne mit zwei möglichen Ausgängen (mit Zurücklegen; Binomial-Experiment, $n = 10, p = 0.5$) hätten wir bei 27-maliger Wiederholung folgende Resultate erhalten (s. Tabelle 30.2.1):

x_i	$H(x_i)$	$h(x_i)$	$P(X_i = x_i)$ für $X_i \sim B(10, 0.5)$
0	2	0.074074074	0.000976563
1	2	0.074074074	0.009765625
2	3	0.111111111	0.043945313
3	4	0.148148148	0.1171875
4	6	0.222222222	0.205078125
5	3	0.111111111	0.24609375
6	2	0.074074074	0.205078125
7	3	0.111111111	0.1171875
8	1	0.037037037	0.043945313
9	0	0	0.009765625
10	1	0.037037037	0.000976563
Summen	27	1	1

Tabelle 30.2.1:

Graphische Darstellung der Tabelle 30.2.1 (s. Abbildung 30.2.2):

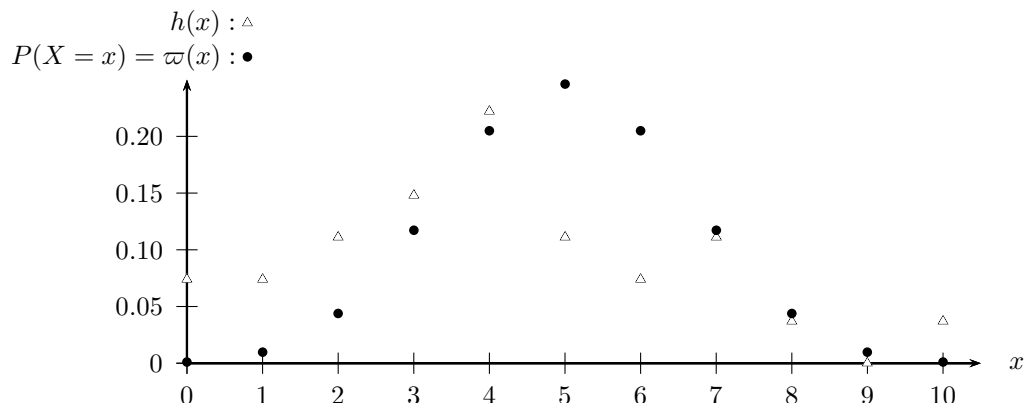


Abbildung 30.2.2: Graphische Darstellung der relativen Häufigkeiten und der geschätzten Wahrscheinlichkeitsfunktion (Binomialverteilung; $n = 10$; $p = 0.5$)

Im Beispiel ist die Übereinstimmung nicht besonders gut, da viel zu wenig Experimente betrachtet wurden. $\diamond$

Man könnte nun den Satz beweisen, dass für $m \rightarrow \infty$ die Wahrscheinlichkeit gegen 1 geht, dass die Abweichung der relativen Häufigkeitsverteilung von der Wahrscheinlichkeitsfunktion mit 0 identisch ist. Formal: Sei h_m die relative Häufigkeitsverteilung bei m -maligem Wiederholen des obigen binomischen Experimentes und f_X die zu schätzende Wahrscheinlichkeitsfunktion. Dann gilt:

$$P\{x_i : \lim_{m \rightarrow \infty} |h_m(x_i) - f_X(x_i)| = 0\} = 1.$$

Somit ist die relative Häufigkeitsverteilung ein konsistenter Schätzer der Wahrscheinlichkeitsfunktion, welche der Wahrscheinlichkeitsverteilung entspricht, laut der die Stichprobenvariablen X_i verteilt sind. Diese Aussage ist einer Aussagen ähnlich, die wir im Kapitel „Gesetz der grossen Zahlen“ gemacht haben, das (Würfelbeispiel und Annäherung an die uniforme Verteilung: je mehr man würfelt, desto mehr nähert sich die relative Häufigkeitsverteilung der theoretischen uniformen Wahrscheinlichkeitsfunktion mit $f_X(x) = \frac{1}{6}$ für alle x . Die Wahrscheinlichkeit, dass eine Abweichung von mehr als k vorliegt, geht gegen 0 für $n \rightarrow \infty$, n = Anzahl Würfelexperimente, s. Seite 226).

Auch hier haben wir uns bisher mit Hilfe von Graphiken davon überzeugt, ob Daten zu einer Verteilung passen (siehe Poissonverteilung). Es gibt allerdings auch ein Testverfahren, um dies zu überprüfen.

Der χ^2 -Anpassungstest

Ziel: Es soll die Anpassung eines Wahrscheinlichkeitsmodells (diskrete Wahrscheinlichkeitsverteilung) an die relative Häufigkeitsverteilung überprüft werden.

Voraussetzungen: Die erwarteten Häufigkeiten sind in mindestens 80% der Klassen grösser als 5. Die folgende Teststatistik ist näherungsweise chi-quadratverteilt:

$\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e}$ ist asymptotisch χ_{m-d-1}^2 -verteilt.
Sei Z eine Zerlegung von $\mathfrak{X}$. $Z = \{H_1, \dots, H_m\}$
$k_i^e = P(X \in H_i) \cdot n$ (absolute theoretische Häufigkeiten = erwartete absolute Häufigkeiten der Klasse i);
$(k_i^e \geq 5$ in mindestens 80% der Fälle)
K_i = Zufallsvariable, die die Anzahl k_i der Objekte $x_i \in K_i$ annimmt
n = Stichprobengrösse
m = Anzahl der Klassen (Elemente von Z).
d = Anzahl aus den Daten geschätzte Parameter

Bemerkung 30.2.3. (1) Werden d Parameter aus den Daten geschätzt, so muss d bei den Freiheitsgraden zusätzlich weggezählt werden. (z.B. Schätzen wir λ bei poissonverteilten Daten aus den Daten, so würde gelten: $\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e}$ ist asymptotisch χ_{m-2}^2 -verteilt; Schätzen wir bei der Binomialverteilung n durch die grösste Zahl, die in den Daten vorkommt und p durch $\bar{x}$, so würden wir berechnen $\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e}$ ist asymptotisch χ_{m-3}^2 -verteilt).

(2) Bei wenigen möglichen Ausprägungen der Zufallsvariable (z.B. $X_i \sim B(5, 0.5)$), wählen wir bei genügend vielen Daten oft als Zerlegung $\{\{x_i\} \mid x_i \in \mathfrak{X}\}$

(3) Der χ^2 -Test kann eine graphische Überprüfung der Anpassung nicht ersetzen (Gefahr von Ausreissern, die das Resultat verfälschen können). $\diamond$

Beispiel 30.2.4. Es wird überprüft, ob ein Würfel gut in dem Sinne ist, dass alle Seiten dieselbe Wahrscheinlichkeit haben, gewürfelt zu werden. Dazu werden 1000 Würfe gemacht. Es ergeben sich die folgenden Daten (s. Tabelle 30.2.2):

Augenzahl	Anzahl Auftreten
1	150
2	160
3	180
4	175
5	174
6	161
Summe	1000

Tabelle 30.2.2: Beispieldaten für Chi-Quadrat-Anpassungstest - Übereinstimmung von postulierter Wahrscheinlichkeitsverteilung und empirischer Häufigkeitsverteilung

H_0 : Die Daten passen zu einer diskreten uniformen Verteilung ($m = 6$).

H_A : Die Daten passen nicht zu einer diskreten uniformen Verteilung. ($\alpha = 0.05$)

Wir testen :

$$\begin{aligned}
 \sum_{i=1}^m \frac{(k_i - k_i^e)^2}{k_i^e} &= \frac{(\frac{1000}{6} - 150)^2}{\frac{1000}{6}} + \frac{(\frac{1000}{6} - 160)^2}{\frac{1000}{6}} + \frac{(\frac{1000}{6} - 180)^2}{\frac{1000}{6}} \\
 &\quad + \frac{(\frac{1000}{6} - 175)^2}{\frac{1000}{6}} + \frac{(\frac{1000}{6} - 174)^2}{\frac{1000}{6}} + \frac{(\frac{1000}{6} - 161)^2}{\frac{1000}{6}} \\
 &= 3.932.
 \end{aligned}$$

$$F_{\chi_5^2}(3.932; 5) = 1 - 0.44075 > 0.025$$

Die Daten passen zu einem guten Würfel. Wir halbieren das Signifikanzniveau, wenn wir Überanpassung auch als nicht zu den Daten passend qualifizieren. Interessiert uns dieser Fall nicht, würden wir einseitig auf dem 0.05-Niveau testen. $\diamond$

Beispiel 30.2.5. Wir wollen überprüfen, ob die Anzahl der Betriebe bezüglich bestimmter Versicherungsmeldungen pro Jahr und pro Betrieb bei 43 Betrieben poisson-verteilt ist ($\lambda = 8$). Es wurden folgende Daten erhoben (Schätztheorie.xls, s. Abbildung 30.2.3)

x_i	k_i
0	1
1	2
2	3
3	4
4	5
5	8
6	8
7	4
8	3
9	1
≥ 10	4

Tabelle 30.2.3: Daten, die auf Poissonverteilung ($\lambda = 7$) überprüft werden sollen

Die Bedingung, dass H_i^e in nicht mehr als 20% der Klassen kleiner als 5 ist, wird nicht erfüllt, wie Sie selber nachrechnen können. Es muss deshalb eine größere Klassierung vorgenommen werden (s. Tabelle 30.2.4):

x_i klassiert	k_i	$n \cdot P(X = xi) = k_i^e$	$\frac{(k_i - k_i^e)^2}{k_i^e}$
[0,5]	23	8.223150669	26.55372435
6	8	5.251943265	1.437908873
7	4	6.002220874	0.667900851
8	3	6.002220874	1.501665861
9	1	5.335307443	3.522738067
≥ 10	4	12.18515687	5.498229834
Summen	43(= n)	43	39.18216783

Tabelle 30.2.4: Berechnungstabelle für reklassierte Daten von Tabelle 30.2.3; $X \sim P(7)$

Wir erhalten als Testwert:

$$\sum_{i=1}^m \frac{(k_i - k_i^e)^2}{k_i^e} = 39.18216783$$

p -Wert: $1 - F_{\chi_5^2}(39.18216783) = 2.18253E - 07 < 0.05$

◇

Der p -Wert ist kleiner als 0.05. Damit sind die Daten nicht als Werte poisson-verteilter Zufallsvariablen zu betrachten (oder wenigstens nicht als Werte poisson-verteilter Zufallsvariablen mit dem Parameter $\lambda = 8$). Beim Betrachten der Tabelle sieht man sofort, dass es z.B. in der letzten Kategorie zu viele, in der ersten zu wenige Daten hat.

30.2.3 Histogramme als Schätzer

Wir betrachten zum Abschluss noch den stetigen Fall: Auch die Dichtefunktion einer stetigen Wahrscheinlichkeitsverteilung kann unter bestimmten Bedingungen geschätzt werden. Ein Histogramm ist definiert als eine Funktion, die auf Intervallen identischer Länge h jeweils konstant ist, so dass die intervallweisen Flächen zwischen der Funktion und der x -Achse der relativen Häufigkeit

der Daten in den Intervallen entspricht (allerdings werden Histogramme gewöhnlich mit Balken dargestellt, die dann keinen Funktionen mehr entsprechen würden). Formal: Sei h die Intervallbreite. Dann stellt $[ih, (i+1)h[$ mit $i \in \mathbb{Z}$ (= Menge der ganzen Zahlen) eine Zerlegung von $\mathbb{R}$ dar. Das Histogramm f^e kann nun definiert werden als:

Definition 30.2.6.

$$\begin{aligned} f^e : \mathbb{R} &\rightarrow [0, 1] \\ f^e(x) &:= \frac{1}{h} \frac{|\{x_j : x_j \in [ih, (i+1)h[\}|}{n} \\ &\text{für } x \in [ih, (i+1)h[\quad (\text{für alle } i \in \mathbb{Z}) \end{aligned}$$

Wir können das Histogramm als eine empirische Dichtefunktion auffassen. Entsprechend muss die Fläche zwischen dem Histogramm und der x-Achse 1 sein:

Satz 30.2.7. *Sei f^e durch Definition 30.2.6 bestimmt. Dann ist die Fläche zwischen f^e und der x-Achse mit 1 identisch.*

Beweis. Es gibt ein k so dass für alle $i > k$ gilt: $\frac{|\{x_j : x_j \in [ih, (i+1)h[\}|}{n} = 0$. Zudem gibt es ein l so dass für alle $i < l$ gilt: $\frac{|\{x_j : x_j \in [ih, (i+1)h[\}|}{n} = 0$. Für diese Intervall ist die Fläche zwischen dem Histogramm und der x-Achse mit 0 identisch. Wir brauchen sie somit nicht zu beachten. Zu berechnen ist nun die endliche Summe

$$\sum_{i=l}^k \left(h \cdot \frac{|\{x_j : x_j \in [ih, (i+1)h[\}|}{n} \right) = \sum_{i=l}^k \frac{|\{x_j : x_j \in [ih, (i+1)h[\}|}{n} = 1, \text{ da die Summe der relativen Häufigkeiten mit 1 identisch ist. } \square$$

Beispiel 30.2.8. *Wir betrachten den Datensatz x mit über 10'000 Daten (s. Tabellenblatt BHisto in Exceltabelle Schätztheorie.xls). Wir wählen als Klassenbreite 0.5, wobei wir mit -4.5 anfangen. Berechnen wir das Histogramm laut obigen Vorgaben und zeichnen wir die Dichtefunktion der Standardnormalfunktion drüber, so erhält man (s. Abbildung 30.2.3):*

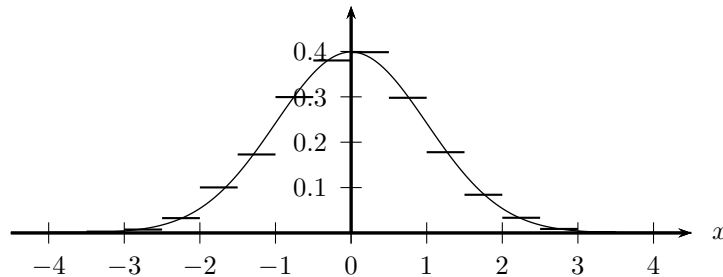


Abbildung 30.2.3: Histogramm Standardnormalverteilter Daten (s. Excelblatt BHisto in der Exceldatei Schätzungen.xls) und Dichtefunktion der Standardnormalverteilung

◇

Man sieht, dass sich die empirische Dichtefunktion der theoretischen nahe kommt (man könnte bei so viel Daten noch kleinere Intervalle wählen, dann wäre die Näherung noch besser). Man kann zeigen, dass die Wahrscheinlichkeit, dass das Histogramm (für $n \rightarrow \infty$ und immer kleineren Intervallen) von der Dichtefunktion der theoretischen Verteilung abweicht, beliebig klein wird - unter bestimmten Bedingungen (u.a. falls die Ableitung der Dichtefunktion der theoretischen Verteilung beschränkt ist und falls die Intervallbreite mit der Anzahl n der Daten passend verkleinert werden (z.B. so dass $h(n) = \frac{1}{\sqrt{n}}$)). Nach dieser Vorschrift könnte man für die Beispieldaten von Abbildung 30.2.3 die folgende Intervallbreite wählen: $\frac{1}{\sqrt{10405}} = 0.0098035 \approx 0.01$.

Bemerkung 30.2.9. Für fixiertes n stellt die Intervallbreite h einen wesentlichen Faktor dar, ob das Histogramm nahe bei der zu schätzenden Verteilung liegt oder nicht. $\diamond$

Auch die Anpassung von Histogrammen an die Dichtefunktion kann mit Hilfe eines Chi-Quadrat-Testes überprüft werden: Die folgende Teststatistik ist näherungsweise chi-quadrat-verteilt (Die erwarteten Häufigkeiten sind in mindestens 20% der betrachteten Intervalle grösser als 5).

$\sum_{i=1}^m \frac{(K_i - k_i^e)^2}{k_i^e}$ ist asymptotisch χ_{m-d-1}^2 -verteilt.
$k_i^e = P(X \in \text{Intervall } i) \cdot n$ (absolute theoretische Häufigkeiten = erwartete Häufigkeiten des Intervalls i)
K_i = Zufallsvariable, welche die absoluten empirische Häufigkeiten k_i des Intervalls i annimmt
n = Stichprobengrösse
m = Anzahl der betrachteten Intervalle
(die Intervalle l , so dass $\frac{ \{x_j: x_j \in [ih, (i+1)h]\} }{n} = 0$ für alle $i < l$ gilt und die Intervalle k , so dass $\frac{ \{x_j: x_j \in [ih, (i+1)h]\} }{n} = 0$ für alle $i > k$ fassen wir wir jeweils zu einem Intervall zusammen).
d = Anzahl aus den Daten geschätzte Parameter

Den Chi-Quadrat-Test können wir auch anwenden, indem wir unterschiedlich lange Intervalle verwenden. Allerdings hat die Wahl der Intervalle einen Einfluss auf das Ergebnis. Man kann hier u.U. das Resultate manipulieren.

Beispiel 30.2.10. Wir möchten testen, ob die Daten des Beispiels 30.2.8 normalverteilt sind (BHisto, Schätztheorie.xls). Wir nehmen folgende Klassierung vor: $[-4.5, -3]$, $]-3, -2]$, ..., $]2, 3]$, $]3, 4.5]$. Wir erhalten folgende Tabelle:

Intervall mit Grenzen x_i, x_j	k_i	$P(x_i < X \leq x_j)$ $X \sim N(0, 1)$	k_i^e	$\frac{(k_i - k_i^e)^2}{k_i^e}$
$]-\infty, -3]$	18	0.0013465	14.01033623	1.136119557
$]-3, -2]$	209	0.021400234	222.6694339	0.839151651
$]-2, -1]$	1422	0.135905122	1414.092794	0.044214852
$]-1, 0]$	3536	0.341344746	3551.692083	0.069330747
$]0, 1]$	3623	0.341344746	3551.692083	1.431661003
$]1, 2]$	1364	0.135905122	1414.092794	1.774486119
$]2, 3]$	220	0.021400234	222.6694339	0.032002046
$]3, +\infty]$	13	0.001349898	14.04568902	0.077850615
Summen	10405	1	10405	5.381962128

Wir ordnen die Daten, bestimmen Intervallgrenzen, zählen die Anzahl der Daten pro Intervall und berechnen die theoretischen Wahrscheinlichkeiten für die Intervalle. $\diamond$

$$1 - F_{\chi_4^2}(5.381962128) = 0.613456982 > 0.05$$

Damit ist die Normalverteilungsannahme haltbar.

Bemerkung 30.2.11. Bei stetig verteilten Daten ist die Verwendung des Kolmogorov-Smirnov-Testes dem Chi-Quadrat-Anpassungstest aus praktischen Gründen vorzuziehen, da der K-S-Test

gewöhnlich von Statistik-Programmen angeboten wird. Allerdings sollte letzterer nicht bei geschätzten Modellen Verwendung finden. $\diamond$

30.2.4 Kerndichteschätzer (fakultativer Stoff)

Neben Histogrammen kann man auch sogenannte Kerndichteschätzer für das Schätzen stetiger Dichtefunktionen verwenden. Während Histogramme als Treppenfunktion nicht stetig sind, sind Kerndichteschätzer stetige Schätzer. Wir gehen von der Dichtefunktion der Normalverteilung aus (etliche andere stetige Dichtefunktionen wären auch geeignet). Sei $x = (x_1, \dots, x_n)$ die Realisierung einer Stichprobe. Dann können wir die n Dichtefunktionen

$$f_{x_i}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x - x_i)^2\right)$$

betrachten. Diese Dichtefunktionen f_{x_i} können wir addieren und dann durch n teilen. Wir erhalten die Funktion

$$\bar{f}_x(x) = \frac{1}{n} \sum_{i=1}^n f_{x_i}(x).$$

Diese ist wiederum eine Dichte, da die Fläche unter der Summe der n Dichtefunktionen $= n \cdot 1 = n$ ist und damit die Fläche unter der mit $\frac{1}{n}$ gewichteten Summe $\sum_{i=1}^n f_{x_i}$ die Fläche $\frac{n}{n} = 1$ ergibt. Die berechnete Dichtefunktion nimmt an x jeweils als Wert das arithmetische Mittel der Werte der n Dichtefunktionen f_{x_i} an - für $x \in \mathbb{R}$. Wir nennen diese „mittlere“ Dichtefunktion „Kerndichteschätzer“.

Intuitiv können wir wie folgt plausibel machen, dass die mittlere Dichtefunktion als Schätzer der Dichte von P ($X_i \sim P$) geeignet ist. Die Dichtefunktionen f_{x_i} nehmen ihr Maximum jeweils in x_i an. Zudem weisen sie zwischen sich und der x-Achse am meisten Fläche in einer Umgebung von x_i auf. Je nachdem wie wir σ^2 festlegen, werden die Dichtefunktionen f_{x_i} mehr oder weniger stark bei den x_i zusammengestaucht und wir erhalten unterschiedliche Schätzer. Liegen die Punkte x_i in einem Intervall nahe beieinander, so werden sich viele Dichten in Bereichen überlagern, wo sie viel Fläche zwischen sich und der x-Achse aufweisen, und die mittlere Dichte wird dort relativ grosse Werte annehmen. Dies ist erwünscht: dort wo viele Daten anfallen, wird dem gesuchten Modell entsprechend die Wahrscheinlichkeit gross sein, dass dort Daten anfallen. Liegen die Punkte x_i in einem Intervall weit auseinander, so werden zu den grössten Werten einer Dichte nur kleine Werte anderer Dichten hinzugezählt. Diese Summe geteilt durch n ergibt dann relativ kleine Werte der mittleren Dichte (= Kerndichteschätzer). Wie stark sich die Dichten f_{x_i} überlagern, hängt einerseits von der Lage und der Anzahl der Daten ab und andererseits von der Wahl von σ^2 . Die Güte des Kerndichteschätzers hängt von diesen beiden Grössen ab (Bei der Verwendung des Schätzers und gegebenem n experimentiert man am besten mit der Grösse von σ^2 , um eine günstige Kurve zu erlangen).

Beispiel 30.2.12. (für die Überlagerung und die Abhängigkeit von der Varianz σ^2):

$$x = (3, 4, 4.5, 5, 10, 12, -2, -7)$$

Dichtefunktionen mit $\sigma = 0.5$:

$$\frac{1}{0.5\sqrt{2\pi}} \exp\left(\frac{1}{2 \cdot 0.5^2}(x - x_i)^2\right)$$

und das „Mittel“ dieser Dichtefunktionen (s. Abbildung 30.2.4):

$$\frac{1}{8} \sum_{i=1}^8 \frac{1}{0.5\sqrt{2\pi}} \exp\left(\frac{1}{2 \cdot 0.5^2}(x - x_i)^2\right)$$

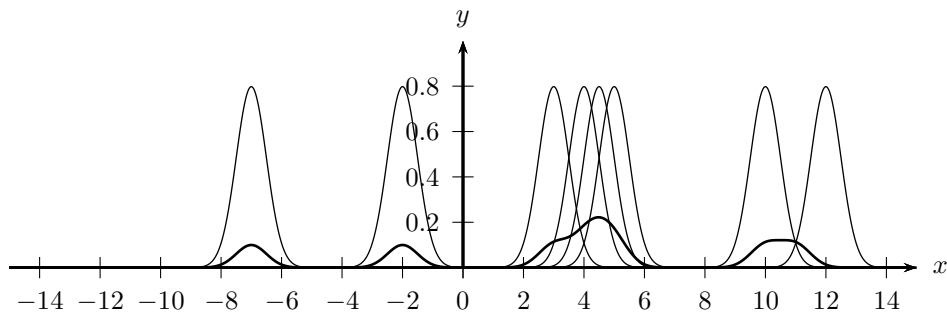


Abbildung 30.2.4: Dichtefunktionen mit $\mu = x_i \in \{-7, -2, 3, 4, 4.5, 5, 10, 12\}$ und $\sigma = 0.5$ und deren Mittel (fett; Kerndichteschätzer)

◇

Je grösser die Varianz, desto runder wird die Kurve bei wenigen Daten. Bei vielen Daten wird die Varianz kleiner gewählt, damit die Schätzung besser wird. Für $\sigma^2 = 4$ erhalten wir für die obigen Daten (s. Abbildung 30.2.5):

$$\frac{1}{8} \sum_{i=1}^8 \frac{1}{2\sqrt{2\pi}} \exp\left(\frac{1}{2 \cdot 4}(x - x_i)^2\right)$$

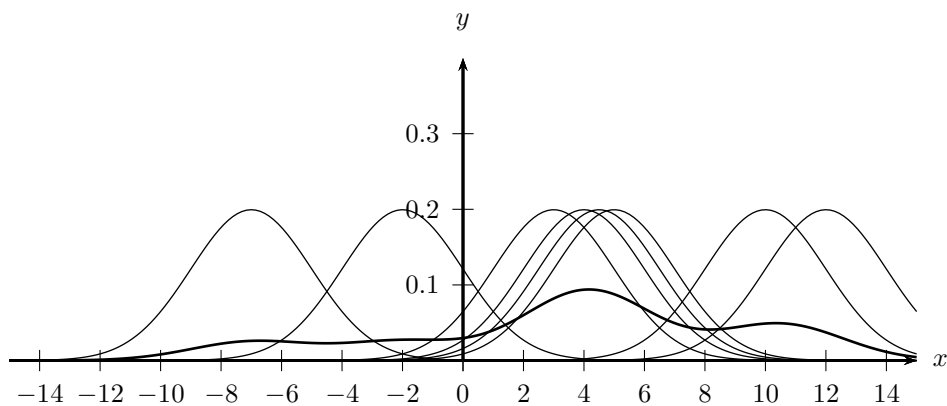


Abbildung 30.2.5: Dichtefunktionen mit $\mu = x_i \in \{-7, -2, 3, 4, 4.5, 5, 10, 12\}$ und $\sigma = 2$ und deren Mittel (fett; Kerndichteschätzer)

Auch bei Kerndichteschätzern gilt, dass bei Vorliegen bestimmter Bedingungen (u.a. Ableitung der Dichtefunktion ist beschränkt, angemessene Wahl des Streckungsfaktors h (bei Verwendung der Normalverteilung ist $h = \sigma$) in Abhängigkeit von n , z.B. $h_n = \frac{1}{\sqrt[4]{n}}$) die Wahrscheinlichkeit, dass die Differenz von Kerndichteschätzer und zu schätzender Funktion = 0 ist, gegen 1 geht.

30.2.5 Übungen

1. Wir nehmen an, im Oberwallis wisse die kinderreichste Familie 12 Kinder auf. Es gibt sich dabei für den Kinderreichtum der Familien die folgende Tabelle (so haben z.B. 706 Familien je 2 Kinder und 448 Familien haben keine Kinder; die Daten sind frei erfunden, s. Tabelle 30.2.5).

Anzahl Kinder	Anzahl Familien
0	448
1	823
2	706
3	359
4	133
5	13
6	3
7	2
8	0
9	0
10	0
11	0
12	1

Tabelle 30.2.5: Übungsdaten für eine Überprüfung auf Binomialverteilung

Überprüfen Sie, ob die Anzahl der Familien pro spezifische Kinderzahl binomialverteilt ist ($B(12; 0.13)$) ; Übung 1, Schätztheorie.xls)

- Die Unfallhäufigkeit in den KMUs der Schweiz weise folgende Verteilung auf: (frei erfundene Daten, (Übung 2, Schätztheorie.xls). Versuchen Sie eine Poisson-Verteilung an die Häufigkeiten der Betriebe pro Unfallhäufigkeit anzupassen und testen Sie die Angemessenheit.
- Überprüfen Sie die folgenden Daten auf Normalverteilung (Übung 3, Schätztheorie.xls) (Wenn Sie zwecks Überprüfung dieselben Klasseneinteilungen vornehmen wollen wie in der Antwort verwendet: $]-\infty, 20'000[$, $[20'000, 30'000[$, $[30'000, 40'000[$, $[40'000, 50'000[$, $[50'000, 60'000[$, $[60'000, \infty[$). Sollten die Daten nicht normalverteilt sein, versuchen Sie eine andere Verteilungsfunktion an die Daten anzupassen. Führen Sie den K-S-Test mit einer Statistiksoftware durch und vergleichen Sie das Resultat.
- Überprüfen Sie die folgenden Daten auf Exponentialverteilung (Übung 4, Schätztheorie.xls).
- Es wird behauptet, ein fallengelassener Fünffiber bleibe mit einer Wahrscheinlichkeit von 0.6 mit dem Wilhelm Tell nach unten liegen. Die These wird überprüft, indem ein Fünffiber 50 mal geworfen wird. Dabei bleibt er 25 mal mit dem Tell nach unten liegen, 25 mal mit der Zahl nach unten. Ist die These haltbar?
- Schätzen Sie für die Daten (Übung 6, Schätztheorie.xls) von Auge ein Modell. Machen Sie einen P-P-Plot.
- Erstellen Sie für die Daten (Schätztheorie.xls) ein Histogramm (Intervallbreite $h = 15$ und $h = 30$). Zeichnen Sie die theoretische Dichtefunktion hinzu, die Sie aus den Daten schätzen (Normalverteilung).

30.2.6 Lösungen

- Die Lösung finden Sie unter Tabelle 2, (Übung 1, Schätztheorie.xls). Der p-Wert ist 0 (12 Freiheitsgrade, da es 13 Ausprägungen von $X \sim B(12, p)$ gibt). Das Modell passt nicht. Dies

ist vor allem dem letzten Wert zuzuschreiben, der weitaus am meisten zum grossen Testwert beiträgt. Vermutlich müssten wir, um ein Modell anzupassen, den Ausreisser weglassen und es mit den restlichen Daten versuchen. Zu berechnen ist: $F_{\chi^2_{12}}(16.49924021) = 1 - (p\text{-Wert}) = 1 - 0.169424562$. Der p -Wert von 0.169424562 (Lösungsweg siehe Anpassungstests.xls) führt nicht zur Verwerfung des Modells. Das Beispiel zeigt die Bedeutung von Ausreissern.

2. Wir berechnen die durchschnittliche Unfallhäufigkeit pro Betrieb und setzen diese mit λ gleich: es gilt also $\lambda = 3.160287081$. Dann berechnen wir die theoretischen absoluten Häufigkeiten und den Chi-Quadrat-Testwert (siehe Schätztheorie.xls). Zu berechnen ist: $F_{\chi^2_{13}}(622.4127309) = 1 - 1.2737E-124$. Es gibt 15 Kategorien, somit verwenden wir 14 Freiheitsgrade. Da wir zusätzlich das λ aus den Daten schätzen, müssen wir einen zusätzlichen Freiheitsgrad abziehen. Der p -Wert ist: $1.2737E-124 < 0.05$. Die Daten sind nicht mit dem Modell verträglich.
3. Wir erhalten den Testwert: 13.12573332 und mit $df = 6 - 3$ (Anzahl Klassen minus 1 minus Anzahl geschätzter Parameter erhält man den p -Wert: $0.004372441 < 0.05$. Die Daten sind kaum normalverteilt (Lösungsweg: Tabelle 2; die erwartete Häufigkeit unterschreitet bei den gewählten Klassen in einem Intervall 5. Wir würden den Test trotzdem durchführen - nicht besonders starke Abweichung nach unten, weniger als 20% der Zellen sind betroffen). Ein Blick auf den P-P-Plot normal führt zur selben Antwort, s. Abbildung 30.2.6):

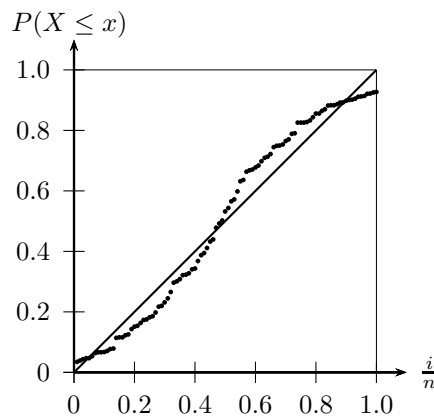


Abbildung 30.2.6: P-P-Plot der Daten bei Vermutung von Normalverteilung

Um eine passende Verteilungsfamilie zu finden, zeichnen wir eine Plot der Eckpunkte der empirischen, kumulativen, relativen Häufigkeitsverteilung (s. Abbildung 30.2.7)

Die Zeichnung lässt eher eine uniforme Verteilung oder eine polynomiale Kurve vermuten. Da recht viele Daten vorliegen, berechnen wir eine Gerade durch die zwei Punkte durch $(18337.93852, 0)$ und $(64806.36625, 1)$ berechnen (man könnte die Gerade auch durch $(18337.93852, 0.01)$ und $(64806.36625, 1)$ berechnen):

$$0 = 18337.93852a + b$$

$$1 = 64806.36625a + b$$

Wir erhalten: $a = 2.1520 \times 10^{-5}$, $b = -0.39463$

Dies ergibt die Gleichung: $y = 2.1520 \times 10^{-5}x - 0.39463$.

Dies ergibt die folgende Verteilungsfunktion:

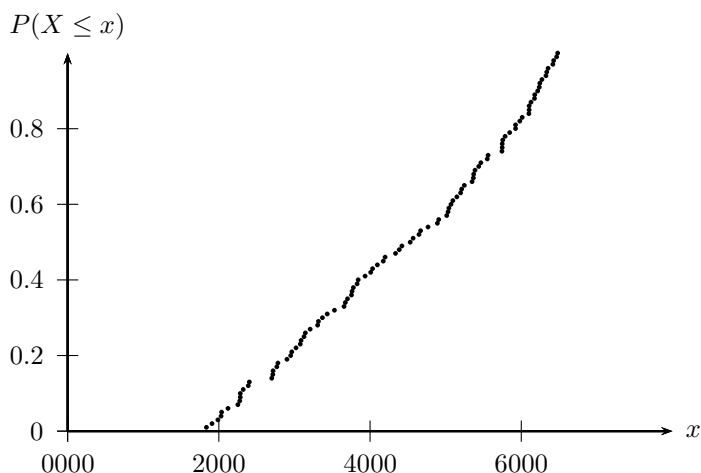


Abbildung 30.2.7: Empirische Verteilungsfunktion der Daten der Übung

$$F_X(x) = \begin{cases} 0 & \text{für } x \leq 18337.93852 \\ y = 2.1520 \times 10^{-5}x - 0.39463 & \text{für }]18237.93852, 64806.36625] \\ 1 & \text{für } x > 64806.36625 \end{cases}$$

Wir überprüfen das Modell: Wir erhalten bei der vorgenommenen Klasseneinteilung (s. Tabelle 30.2.6):

Klassen	$k(x_i)$	$k^e(x_i)$
$] -\infty, 30000]$	21	25.097
$]30000, 40000]$	20	21.52
$]40000, 50000]$	15	21.52
$]50000, 60000]$	26	21.52
$]60000, \infty[$	18	10.343
Summen	100	100

Tabelle 30.2.6: Vorbereitende Berechnungen für einen χ^2 -Test bei Vermutung einer stetigen uniformen Verteilung $X \sim U(18.237.938, 64806.366)$

den Testwert:

$$\frac{(21 - 25.097)^2}{25.097} + \frac{(20 - 21.52)^2}{21.52} + \frac{(15 - 21.52)^2}{21.52} + \frac{(26 - 21.52)^2}{21.52} + \frac{(18 - 10.343)^2}{10.343} = 9.3527$$

Es wurden 5 Klassen gewählt. Entsprechend ergeben sich 4 Freiheitsgrade. Zudem haben wir aus den Daten 2 Parameter geschätzt. Dies ergibt 2 Freiheitsgrade und damit den p-Wert: 0.009312944. Die Daten sind mit der konstruierten theoretischen Verteilung nicht verträglich. Die Kurve ist in der Tat gekrümmt. Man sollte es noch mit einer polynomialen Anpassung versuchen.

4. Zuerst muss man λ mit Hilfe der Daten schätzen. Dann werden Klassen gebildet: im Lösungsbeispiel (Tabelle 2): $[-\infty, 0.1[$, $[0.1, 0.3[$, $[0.3, 0.7[$, $[0.7, \infty[$
Testwert: 0.504700406; p-Wert (df = 2): 0.776972593 > 0.05 (4 Klassen, 3 df; da wir einen Parameter aus den Daten geschätzt haben, wird ein zusätzlicher Freiheitsgrad abgezählt).

Das Modell ist mit den Daten verträglich.

$$5. \frac{(25-30)^2}{30} + \frac{(25-20)^2}{20} = 2.0833$$

$$1 - F_{\chi^2_1}(2.0833) = 0.148917989$$

Das Ergebnis ist mit der Hypothese verträglich. Es wäre hier allerdings besser, einen Binomialtest durchzuführen. $P(X \leq 25)$ mit $X \sim B(50, 0.6)$. $\sum_{i=0}^{25} \binom{50}{i} 0.6^i 0.4^{50-i} = 0.097807$.

Die Nullhypothese wird auch mit dem Binomialtest nicht verworfen. Allerdings sieht man deutlich, dass der Binomialtest mächtiger ist.

6. Für Lösung siehe Excel-Tabelle.

7. Für Lösung siehe Excel-Tabelle.

30.3 Lernziele

- Die Problemstellung des Schätzens einer theoretischen Verteilung erläutern können.
- Die Grundideen der Begriffe „Erwartungstreue“, „Konsistenz“ und „Effizienz“ erläutern können.
- Situationen erkennen können, in denen Anpassungstests durchgeführt werden können.
- die Tests korrekt durchführen können (unter Prüfung der Bedingungen)
- Die empirische Dichtefunktion (= Histogramm) von Daten für gegebene Intervallbreite berechnen und zeichnen können.
- Übungen von der obigen Art durchführen können (auch für andere Verteilungen).

Kapitel 31

Stichprobentheorie

Oft will man eine Kennzahl einer Grundgesamtheit G auf Grund einer Stichprobe schätzen (z.B. in einer Umfrage möchte man den Anteil der Befürworter, Gegner und Unentschiedenen bezüglich einer Abstimmung oder die Interessenten eines Produktes oder einer Dienstleistung schätzen; man möchte das Durchschnittseinkommen einer Population kennen, usw.). Bei Schätzungen ist jeweils ein Wert anzugeben sowie ein Vertrauensintervall. Ergibt sich bei Meinungsumfragen, die zeitlich gestaffelt vorgenommen werden, ein Unterschied von 1.5% und beträgt das Vertrauensintervall für beide Werte $]\hat{\theta} - 2\%; \hat{\theta} + 2\%[$, wobei $\hat{\theta}$ der geschätzte Anteil an Befürwortern ist, so kann da offensichtlich kein Trend herausgelesen werden - ein häufiger Fehler bei der Interpretation von Umfrageergebnissen. Entsprechend sind Umfrageergebnisse ohne die Angabe von Vertrauensintervallen nicht sinnvoll.

Wir betrachten eine Population G mit N Objekten. Aus dieser wird eine *Zufallsstichprobe* $\mathfrak{S}$ (ohne Zurücklegen) gezogen. Dazu werden z.B. den Objekten der Population Zufallszahlen zugeordnet. Die Zufallszahlen werden dann der Grösse nach geordnet. Bei einer Stichprobengrösse von n wählen wir die n Objekte mit den n kleinsten Zufallszahlen.

Es gibt $\binom{N}{n}$ mögliche Stichproben $\mathfrak{S}_i$ der Grösse n , die auf diese Art gezogen werden können. Jede dieser Stichproben $\mathfrak{S}_i$ ist gleich wahrscheinlich. Entsprechend ist die Wahrscheinlichkeit, eine spezifische Stichprobe $\mathfrak{S}_i$ zu ziehen, $P(\mathfrak{S}_i) = \frac{1}{\binom{N}{n}}$ für $i \in \mathbb{N}_{\binom{N}{n}}^*$.

Die Wahrscheinlichkeit eines spezifischen Objektes x_j , Element der Stichprobe $\mathfrak{S}_i$ der Grösse n zu sein, kann ebenfalls angegeben werden. Es gibt $\binom{N-1}{n-1}$ Stichproben $\mathfrak{S}_i^{n-1}$ der Grösse $n-1$, die ohne x_j gebildet werden können. Entsprechend gibt es so viele Stichproben aus den N Objekten, die x_j enthalten, da x_j jede dieser $\binom{N-1}{n-1}$ Stichproben zu Stichproben der Grösse n auffüllen kann. Da jede dieselbe Wahrscheinlichkeit hat, gewählt zu werden, gilt

Satz 31.0.1. $P(\mathfrak{S}_i \mid \mathfrak{S}_i \subset G \text{ und } x_j \in \mathfrak{S}_i) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{n}{N}$

Beweis.

$$\begin{aligned} \frac{\binom{N-1}{n-1}}{\binom{N}{n}} &\stackrel{\text{Def. Bin.koeff.}}{=} \frac{\frac{(N-1)!}{(N-1-(n-1))!(n-1)!}}{\frac{N!}{(N-n)!n!}} \\ &= \frac{(N-1)!(N-n)!n!}{(N-1-(n-1))!(n-1)!N!} \\ &= \frac{(N-1)!(N-n)!n!}{(N-1-n+1)!(n-1)!N!} \\ &= \frac{(N-1)!(N-n)!n!}{(N-n)!(n-1)!N!} = \frac{(N-1)!(N-n)!n}{(N-n)!N!} \\ &= \frac{(N-1)!n}{N!} = \frac{n}{N} \end{aligned}$$

□

Bemerkung 31.0.2. Statistiken, die Stichproben Kennzahlen zuordnen, werden wir künftig „Schätzstatistik“ nennen. In der Stichprobentheorie werden oft Parameter der Population mit Grossbuchstaben zu bezeichnen und Parameter der Stichprobe mit Kleinbuchstaben. Oft wird zudem nicht zwischen Statistik und Wert der Statistik unterschieden - beides „Schätzer“ genannt. Hier werden wir Zufallsvariablen weiterhin mit Grossbuchstaben bezeichnen, die Werte der Zufallsvariablen mit Kleinbuchstaben sowie die Parameter der Grundgesamtheit mit kleinen griechischen Buchstaben. In der Stichprobentheorie werden die Messwerte der untersuchten Objekte als fix betrachtet - also nicht als Werte von Zufallsvariablen, wie etwa in der Testtheorie. In der Testtheorie ist die Statistik eine Zufallsvariable, die auf das kartesische Produkt von Werten von Zufallsvariablen definiert ist. Der Zufall kommt also mittels der einzelnen Zufallsvariablen hinein, auf deren Werte die Statistik definiert ist. In der Stichprobentheorie hingegen kommt der Zufall nur durch die Wahl der Stichprobe hinein, wodurch die Werte, welche die Schätzstatistik annimmt, zufällig werden. ◇

Sei $\bar{y}_i$ das Stichprobenmittel der metrisch skalierten Merkmale y_j der Stichprobe $\mathfrak{S}_i$. Dann ist $\bar{y}_i$ als Wert einer Zufallsvariable $\bar{Y}$ zu betrachten, deren Wert von der zufällig gezogenen Stichprobe abhängt. Wie verhält sich $\bar{Y}$ in bezug auf die Populationsparameter μ (= Mittelwert der Ausprägung y_j der Gesamtpopulation)? An einem Beispiel wollen wir zuerst zeigen, dass der Erwartungswert von $\bar{Y}$ mit μ übereinstimmt. Zudem ist die Standardabweichung von $\bar{Y}$ (= Standardfehler) kleiner als die Standardabweichung der Grundgesamtheit. Obwohl eigentlich die $x_j \in \mathfrak{S}_i$, schreiben wir auch $y_i \in \mathfrak{S}_i$.

Beispiel 31.0.3. Gegeben sei die Minipopulation $G = (a, b, c, d, e, f)$ mit den Ausprägungen $(y_1, y_2, y_3, y_4, y_5, y_6) = (3, 1, 10, 4, 8, 2)$. Der Mittelwert μ der Population bezüglich der Variable y ist dann:

$$\mu = \frac{3 + 1 + 10 + 4 + 8 + 2}{6} = 4.\bar{6}$$

Die Standardabweichung in der Population ist:

$$\sigma = \sqrt{\frac{(3 - 4.\bar{6})^2 + (1 - 4.\bar{6})^2 + (10 - 4.\bar{6})^2 + (4 - 4.\bar{6})^2 + (8 - 4.\bar{6})^2 + (2 - 4.\bar{6})^2}{6 - 1}} = 3.559$$

$$\sigma^2 = 3.559^2 = 12.666$$

Wir ziehen nun alle möglichen Stichproben $\mathfrak{S}_i$ der Grösse $n = 2$:

Es gibt $\binom{6}{2} = 15$ solcher Stichproben $\mathfrak{S}_i$.

Wir listen diese auf (samt ihrem jeweiligen Mittelwert $\bar{y}_i$ und der Varianz s_i^2 (s. Tabelle 31.0.1)

i	Elemente y_j von $\mathfrak{S}_i$; $n = 2$		Mittelwert $\bar{y}_i$ von $\mathfrak{S}_i$	Varianz s_i^2 von $\mathfrak{S}_i$
1	3	1	2	2
2	3	10	6.5	24.5
3	3	4	3.5	0.5
4	3	8	5.5	12.5
5	3	2	2.5	0.5
6	1	10	5.5	40.5
7	1	4	2.5	4.5
8	1	8	4.5	24.5
9	1	2	1.5	0.5
10	10	4	7	18
11	10	8	9	2
12	10	2	6	32
13	4	8	6	8
14	4	2	3	2
15	8	2	5	18
Mittelwert			4.666666667	12.66666667
Varianz			4.222222222	

Tabelle 31.0.1: Auflistung aller Stichproben der Grösse $n = 2$ aus der Grundgesamtheit $G = \{3, 1, 10, 4, 8, 2\}$ mit den Stichprobenmittel und Stichprobenvarianzen, sowie dem Mittel und der Varianz der Stichprobenmittel sowie dem Mittel der Stichprobenvarianzen

Wir stellen fest: der Erwartungswert

$$\begin{aligned}
 E(\bar{Y}) & \underset{\text{laut Def. Erwartungswert}}{=} \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \cdot \bar{y}_i \\
 &= \sum_{i=1}^{15} \frac{1}{\binom{6}{2}} \bar{y}_i = \frac{1}{15} \sum_{i=1}^{15} \bar{y}_i = 4.\bar{6} = \mu
 \end{aligned}$$

stimmt mit dem Populationsmittel μ überein. Mit anderen Worten: die Statistik Stichprobenmittel $\bar{Y}$ schätzt den Populationsmittelwert μ erwartungstreu. Der Erwartungswert der Zufallsvariable S^2 , die konkrete Stichprobenvarianzen s_i^2 als Werte annimmt, fällt im Beispiel mit der Varianz σ^2 der Grundgesamtheit zusammen:

$$E(S^2) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) s_i^2 = \frac{1}{\binom{6}{2}} \sum_{i=1}^{15} s_i^2 = 12.\bar{6} = \sigma^2.$$

Damit ist Zufallsvariable S^2 im Beispiel eine erwartungstreue Schätzstatistik von σ^2 . Die Varianz der Mittelwerte der Stichproben ist

$$V(\bar{Y}) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \cdot (\bar{y}_i - \mu)^2 = \frac{1}{15} \sum_{i=1}^{\binom{N}{n}} (\bar{y}_i - \mu)^2 = 4.222222$$

im Beispiel hingegen kleiner als die Varianz der Daten der Population. Wie man mittels einer Formel auf die Varianz der Mittelwerte kommt, wird in Kürze zu diskutieren sein. Wir formulie-

ren diese und weitere Resultate nun formal und leiten manche davon her. $\diamond$

31.1 Schätzen des Mittelwertes und des Totals einer Grundmenge

In Bezug auf die Statistik „Summe“ gilt:

Satz 31.1.1. Sei T_s die Zufallsvariable, welche als Werte die Summe der y_j einer Stichprobe $\mathfrak{S}_i$ annimmt. Dann gilt:

$$E(T_s) = \frac{n}{N} \sum_{j=1}^N y_j = n\mu$$

Dabei ist $\frac{1}{N} \sum_{j=1}^N y_j = \mu$ das arithmetische Mittel der Grundgesamtheit. Entsprechend ist $E(T_s)$ das n -fache ($n = \text{Stichprobengrösse}$) des Mittels der Grundgesamtheit. Für eine Beweis, s. S. 578. Sei $\bar{Y}$ die Zufallsvariable, die Stichprobenmittel als Werte annimmt. Der nächste Satz hält fest, dass $\bar{Y}$ eine erwartungstreue Schätzstatistik für das arithmetische Mittel μ der Grundgesamtheit ist (für einen Beweis s. S. 578):

Satz 31.1.2. Sei $\bar{Y}$ die Zufallsvariable, die Stichprobenmittel als Werte annimmt. Dann gilt:

$$E(\bar{Y}) = \mu$$

(μ Mittelwert der Gesamtpopulation).

Beispiel 31.1.3. Wir möchten den durchschnittlichen Weinkonsum einer Stadt erfassen (100'000 Einwohner). Dazu ziehen wir eine Zufallsstichprobe von 200 Personen und erheben ihren Weinkonsum (z.B. durch Befragung). Wir erhalten 200 Ergebnisse y_i . Wir bilden den Mittelwert $\bar{y}_i$ der 200 Ergebnisse y_i und können diesen als Wert einer erwartungstreuen Schätzstatistik für den durchschnittlichen Weinkonsum der 100'000 Einwohner betrachten. $\diamond$

Wir behandeln nun das Schätzen von Totalen. Sei τ die Summe der y_i , $i \in N$ (τ ist somit das Populationstotal der y_i , d.h. $\tau = \sum_{i=1}^N y_i$). Es gilt offensichtlich $\tau = N\mu$, denn $\mu = \frac{1}{N} \sum_{i=1}^N y_i$. $T := N\bar{Y}$ sei die Statistik, welche die Totale

$$t_i := \frac{N}{n} \sum_{j=1}^n y_j = N\bar{y}_i$$

der Stichproben $\mathfrak{S}_i$ annimmt. Das mit der Grösse der Grundgesamtheit multiplizierte Stichprobenmittel ist eine erwartungstreue Schätzstatistik T für das Total der Grundgesamtheit (für einen Beweis s. S. 578):

Satz 31.1.4. $E(T) = E(N\bar{Y}) = \tau$

($N = \text{Anzahl Objekte der Gesamtpopulation}$; $\bar{Y} = \text{Zufallsvariable mit den Mittelwerten der Stichproben der Grösse } n \text{ als Werte}$; $\tau = \text{Total der Gesamtpopulation}$).

Beispiel 31.1.5. Wir führen das vorangehende Beispiel 31.1.3 fort. Wir möchte wissen, wieviel der Gesamtweinkonsum der Stadt ist. Wir multiplizieren den geschätzten durchschnittlichen Weinkonsum mit 100'000 und erhalten den geschätzten totalen Weinkonsum. $\diamond$

31.1.1 Vertrauensintervalle für Schätzungen des Mittelwertes und des Totals

Um Vertrauensintervalle für das geschätzte Mittel zu berechnen, müssen wir die Varianz der Statistik Mittelwert kennen. Um die entsprechenden Formeln zu beweisen, werden zwei Hilfssätze 31.1.6 und 31.1.7 benötigt (für deren Beweise s. S. 578 und S. 579):

Satz 31.1.6. Sei Q die Zufallsvariable, welche als Werte $\sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2$ annimmt, dann gilt:

$$E(Q) = \frac{n}{N} \sum_{j=1}^N (y_j - \mu)^2$$

Satz 31.1.7. Sei K die Zufallsvariable, welche als Werte $k_i := \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu)$ annimmt.

Dann gilt:

$$E(K) = \frac{n(n-1)}{N(N-1)} \sum_{j=1}^N \sum_{\substack{m=1 \\ j \neq m}}^N (y_j - \mu)(y_m - \mu)$$

Nun zur Varianz der Statistik Mittelwert (für einen Beweis des Satzes s. S. 579):

Satz 31.1.8. $V(\bar{Y}) = \frac{N-n}{nN} \sigma^2$ ($= (\frac{1}{n} - \frac{1}{N}) \sigma^2 = \frac{1}{n} (1 - \frac{n}{N}) \sigma^2$)
 ($\bar{Y}$ = Zufallsvariable mit Stichprobenmitteln als Werte; $\sigma^2 := \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu)^2$ Varianz der Grundgesamtheit)

Bemerkung 31.1.9. In Beispiel 31.0.3 wurde die Varianz der Grundgesamtheit mit dem Faktor $\frac{1}{N-1}$ berechnet (N = Anzahl Objekte der Grundgesamtheit). Dieser Faktor wird im Kapitel „Schätztheorie“ durch den Umstand begründet, dass bei Stichprobenvarianzen die Berechnung mit diesem Faktor und mit i.i.d. Stichprobenvariablen Werte einer erwartungstreuen Schätzstatistik liefert. Da eine Grundgesamtheit keine Stichprobe ist und es entsprechend auch nicht darum geht, Schätzungen vorzunehmen, kann man argumentieren, dass der Faktor $\frac{1}{N}$ für Grundgesamtheiten mehr Sinn macht - durch ihn wird das exakte arithmetische Mittel der quadrierten Abweichungen berechnet. Dieser Standpunkt ist durchaus vertretbar und wird in manchen Büchern auch eingenommen. Die obige Formel für die Schätzung der Stichprobenvarianz des Schätzstatistik Mittelwert wird bei dieser Berechnung zu $V(\bar{Y}) = \frac{N-n}{n(N-1)} \sigma_v^2$ mit $\sigma_v^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2$, eine Formel, die in manchen Büchern geliefert wird. Es gilt

$$\frac{N-n}{n(N-1)} \sigma_v^2 = \frac{N-n}{n(N-1)} \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2 = \frac{N-n}{nN} \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu)^2 = \frac{N-n}{nN} \sigma^2 = \frac{1}{n} \left(1 - \frac{n}{N}\right) \sigma^2.$$

◇

Um zu einem praktikablen Verfahren für die Berechnung von Vertrauensintervallen zu gelangen, müssen wir bei Unkenntnis von σ^2 diese Grösse noch schätzen können. In der Tat kann gezeigt werden, dass die Stichprobenvarianz eine erwartungstreue Schätzstatistik der Varianz der Grundgesamtheit ist (für einen Beweis s. S. 581):

Satz 31.1.10. Sei S^2 die Zufallsvariable, die als Werte konkrete Stichprobenvarianzen $s_j^2 := \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_j)^2$ annimmt. Es gilt:

$$E(S^2) = \sigma^2$$

(σ^2 = Varianz der Grundgesamtheit).

Indem wir diese Ergebnisse zusammenfassen, erhalten wir den folgenden Satz (für einen Beweis s. S. 583):

Satz 31.1.11. $\frac{1}{n}\left(1 - \frac{n}{N}\right) S^2$ ist eine erwartungstreue Schätzstatistik der Varianz $V(\bar{Y})$ der Statistik Mittelwert $\bar{Y}$, d.h.

$$E\left(\frac{1}{n}\left(1 - \frac{n}{N}\right) S^2\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right) \sigma^2.$$

(n = Stichprobengrösse; N = Grösse der Gesamtpopulation; σ^2 Varianz der Gesamtpopulation).

Die Zufallsvariable S^2 nimmt im konkreten Fall den Wert

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_j)^2$$

= Varianz in der Stichprobe $\mathfrak{S}_j$ an.

Beispiel 31.1.12. Im betrachteten Beispiel 31.0.1 erhalten wir für die Varianz des Mittelwertes:

$$\frac{1}{2}\left(1 - \frac{2}{6}\right)12.667 = 4.2222$$

Überprüfen Sie die Übereinstimmung mit dem Ergebnis in der Tabelle (Achtung! $N = 6$, nicht $N = 15$. 15 ist die Anzahl der möglichen Stichproben, wenn $n = 2$ und $N = 6$). Die Standardabweichung des Mittelwertes - oft „Standardfehler des Mittelwertes“ genannt - ist $\sqrt{\frac{1}{2}\left(1 - \frac{2}{6}\right)12.667} = 2.0548$.
 $\diamond$

Der Faktor

$$\left(1 - \frac{n}{N}\right)$$

in der Berechnungsformel für die Varianz des Mittelwertes wird „Endlichkeitskorrektur“ genannt. In der Praxis kann er vernachlässigt werden, d.h. mit 1 identifiziert werden, wenn $\frac{n}{N} < 0.1$ ist (die Stichprobe im Verhältnis zur Grundgesamtheit klein ist). Beim Vernachlässigen der Endlichkeitskorrektur wird die Varianz der Schätzstatistik etwas überschätzt. Die Endlichkeitskorrektur zeigt zudem, dass beim Schätzen von Parametern einer Grundgesamtheit und bei $\frac{n}{N} < 0.1$ die Grösse der Grundgesamtheit kaum mehr eine Rolle spielt. Ob wir mit 1000 Daten den Mittelwert einer Grundgesamtheit von 12'000 Objekten schätzen oder einer Grundgesamtheit von 1'000'000, spielt bezüglich der Präzision der Schätzung keine nennenswerte Rolle mehr - ein erstaunliches Ergebnis. Umgekehrt braucht es bei sehr kleinen Grundgesamtheiten fast so grosse Stichproben wie bei sehr grossen Grundgesamtheiten, um einen gegebenen Präzisionsgrad zu erreichen. In der Folge werden jeweils nur die Formeln mit Endlichkeitskorrektur geliefert, um das Skriptum übersichtlicher zu halten.

Bei der Berechnung von Vertrauensintervallen geht man davon aus, dass die Zufallsvariable $\bar{Y}$ normalverteilt ist. Dies ist gerechtfertigt, da die Stichproben im Allgemeinen mehrere Hundert Daten umfasst. Die Normalverteilungsannahme ist jedoch nicht gerechtfertigt, wenn extrem schiefe Verteilungen mit extremen Ausreissern vorliegen (häufig bei Einkommensdaten oder Investitionen). Bei kleinen Stichproben wird die t-Verteilung verwendet ($n < 50$). Dies würde zwar nur gerechtfertigt sein, wenn die Daten normalverteilt sind. Allerdings sind leichte Abweichungen von dieser Annahme nicht schwerwiegend.

Zur Berechnung von β -Vertrauensintervallen können wir bei Voraussetzung einer Normalverteilung entwickeln:

$$\frac{\bar{Y} - \mu}{\sqrt{V(\bar{Y})}} \sim N(0, 1)$$

wobei $\sqrt{V(\bar{Y})}$ durch $\sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right) S^2}$ geschätzt wird. Für die Berechnung der Grenzen des Vertrauensintervalls setzen wir $\Phi\left(\frac{\bar{Y} - \bar{Y}_g}{\sqrt{V(\bar{Y})}}\right) = \frac{1-\beta}{2}$ sowie $\Phi\left(\frac{\bar{Y} - \bar{Y}_g}{\sqrt{V(\bar{Y})}}\right) = \beta + \frac{1-\beta}{2}$ und Lösen nach $\bar{Y}_{g_i}$ auf, erhalten (rechnen Sie das übungshalber nach):

$\bar{Y}_{g_i} = \bar{Y} \pm \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right) \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right) S^2}$ (mit Endlichkeitskorrektur)
β = Wahrscheinlichkeit, dass Grösse im Vertrauensintervall liegt (z.B. 0.95)
$\bar{Y}_{g_i}$ = Zufallsvariable „Intervallgrenzen“ mit konkreten Intervallgrenzen $\bar{y}_g$ des Vertrauensintervalls als Wert.
$\bar{Y}$ = Zufallsvariable mit den Stichprobenmittelwerten $\bar{y}_j$ als Werten.
S = Zufallsvariable „Standardabweichung“ mit der Standardabweichung der Stichprobe s_j als Wert.
n = Stichprobengrösse
N = Grösse der Grundgesamtheit
Φ^{-1} Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung

Beispiel 31.1.13. Wir wollen den durchschnittlichen Konsum elektronischer Unterhaltungsgüter einer Population schätzen (in Franken). Dazu soll ein Vertrauensintervall von 95% angegeben werden. Die Population weise 500'000 Personen auf. Es wird eine Zufallsstichprobe von 500 Personen gezogen mit einem Mittelwert von 700 und einer Stichprobenvarianz von 200^2 . Mit Hilfe der Formel erhalten wir durch Einsetzen in $\bar{Y}_{g_i} = \bar{Y} \pm \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right) \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right) S^2}$:

$$700 \pm 1.96 \cdot \sqrt{\frac{1}{500} \left(1 - \frac{500}{500000}\right) \cdot 200} = 682.48 \text{ für „-“ und } 717.52 \text{ für „+“}.$$

Das Vertrauensintervall ist somit:]682.48;717.52[.

◇

31.1.2 Berechnung der Stichprobengrösse beim Schätzen des Mittelwertes

Sofern wir aus früheren Daten die vermutliche Varianz der Grundgesamtheit schätzen können, können wir mit den obigen Formeln auch die Stichprobengrösse festlegen, die zu einer spezifischen Breite des Vertrauensintervalls führt. Mit $q := \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right)$ und der Hälfte d der Vertrauensintervalllänge gilt:

$$\begin{aligned} d &= q \sqrt{V(\bar{Y})} \\ &= q \sqrt{\frac{N-n}{nN} \sigma^2} \end{aligned}$$

Damit gilt dann, wie Sie selber nachrechnen können:

$n = \frac{N \sigma^2 q^2}{\sigma^2 q^2 + d^2 N}$ (mit Endlichkeitskorrektur)
d = die Hälfte der Breite des Vertrauensintervalls
σ = Standardabweichung der Gesamtpopulation (gewöhnlich geschätzt durch frühere Stichprobe)
n = Stichprobengrösse
$q := \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right)$
β = Wahrscheinlichkeit, dass Grösse im Vertrauensintervall liegt (z.B. 0.95)
Φ^{-1} Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung

Beispiel 31.1.14. Wir führen das Beispiel 31.1.13 fort. Wir nehmen an, wir würden nochmals eine Untersuchung vornehmen und möchten ein 95%-Vertrauensintervall mit einer Breite von ± 10 haben (oben betrug die Breite ± 17.52). Wie gross muss die Stichprobe sein?

Mit Hilfe der Formel mit Endlichkeitskorrektur:

$$n = \frac{\sigma^2 N q^2}{\sigma^2 q^2 + d^2 N} = \frac{200^2 \cdot 500000 \cdot 1.96^2}{200^2 \cdot 1.96^2 + 10^2 \cdot 500000} = 1531.9$$

Wir sehen, dass die Grösse der Stichprobe recht massiv erhöht werden muss, um das Vertrauensintervall in dieser Grössenordnung zu senken (Faustregel: um das Vertrauensintervall zu halbieren, muss man die Stichprobengrösse vervierfachen). $\diamond$

31.1.3 Vertrauensintervalle für Schätzungen des Totals

Bei der Berechnung des Vertrauensintervalls von Totalen gilt (für einen Beweis s. S. 583) :

Satz 31.1.15. $V(T) = V(N\bar{Y}) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) \sigma^2$

Wiederum schätzen wir σ^2 durch den Wert von S^2 .

Somit erhalten wir als Schätzstatistik für $V(T) = V(N\bar{Y})$:

$$\frac{N^2}{n} \left(1 - \frac{n}{N}\right) S^2$$

Wir können wiederum davon ausgehen, dass näherungsweise

$$\frac{T - Y}{\sqrt{V(T)}} \sim N(0, 1).$$

Wir erhalten (rechnen Sie dies übungshalber nach):

$Y_{gi} = T \pm \Phi^{-1}(\beta + \frac{1-\beta}{2}) \sqrt{\frac{N^2}{n} \left(1 - \frac{n}{N}\right) S^2}$ (mit Endlichkeitskorrektur)
Y_{gi} = Zufallsvariable „Intervallgrenze“ mit konkreten Intervallgrenzen als Wert.
$T = N\bar{Y}$ = Zufallsvariable „Total“ mit konkreten Totalen als Wert
S = Zufallsvariable „Standardabweichung“ mit den konkreten Stichprobenstandardabweichungen s_j als Wert.
N = Anzahl Objekte der Gesamtpopulation
n = Anzahl Objekte der Stichprobe
β = Wahrscheinlichkeit, dass Grösse im Vertrauensintervall liegt (z.B. 0.95)
Φ^{-1} Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung

Beispiel 31.1.16. Wir führen das Beispiel 31.1.13 fort. Wir wollen den Gesamtkonsum elektronischer Unterhaltungsgüter einer Population schätzen (in Franken). Dazu soll ein 95%-Vertrauensintervall angegeben werden. Die Population weise 500'000 Personen auf. Es wird eine Zufallsstichprobe von 500 Personen gezogen. Wir nehmen an, $\bar{y}_s = 700.-$ und $s_s = 200.-$. Wir berechnen: $\frac{500}{500000} = 0.001 < 0.1$. Wir könnten somit die Endlichkeitskorrektur $1 - \frac{n}{N}$ vernachlässigen. Wir rechnen trotzdem mit der vollständigen Formel:

$N\bar{Y}$ nimmt den Wert $500000 \cdot 700 = 350\,000\,000$ an.

$$1.96 \sqrt{\frac{500000^2}{500} \left(1 - \frac{500}{500000}\right) 200^2} = 8761000$$

$$350\,000\,000 - 8761000 = 341\,239\,000$$

$$350\,000\,000 + 8761000 = 358\,761\,000$$

Das Vertrauensintervall ist:

$$]341\,239\,000; 358\,761\,000[$$

Mit einer Wahrscheinlichkeit von 95% befindet sich die tatsächliche Gesamtsumme in diesem Intervall, wenn der geschätzte Wert $3.5 \cdot 10^8$ ist. $\diamond$

Wie im Falle des Mittelwertes, kann auch hier bei bekannter Varianz eine Formel für die nötige Stichprobengröße für ein spezifisches Vertrauensintervall angegeben werden (das Vorgehen ist dabei mit dem obigen identisch):

31.1.4 Berechnung der Stichprobengröße beim Schätzen des Totals

Mit $q := \Phi^{-1}(\beta + \frac{1-\beta}{2})$ und der Hälfte d der Intervalllänge gilt

$$\begin{aligned} d &= q\sqrt{V(T)} \\ &= q\sqrt{\frac{N^2}{n} \left(1 - \frac{n}{N}\right) \sigma^2} \\ \implies d &= q\sqrt{\frac{N^2}{n} \left(1 - \frac{n}{N}\right) \sigma^2} \\ \implies d^2 &= q^2 \frac{N^2}{n} \left(1 - \frac{n}{N}\right) \sigma^2 \\ \implies \frac{d^2}{q^2 \sigma^2 N^2} &= \frac{1}{n} \left(1 - \frac{n}{N}\right) = \frac{1}{n} - \frac{1}{N} \\ \implies \frac{d^2}{q^2 \sigma^2 N^2} + \frac{1}{N} &= \frac{1}{n} \end{aligned}$$

Da

$$\frac{d^2}{q^2 \sigma^2 N^2} + \frac{1}{N} = \frac{d^2 + q^2 \sigma^2 N}{q^2 \sigma^2 N^2}$$

(gleichnamig machen) gilt

$$n = \frac{q^2 \sigma^2 N^2}{d^2 + q^2 \sigma^2 N}$$

$n = \frac{q^2 \sigma^2 N^2}{d^2 + q^2 \sigma^2 N}$ (mit Endlichkeitskorrektur)
σ^2 = Varianz in der Grundgesamtheit, gewöhnlich geschätzt
aus früheren Erhebungen
d = Hälfte der Vertrauensintervalllänge
N = Anzahl Objekte der Gesamtpopulation
n = Anzahl Objekte der Stichprobe
$q = \Phi^{-1}(\beta + \frac{1-\beta}{2})$
β = Wahrscheinlichkeit, dass Grösse im Vertrauensintervall liegt (z.B. 0.95)
Φ^{-1} Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung

Beispiel 31.1.17. Sei $d = 50000$, $\beta = 0.95$, $\sigma^2 = 200$ und $N = 150000$. Damit erhält man für ein 95%-VI:

$$n = \frac{1.96^2 \cdot 200 \cdot 150000^2}{50000^2 + 1.96^2 \cdot 200 \cdot 150000} = 6610.2 \approx 6611.$$

$\diamond$

31.2 Schätzen von Anteilen

Die entsprechenden Sätze für Anteile kann man aus den Aussagen für Mittelwerte herleiten. Setzen wir nämlich

$$\begin{aligned} y_i &= 1, \text{ sofern ein Objekt } x_i \text{ das Merkmal aufweist} \\ y_i &= 0, \text{ sofern ein Objekt } x_i \text{ das Merkmal nicht aufweist} \end{aligned}$$

so entspricht der Anteil dem Mittelwert von Daten. Sei G die Grundgesamtheit und A die Menge der Objekte, welche ein Merkmal aufweisen ($A \subset G$), N die Anzahl der Objekte von G , N_A Anzahl Objekte von A . Der Anteil π_A von A an G ist somit:

$$\pi_A := \frac{N_A}{N}$$

Sei zudem

$$p_A = \frac{n_A}{n}$$

der Anteil in der Stichprobe (n_A = Anzahl der Elemente der Menge $A_{\mathfrak{S}} \subset A$ in der Stichprobe $\mathfrak{S}$; n = Stichprobengröße). Mit diesen Festlegungen gilt:

Satz 31.2.1. *Sei $y_i = 1$, wenn $y_i \in A$ und $y_i = 0$ sonst. Dann ist*

$$\begin{aligned} (i) \quad & \sum_{i=1}^N y_i = N_A; \quad \sum_{i=1}^n y_i = n_A \\ (ii) \quad & \pi_A = \frac{N_A}{N} = \frac{\sum_{i=1}^N y_i}{N} = \frac{1}{N} \sum_{i=1}^N y_i = \mu \quad (= \text{Mittel der Grundgesamtheit}) \\ (iii) \quad & p_A = \frac{n_A}{n} = \frac{\sum_{i=1}^n y_i}{n} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}_i \quad (= \text{Stichprobenmittel}) \end{aligned}$$

Der Anteil π_A wird erwartungstreu geschätzt durch die Schätzstatistik „Anteil“ P_A , wie der folgende Satz festhält - für einen Beweis s. S. 583:

Satz 31.2.2. *Sei P_A die Zufallsvariable, die für eine Stichprobe $\mathfrak{S}_i$ den Stichprobenanteil p_i^A annimmt und sei π_A der Anteil von A an G . Dann gilt*

$$E(P_A) = \pi_A.$$

Für die Varianz der Zufallsvariable P_A lässt sich zeigen (für eine Beweis s. S. 583):

Satz 31.2.3. $V(P_A) = \frac{N-n}{nN} \sigma^2 = \frac{N-n}{N-1} \frac{\pi_A(1-\pi_A)}{n}$ mit $\sigma^2 = \frac{N}{N-1} \pi_A(1-\pi_A)$

σ^2 müssen wir in der Praxis wiederum schätzen. Dabei gilt gemäss Satz 31.1.10, dass S^2 eine erwartungstreue Schätzstatistik für σ^2 ist, wobei S^2 bei spezifischen Stichproben $\mathfrak{S}_i$ den Wert

$$\frac{1}{n-1} \sum_{j=1}^n (y_j - \bar{y}_i)^2 = \frac{1}{n-1} (np_i^A - n(p_i^A)^2) = \frac{n}{n-1} (p_i^A(1-p_i^A))$$

annimmt. Mit

$$\begin{aligned} & \frac{N-n}{nN} \frac{n}{n-1} P_A(1-P_A) \\ &= \frac{N-n}{N} \frac{P_A(1-P_A)}{n-1} \\ &= \left(1 - \frac{n}{N}\right) \frac{P_A(1-P_A)}{n-1} \end{aligned}$$

ist also

$$\left(1 - \frac{n}{N}\right) \frac{P_A(1-P_A)}{n-1}$$

eine erwartungstreue Schätzstatistik für $V(P_A)$, d.h. es gilt:

Satz 31.2.4. $E\left((1 - \frac{n}{N}) \frac{P_A(1-P_A)}{n-1}\right) = V(P_A)$

Dabei ist $(1 - \frac{n}{N})$ auch hier die Endlichkeitskorrektur. Wir können gemäss zentralem Grenzwertsatz davon ausgehen, dass

$$\frac{P_A - \pi_A}{\sqrt{(1 - \frac{n}{N}) \frac{P_A(1-P_A)}{n-1}}} \sim N(0, 1).$$

Wir berechnen die Grenzen des Vertrauensintervalls mittels den Gleichungen

$$\frac{1-\beta}{2} = \Phi\left(\frac{P_A - P_A^{gu}}{\sqrt{(1 - \frac{n}{N}) \frac{P_A(1-P_A)}{n-1}}}\right) \text{ und } \Phi\left(\frac{P_A - P_A^{go}}{\sqrt{(1 - \frac{n}{N}) \frac{P_A(1-P_A)}{n-1}}}\right) = \left(\beta + \frac{1-\beta}{2}\right).$$

Analog zur Herleitung im Kapitel „Vertrauensintervalle“ erhalten wir (P_A wird durch den Wert der Schätzstatistik P_A geschätzt):

$P_A^{gi} = P_A \pm \Phi^{-1}(\beta + \frac{1-\beta}{2}) \sqrt{(1 - \frac{n}{N}) \frac{P_A(1-P_A)}{n-1}}$
mit Endlichkeitskorrektur
β = Wahrscheinlichkeit, dass Grösse im Vertrauensintervall liegt (z.B. 0.95)
P_A^{gi} = Zufallsvariablen „Grenze des Vertrauensintervalls“ mit konkreten
Grenzen als Wert.
P_A = Zufallsvariable mit den Anteilen p_i^A an der Stichprobe als Werten
N = Grösse der Gesamtpopulation
n = Stichprobengrösse
Φ^{-1} Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung

Beispiel 31.2.5. Es soll der Anteil an Jugendlichen bei den Konsumenten von Mobiltelefonie in einer Stadt erhoben werden ($N = 200'000$). Dazu wird eine Zufallsstichprobe von 400 Personen unter den Mobiltelefonbenutzern gezogen. Es ergebe sich ein Anteil von 100 Jugendlichen. Wir schätzen den Anteil der Jugendlichen bei den Konsumenten durch den Stichprobenanteil $\frac{100}{400} = 0.25$ und berechnen das 95%- (das 98%-) Vertrauensintervall:

$$\frac{100}{400} \pm 1.96 \cdot \sqrt{(1 - \frac{400}{200000}) \frac{\frac{100}{400}(1 - \frac{100}{400})}{400 - 1}} = 0.29245$$

für „+“. Da $\frac{100}{400} - 0.29245 = -0.04245$, erhalten wir für die linke Intervallgrenze: $\frac{100}{400} - 0.04245 = 0.20755$. Damit erhalten wir das Intervall: $]0.20755, 0.29245[$. Mit einer Wahrscheinlichkeit von 95% befindet sich der tatsächliche Anteil von Jugendlichen in diesem Intervall, wenn der Stichprobenanteil $\frac{100}{400} = 0.25$ beträgt. $\diamond$

31.2.1 Stichprobengrösse bei Anteilschätzung

Es gilt mit $q := \Phi^{-1}(\beta + \frac{1-\beta}{2})$ und der Hälfte d der Vertrauensintervalllänge:

$$\begin{aligned} d &= q \sqrt{\frac{N-n}{N-1} \frac{P_A(1-P_A)}{n}} \\ \implies d^2 &= q^2 \frac{N-n}{N-1} \frac{P_A(1-P_A)}{n} \\ \implies \frac{d^2}{q^2} \frac{N-1}{P_A(1-P_A)} &= \frac{N-n}{n} = \frac{N}{n} - 1 \\ \implies \frac{N}{n} &= \frac{(N-1)d^2}{q^2 P_A(1-P_A)} + 1 \\ \implies n &= \frac{N}{\frac{(N-1)d^2}{q^2 P_A(1-P_A)} + 1} \end{aligned}$$

Für das unbekannte $P_A(1-P_A)$ können wir alte Stichprobenergebnisse oder das Maximum einsetzen. Ein Maximum von 0.25 wird in $P_A = 0.5$ erreicht (Beweis: Ableiten von $f(P_A) = P_A(1-P_A) = P_A - (P_A)^2$ ergibt mit Nullsetzen: $f'(P_A) = 1 - 2P_A = 0 \implies P_A = 0.5$. Mit der Ableitung $f''(P_A) = -2$ kontrollieren wir, ob Maximum vorliegt. Damit liegt in $P_A = 0.5$ ein Maximum vor. Diese ist: $f(0.5) = 0.5(1-0.5) = 0.25$). n ist nun nur mehr durch die Breite des β -Vertrauensintervalls bezüglich der Anteile bestimmt.

$n = \frac{N}{\frac{(N-1)d^2}{q^2 P_A(1-P_A)} + 1}$ (mit Endlichkeitskorrektur)
d = Hälfte der Vertrauensintervalllänge
N = Grösse der Gesamtpopulation
n = Stichprobengrösse
$q = \Phi^{-1}(\beta + \frac{1-\beta}{2})$
β = Wahrscheinlichkeit, dass Grösse im Vertrauensintervall liegt (z.B. 0.95)
Φ^{-1} Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung

Beispiel 31.2.6. Annahmen: $N = 2000$; $d = 0.025$; $\beta = 0.95$. Dann muss:

◇

$$n = \frac{2000}{\frac{(2000-1)0.025^2}{1.96^2 0.5^2} + 1} = 869.23$$

(d.h. $n = 870$)

31.3 Schätzen des Medians und von Quartilen

Der Median der Populationswerte kann durch den Median der Stichprobe geschätzt werden. Die Quartile der Populationswerte werden durch die Quartile der Stichprobe geschätzt. Der Median hat gegenüber dem arithmetischen Mittel den Vorteil, durch Ausreisser nicht beeinflusst zu werden. Entsprechend ist die Verwendung des Medians angemessen, wenn die Stichprobe Ausreisser enthalten (z.B. beim Schätzen des Einkommens einer Kleinstadt, wenn ein Milliardär in der Stichprobe vorkommt).

Eine grobe Abschätzung der Standardabweichung des Stichproben-Medians unter Annahme einer Normalverteilung liefert

$$\frac{0.93(Q_3 - Q_1)}{\sqrt{n}}$$

(Q_i = i-tes Quartil). Deshalb versagt diese Angabe genau dann, wenn die Verwendung des Medians vorteilhaft ist. Damit wird die Angabe von Vertrauensintervallen bei nicht-normalverteilten

Zufallsvariablen schwierig. Man kann allerdings die Streuung des Medians simulieren, indem man aus den Stichprobendaten Stichproben zieht und dann das Vertrauensintervall aus der erhaltenen empirischen Verteilung des Medians bestimmt (Monte-Carlo-Verfahren, s. S. 20.9, Unterkapitel „Tests auf Kennzahlen bei bekannter, nicht-normaler Verteilung oder ohne bekannte Verteilung“). Es gibt dabei verschiedene Verfahren (Bootstrap; Jackknife).

31.4 Geschichtete Zufallsstichproben

Weiss man, dass bestimmte Merkmale der Population korrelieren (z.B. Frauen kaufen mehr Äpfel als Männer) und möchte man eine Erhebung über den Apfelkonsum der Bevölkerung machen, so kann man die Vertrauensintervalle bei gegebener Stichprobengrösse verkleinern (oder die Stichprobengrösse bei gegebenem Vertrauensintervall verkleinern), wenn man die verschiedenen Gruppen gesondert betrachtet. Die Untergruppen werden „Schichten“ genannt. Die Schichten können wiederum geschichtet werden. Wenige Schichtungsebenen können die Vertrauensintervalle stark reduzieren, wenn die geeigneten Schichtungen vorgenommen werden, d.h. wenn die Varianz innerhalb der Schichten klein und die zwischen den Schichten gross ist. Wächst die Anzahl der Schichtungsebenen aber allzu sehr, so wird der Nutzen schnell kleiner. Die Anzahl Objekte pro Schicht muss grösser als 2 sein, da sonst keine Varianz berechnet werden kann. Oft muss man bei Erhebungen damit rechnen, dass nur 50% der Personen erfasst werden können. Dieser Umstand muss bei der Berechnung der Stichprobengrösse berücksichtigt werden - um so mehr bei Schichtungen.

Innerhalb der Schichten können die bisher betrachteten Techniken verwendet werden. So ist z.B. die Zufallsvariable Schichtmittel $\bar{Y}_h$ der Schicht h eine erwartungstreue Schätzstatistik des Mittelwertes der entsprechenden Untergruppe der Gesamtpopulation. Ebenso können die schichtspezifischen Anteile, Varianzen, etc. berechnet werden. Für die Berechnung der entsprechenden Gesamtwerte sind die Werte der Schichten mit den Anteilen der Schichten an der Grundgesamtheit zu gewichten und dann zusammenzuzählen. Als Gewicht wählen wir beim Mittelwert $\frac{N_h}{N}$ (N_h = Anzahl Objekte der Schicht h , N = Anzahl Objekte der jeweiligen Obermenge (Bei nur eine Schichtebene ist N = Anzahl der Objekte der Grundgesamtheit)).

Wir erhalten also folgende Formeln:

Mittelwert:

$\bar{Y} = \frac{1}{N} \sum_{h=1}^L N_h \bar{Y}_h = \sum_{h=1}^L \frac{N_h}{N} \bar{Y}_h$
N = Anzahl Objekte der Grundgesamtheit
L = Anzahl Schichten pro Schichtebene
N_h = Anzahl Objekte der Schicht h (Grundgesamtheit)
$\bar{Y}_h$ = Zufallsvariable mit Stichprobenmittelwert der Schicht h als Wert

Für die Varianz des Stichprobenmittels kann folgender Term hergeleitet werden:

$V(\bar{Y}) = \sum_{h=1}^L \left(W_h^2 \cdot \frac{1}{n_h} \cdot \left(1 - \frac{n_h}{N_h} \right) \sigma_h^2 \right)$
L = die Anzahl der Schichten einer Schichtebene
$W_h = \frac{N_h}{N}$ = Anteil der Schicht h in der Gesamtpopulation
n_h = Anzahl der Objekte der Schicht h in der Stichprobe
N_h = Anzahl der Objekte der Schicht h in der Gesamtpopulation
N = Anzahl Objekte der Gesamtpopulation
$\sigma_h^2 = \frac{1}{N_h - 1} \sum_{i \in U_h} (y_i - \mu_h)^2$ die Varianz der Objekte der Schicht h in der Gesamtpopulation
U_h = Menge der Indizes der Objekte der Schicht h in der Gesamtpopulation
μ_h = Mittelwert der Ausprägung y_i der Schicht h in der Gesamtpopulation

Konkret verfahren wir bei der Berechnung wie folgt:

σ_h^2 wird geschätzt durch $S_h^2 = \frac{1}{n_h-1} \sum_{i \in I_h} (y_i - \bar{Y}_h)^2$
I_h = Menge der Indizes der Objekte der Schicht h in der Stichprobe
S_h^2 = Zufallsvariable mit der Stichprobenvarianz der Schicht h als Wert.
$\bar{Y}_h$ = Zufallsvariable mit dem Stichprobenmittel der Schicht h als Wert.
y_i = Stichprobenwerte

Wir erhalten das Vertrauensintervall:

$$\bar{Y}_{g_i} = \bar{Y} \pm \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right) \sqrt{V(\bar{Y})}$$

Mit Hilfe eines Beispiels können wir zeigen, dass die Varianz der Mittelwerte bei Schichtung (oft) kleiner ist als ohne Schichtung (dies liegt daran, dass bei Schichtbildung die Varianz *zwischen* den Schichten wegfällt!).

Beispiel 31.4.1. In einer Studie werden die Ausgaben für Musikprodukte in einem Kanton erhoben. Es wird vermutet, dass junge Leute mehr für Musikprodukte ausgeben als Ältere. Der Anteil der Jungen an der Population betrage 20%. Die Stichprobengrösse betrage 500, $N = 250'000$ ($\Rightarrow N_h$ der Jungen = $0.2 \cdot 250000 = 50000$, N_h der Älteren = 200000). Wir ziehen eine Zufallsstichprobe unter den Jungen und eine Zufallsstichprobe unter den Älteren (Resultate siehe math.logik.ch; Stichproben, Beispiel 8).

100.874 ist der Mittelwert der Stichprobe 1

49.486 ist der Mittelwert der Stichprobe 2

Das geschätzte Mittel ist : $\frac{100.874 \cdot 50000 + 49.486 \cdot 200000}{250000} = 59.7636$

Die Varianz in der Gesamtstichprobe beträgt 634.0460471.

Die Varianz in der Stichprobe bei den Jungen beträgt: 237.3241657

Die Varianz in der Stichprobe bei den Älteren beträgt: 204.6009063

Wir erhalten als geschätzte Gesamtvarianz von $\bar{Y}$ (bei Schichtung):

$$0.2^2 \cdot \frac{1}{100} \cdot \left(1 - \frac{100}{50000}\right) \cdot 237.3241657 + 0.8^2 \cdot \frac{1}{400} \cdot \left(1 - \frac{400}{200000}\right) \cdot 204.6009063 = 0.42145$$

Wir erhalten als geschätzte Gesamtvarianz von $\bar{Y}$ (ohne Schichtung): $\frac{1}{500} \cdot \left(1 - \frac{500}{250000}\right) \cdot 634.0460471 = 1.2656$.

Die Schichtbildung wirkt sich auf die 95%-Vertrauensintervalle am Beispiel wie folgt aus:

Ohne Schichtung:

$$\bar{Y}_{g_i} = \bar{Y} \pm 1.96 \cdot \sqrt{1.2656} \Rightarrow 59.7636 \pm 2.2050$$

Das VI ist also]59.7636 - 2.2050, 59.7636 + 2.2050[=]57.559, 61.969[

Mit Schichtung erhalten wir:

$$\bar{Y}_{g_i} = \bar{Y} \pm 1.96 \cdot \sqrt{0.42145} \Rightarrow \bar{y}_i \pm 1.2724$$

Das VI ist also]59.7636 - 1.2724, 59.7636 + 1.2724[=]58.491, 61.036[

Dies ist eine Verkleinerung des Vertrauensintervalls um einen Faktor von $\frac{1.2724}{2.2050} = 0.57705$

Um diese Verkleinerung des Vertrauensintervalls ohne Schichtung durch eine Erhöhung der Stichprobe zu erhalten, müssten wir folgendes n erreichen:

$$\begin{aligned}
n &= s_S^2 N \frac{\left(\Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right)\right)^2}{s_S^2 \left(\Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right)\right)^2 + (\bar{Y}_{gi} - \bar{Y})^2 N} \\
&= 634.0460471 \cdot 250000 \cdot \frac{1.96^2}{634.0460471 \cdot 1.96^2 + 1.272 \cdot 4^2 \cdot 250000} \\
&= 1495.5
\end{aligned}$$

Für Totale erhalten wir die Schätzstatistik:

$T = \sum_{h=1}^L N_h \bar{Y}_h$
$L = \text{Anzahl Schichten}$
$N_h = \text{Anzahl Objekte der Schicht } i \text{ in der Grundgesamtheit.}$
$\bar{Y}_h = \text{Zufallsvariable mit Stichprobenmittel der Schicht } i \text{ als Wert}$

Es handelt sich um die Summe der geschätzten Totale pro Schicht. ◇

Die Varianz ist:

$$V(T) = \sum_{h=1}^L \left(N_h^2 \cdot \frac{1}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \sigma_h^2 \right),$$

also wiederum einfach die Summe der Varianzen pro Schicht, denn

$$\begin{aligned}
V(T) &= N^2 V(\bar{Y}) = N^2 \sum_{h=1}^L \left(W_h^2 \cdot \frac{1}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \sigma_h^2 \right) \\
&= N^2 \sum_{h=1}^L \left(\left(\frac{N_h}{N} \right)^2 \cdot \frac{1}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \sigma_h^2 \right) = \sum_{h=1}^L \left(N_h^2 \cdot \frac{1}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \sigma_h^2 \right)
\end{aligned}$$

und das Vertrauensintervall

$$T_g = T \pm \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right) \sqrt{V(T)}$$

Dabei schätzen wir σ_h^2 wiederum durch S_h^2 (Zufallsvariable mit den Varianz der Schicht h als Wert).

Beispiel 31.4.2. Für die Angaben des Beispiels 31.4.1 erhält man:

Geschätztes Total: $50000 \cdot 100.874 + 200000 \cdot 49.486 = 14940900$

Geschätzte Varianz der Statistik Total

$$50000^2 \frac{1}{100} \left(1 - \frac{100}{50000}\right) 237.3241657 + 200000^2 \frac{1}{400} \left(1 - \frac{400}{200000}\right) = 26340408378$$

$$0.95\text{-VI: } 14940900 \pm 1.96 \cdot \sqrt{26340408378} :]14622803.17, 15258996.83[$$

◇

Für die Anteilsschätzung erhalten wir für geschichtete Stichproben:

$P_A = \sum_{h=1}^L \frac{N_h P_h}{N}$ ist eine erwartungstreue Schätzstatistik eines Anteils der Grundgesamtheit.
$P_A = \text{Zufallsvariable, die einen konkreten Stichprobenanteil als Wert annimmt}$
$N_h = \text{Grösse der Schicht in der Grundgesamtheit.}$
$P_h = \text{Zufallsvariable, die einen konkreten Schicht-Stichprobenanteil annimmt.}$
$N = \text{Grösse der Grundgesamtheit.}$

und

$V(P_A) = \frac{1}{N^2} \sum_{h=1}^L \frac{N_h^2(N_h - n_h)}{N_h - 1} \frac{P_h(1 - P_h)}{n_h}.$
P_A = Zufallsvariable, die einen konkreten Stichprobenanteil als Wert annimmt
N_h = Grösse der Schicht h in der Grundgesamtheit.
P_h = Geschätzt durch p_h - Wert der Statistik P_h .
N = Grösse der Grundgesamtheit.
n_h = Grösse der Schicht h in der Stichprobe.

Dies ergibt als Vertrauensintervall:

$$P_g = P_A \pm \Phi^{-1}(\beta + \frac{1-\beta}{2}) \sqrt{V(P_A)}.$$

Beispiel 31.4.3. Einstellung zu einer politischen Massnahme: *math.logik.ch*; Stichproben, Beispiel 8, erste und dritte Spalte).

Anteil der Befürworter an Stichprobe 1: $\frac{15}{100} = 0.15$ (geschätzter Anteil bei Schichtung)

Anteil der Befürworter an Stichprobe 2: $\frac{99}{400} = 0.2475$

Geschätzter Anteil: $\frac{1}{5}0.15 + \frac{4}{5} * \frac{99}{400} = 0.228$

Geschätzte Varianz: $\frac{1}{250000} \left(\frac{50000^2}{50000-1} (50000 - 100) 0.15 \frac{(1-0.15)}{100} + \frac{200000^2}{200000-1} (200000 - 400) \frac{99}{400} \frac{1-\frac{99}{400}}{400} \right) =$

0.000348295

0.95-VI: $0.228 \pm 1.96\sqrt{0.000348295}$:]0.191421877, 0.264578123[

◇

Bezüglich der nötigen Stichprobengrössen lässt sich bemerken: die Berechnung der Stichprobengrösse erfolgt nicht mehr so direkt, wie bei den ungeschichteten Stichproben. Statt uns in dieses - etliche Zusatzüberlegungen benötigende - Unternehmen zu verstricken, ein praktischer Hinweis: Wir können z.B. die Stichprobengrösse pro Schicht anteilmässig festlegen und auf dem Hintergrund einer bekannten Varianz pro Schicht mit den obigen Formeln das Vertrauensintervall berechnen. Dies ziehen wir für unterschiedliche Stichprobengrössen durch (z.B. mit Excel) und können so die bezüglich der gewünschte Genauigkeit nötige Stichprobengrösse berechnen. Oft ist es nützlich, in kleinen Schichten (Grundgesamtheit) mit grosser Varianz eine Vollerhebung vorzunehmen, während dann in grossen Schichten mit kleinerer Varianz eine Zufallsstichprobe gezogen wird.

31.4.1 Nachschichtung

Bei der Schichtung werden für die Zielgruppen gesondert Zufallsstichproben erhoben. Möglich ist ein weiteres Verfahren, das man „Nachschichtung“ nennt. Dabei wird eine Stichprobe in der Gesamtpopulation erhoben und die Schichten erst im Nachhinein unterschieden. Dabei müssen die Anteile der Schichten an der Gesamtpopulation bekannt sein. Bei der Nachschichtung sind die Stichprobengrössen der Schichten Werte einer Zufallsvariable. Trotzdem kann man bei genügend grossen Teilstichproben die oben für Schichtung eingeführten Formeln verwenden.

Beispiel 31.4.4. Wir wollen den durchschnittlichen Konsum elektronischer Unterhaltungsgüter einer Population schätzen (in Franken). Dazu soll ein 95%-Vertrauensintervall angegeben werden. Die Population weise 500'000 Personen auf. Wir vermuten, dass ältere Leute (≥ 60) weniger Unterhaltungselektronik konsumieren als andere Alterssegmente. Es wird eine Zufallsstichprobe von 500 Personen gezogen. Wir untersuchen die beiden Altersgruppen im Nachhinein getrennt. Wir unterteilen entsprechend die Stichprobe von 500 Personen in zwei Schichten A(ter) und B. Auf Grund der Daten (siehe *Stichproben.xls*) erhalten wir folgende Resultate: Zur Schicht der Älteren gehören 93, zur Schicht der Jüngeren gehören $500 - 93 = 407$. Der Anteil der Älteren an der Gesamtpopulation beträgt 22%.

Mittel der Stichprobe: 800.7386095

Wir berechnen die Standardabweichung für die Gesamtstichprobe, um diese mit der Summe der Varianzen der Schichten zu vergleichen.

Varianz der Gesamtstichprobe: 3210.01891

Varianz in der Altersgruppen der Älteren: 141.8137735

Varianz in der Altersgruppe der Anderen: 2094.9

Geschätzte Gesamtvarianz des Mittelwertes bei Schichtung:

$$s_{\bar{y}} = \left(\frac{22}{100}\right)^2 \cdot \frac{1}{93} \cdot \left(1 - \frac{93}{0.22 \cdot 500000}\right) \cdot 141.8137735 + \\ \left(\frac{78}{100}\right)^2 \cdot \frac{1}{407} \cdot \left(1 - \frac{407}{0.78 \cdot 500000}\right) \cdot 2094.9 = 3.202$$

Geschätzte Gesamtvarianz des Mittelwertes ohne Schichtung:

$$s_{\bar{y}} = \frac{1}{500} \left(1 - \frac{500}{500000}\right) 3210.01891 = 6.4136$$

Wir erhalten entsprechend folgende 95%-Vertrauensintervalle mit und ohne Nachschichtung für dieselbe Stichprobe:

Mit Nachschichtung: $\bar{Y}_{g_i} = \bar{Y} \pm 1.96\sqrt{3.202} \Rightarrow \text{Grenzen} = 800.7386095 \pm 3.5073$

Ohne Nachschichtung: $\bar{Y}_{g_i} = \bar{Y} \pm 1.96\sqrt{6.4136} \Rightarrow \text{Grenzen} = 800.7386095 \pm 4.9637$

Das Vertrauensintervall kann somit um einen Faktor von $\frac{3.5073}{4.9637} = 0.70659$ verkleinert werden.

Wollten wir eine solche Verkleinerung ohne Schichtung durch die Erhöhung der Stichprobe erreichen, müssten wir (ohne Endlichkeitskorrektur) eine Stichprobe von $n = \frac{(\Phi^{-1}(\beta + \frac{1-\beta}{2}))^2 \sigma^2}{\beta^2} = \frac{1.96^2 \cdot 3210.01891}{3.5073^2} = 1002.5$ erheben. $\diamond$

31.5 Schlussbemerkungen

Hiermit wurden nur einige kurze erste Schritte in die Stichprobentheorie getan, ein Gebiet, das bei professioneller Anwendung durchaus technisch komplex wird. So wurde die Behandlung von Klumpenstichproben (engl. „cluster samples“, zu unterscheiden von der Clusteranalyse, einem multivariaten Verfahren, welches versucht, in einer Menge von Objekten, die als Werte von Zufallsvariablen betrachtet werden, inhaltlich sinnvolle Teilmengen zu bilden), und die von gemischten oder mehrstufigen Stichprobenplänen (z.B. Klumpen (aus denen Zufallsstichproben gezogen werden) unterlassen. Klumpen sind Mengen von Objekten, die alle in die Untersuchung gelangen (z.B. bei einer Untersuchung werden Schüler ausgewählt, indem auf der Basis eine Liste aller Klassen eine Zufallsstichprobe von Klassen gezogen wird. In die Untersuchung kommen dann alle Schüler der gezogenen Klassen. Es handelt sich hier nicht mehr um eine Zufallsstichprobe). Bei solchen Stichprobenplänen werden oft Monte-Carlo-Verfahren verwendet (s. entsprechende Ausführungen auf Seite 572, Schätzen von Vertrauensintervallen für den Median). Es gibt verschiedene Verfahren wie Bootstrap, Jackknife, Balanced Repeated Replicates (BRR) oder Fay's Methode (s. (Lumley, 2010)).

Heikel ist auch die Behandlung von Ausreißern. Datenmaterial ist immer auf Plausibilität hin zu untersuchen. Es ist zu überprüfen, ob ein Ausreißer durch falsche Datenerhebung oder -eingeabe entstand. Ist der Ausreißer echt, gibt es verschiedene Verfahren, um damit umzugehen. So können Ausreißer bei der Auswertung weggelassen werden, um z.B. den Mittelwert nicht allzu sehr zu verfälschen (gestutztes Mittel). Manchmal werden Ausreißer auch auf eine äussere Grenze zurückgesetzt, oder es werden robuste Schätzstatistiken wie der Median verwendet.

Ein besonderes Problem stellen Antwortausfälle dar. Bei der Befragung von Haushalten rechnet das Bundesamt für Statistik mit Antwortausfällen von 20 bis 50%. Bei Umfragen bei Unternehmen rechnet man sogar mit 40 bis 80% Ausfällen. Das besondere Problem von Antwortausfällen besteht darin, dass Personen antworten oder nicht antworten, die vermutlich spezifische Charakteristiken aufweisen. Dadurch werden die Ergebnisse systematisch verfälscht. Macht man z.B. eine Umfrage über durchschnittliche Kosmetikausgaben, und antworten mehrheitlich nur Mittelschicht- und Oberschichtfrauen, so werden die Resultate eine systematische Verzerrung (= engl. Bias) aufweisen. Im Beispiel könnte man z.B. die Stichprobe nachschichten und die Schichtresultate dann entsprechend den Anteilen an der Gesamtpopulation hochrechnen. Allerdings werden durch Hoch-

rechnen die Varianz und damit die Vertrauensintervalle vergrößert.

Ein spezielles Problem stellt sich zudem mit partiellen Antwortausfällen (Personen beantworten nicht alle Fragen). Manchmal werden solche Antwortausfälle mit Hilfe von Schätzungen kompensiert (Imputation). Angesichts dieser Probleme zu jeweils zufriedenstellenden Resultaten zu gelangen verlangt einiges an Erfahrung und an statistischem Detailwissen, das über den hier gelieferten Apparat hinaus geht. Deshalb ein paar Literaturhinweise: Ein Klassiker ist [Cochran \(1977\)](#). Neuere empfehlenswerte Werke sind [Thompson \(2012\)](#) sowie [Särndal, Swensson und Wretman \(1992\)](#). Eine Einführung in die Verwendung von **R** liefert [Lumley \(2010\)](#).

31.6 Beweise

Beweis von Satz 31.1.1

Mit der Definition des Erwartungswertes und $P(\mathfrak{S}_i) = \frac{1}{\binom{N}{n}}$ erhält man:

$$E(T_s) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \cdot \sum_{y_j \in \mathfrak{S}_i} y_j = \frac{1}{\binom{N}{n}} \cdot \sum_{i=1}^{\binom{N}{n}} \sum_{y_j \in \mathfrak{S}_i} y_j$$

$\sum_{i=1}^{\binom{N}{n}} \sum_{y_j \in \mathfrak{S}_i} y_j$ ist die Summe aller Stichprobensummen. Jedes y_j kommt in $\binom{N-1}{n-1}$ Stichproben vor, da es $\binom{N-1}{n-1}$ Stichproben aus der Grundgesamtheit ohne y_j gibt und jede dieser Stichproben durch y_j zu einer Stichprobe der Grösse n aufgefüllt werden kann. Entsprechend gilt, dass

$$\sum_{i=1}^{\binom{N}{n}} \sum_{y_j \in \mathfrak{S}_i} y_j = \binom{N-1}{n-1} \sum_{j=1}^N y_j.$$

Also ist

$$\begin{aligned} E(T_s) &= \frac{1}{\binom{N}{n}} \binom{N-1}{n-1} \sum_{j=1}^N y_j \\ &= \frac{1}{\frac{N!}{(N-n)!n!}} \frac{(N-1)!}{(N-1-n+1)!(n-1)!} \sum_{j=1}^N y_j \\ &= \frac{(N-n)!n!}{N!} \frac{(N-1)!}{(N-n)!(n-1)!} \sum_{j=1}^N y_j \\ &= \frac{n}{N} \sum_{j=1}^N y_j \end{aligned}$$

Beweis von Satz 31.1.2

Es gilt $\bar{Y} = \frac{1}{n} T_s$ (T_s nimmt die Summe der Werte einer Stichprobe an) und damit $E(\bar{Y}) =$

$$E\left(\frac{1}{n} T_s\right) = \frac{1}{n} E(T_s) \stackrel{\text{Satz 31.1.1}}{=} \frac{1}{n} \frac{n}{N} \sum_{j=1}^N y_j = \frac{1}{N} \sum_{j=1}^N y_j = \mu$$

Beweis von Satz 31.1.4

$$E(T) = E(N\bar{Y}) = NE(\bar{Y}) = N\mu = \tau$$

Beweis von Satz 31.1.6

Mit der Definition des Erwartungswertes und $P(\mathfrak{S}_i) = \frac{1}{\binom{N}{n}}$ gilt:

$$E(Q) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \cdot \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 = \frac{1}{\binom{N}{n}} \cdot \sum_{i=1}^{\binom{N}{n}} \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2$$

$\sum_{i=1}^{\binom{N}{n}} \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2$ ist die Summe aller Stichprobenquadratsummen. Jedes $(y_j - \mu)^2$ kommt in $\binom{N-1}{n-1}$ Stichproben vor. Entsprechend gilt, dass

$$\sum_{i=1}^{\binom{N}{n}} \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 = \binom{N-1}{n-1} \sum_{j=1}^N (y_j - \mu)^2.$$

Also gilt mit den zum Beweis von Satz 31.1.1 analogen Umformungen der Binomialkoeffizienten

$$E(Q) = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} \sum_{j=1}^N (y_j - \mu)^2 = \frac{n}{N} \sum_{j=1}^N (y_j - \mu)^2$$

Beweis von Satz 31.1.7

Mit der Definition des Erwartungswertes und $P(\mathfrak{S}_i) = \frac{1}{\binom{N}{n}}$ gilt:

$$E(K) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \cdot \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu) = \frac{1}{\binom{N}{n}} \sum_{i=1}^{\binom{N}{n}} \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu)$$

Im rechten Ausdruck ist der Term

$$\sum_{i=1}^{\binom{N}{n}} \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu)$$

die Summe aller k_i . In dieser Summe kommt jeder Term $(y_j - \mu)(y_m - \mu)$ genau $\binom{N-2}{n-2}$ mal vor, da aus $G \setminus \{y_j, y_m\}$ genau $\binom{N-2}{n-2}$ Stichproben der Grösse $n-2$ bildbar sind. Diese können mit y_j und y_m zu Stichproben der Grösse n ergänzt werden. Entsprechend gilt

$$\sum_{i=1}^{\binom{N}{n}} \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu) = \binom{N-2}{n-2} \sum_{j=1}^N \sum_{\substack{m=1 \\ m \neq j}}^N (y_j - \mu)(y_m - \mu)$$

und damit

$$E(K) = \frac{1}{\binom{N}{n}} \binom{N-2}{n-2} \sum_{j=1}^N \sum_{\substack{m=1 \\ m \neq j}}^N (y_j - \mu)(y_m - \mu)$$

wobei

$$\frac{1}{\binom{N}{n}} \binom{N-2}{n-2} = \frac{(N-n)!n!}{N!} \frac{(N-2)!}{(N-n)!(n-2)!} = \frac{n(n-1)}{N(N-1)}$$

d.h.

$$E(K) = \frac{n(n-1)}{N(N-1)} \sum_{j=1}^N \sum_{\substack{m=1 \\ m \neq j}}^N (y_j - \mu)(y_m - \mu)$$

Beweis von Satz 31.1.8

Es gilt

$$n(\bar{y}_i - \mu) = n \left(\frac{1}{n} \sum_{y_j \in \mathfrak{S}_i} y_j - \mu \right) = \sum_{y_j \in \mathfrak{S}_i} y_j - n\mu = \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)$$

und damit

$$\begin{aligned} n^2(\bar{y}_i - \mu)^2 &= [n(\bar{y}_i - \mu)]^2 = \left(\sum_{y_j \in \mathfrak{S}_i} (y_j - \mu) \right)^2 \\ &= \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 + \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu) \end{aligned}$$

und

$$(\bar{y}_i - \mu)^2 = \frac{1}{n^2} \left(\sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 + \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu) \right) \quad (31.6.1)$$

Nun gilt laut der Definition der Varianz

$$V(\bar{Y}) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) (\bar{y}_i - \mu)^2$$

und mit dem Zwischenergebnis (31.6.1)

$$V(\bar{Y}) = \frac{1}{n^2} \left(\sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 + \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \sum_{\substack{y_j, y_m \in \mathfrak{S}_i \\ j \neq m}} (y_j - \mu)(y_m - \mu) \right)$$

Daraus erhält man mit den Definitionen von Q und K (s. Sätze 31.1.6 und 31.1.7):

$$V(\bar{Y}) = \frac{1}{n^2} (E(Q) + E(K))$$

und mit den Hilfssätzen 31.1.6 und 31.1.7

$$\begin{aligned} V(\bar{Y}) &= \frac{1}{n^2} \left(\frac{n}{N} \sum_{j=1}^N (y_j - \mu)^2 + \frac{n(n-1)}{N(N-1)} \sum_{j=1}^N \sum_{\substack{m=1 \\ j \neq m}}^N (y_j - \mu)(y_m - \mu) \right) \\ &= \frac{1}{nN} \left(\sum_{j=1}^N (y_j - \mu)^2 + \frac{(n-1)}{(N-1)} \sum_{j=1}^N \sum_{\substack{m=1 \\ j \neq m}}^N (y_j - \mu)(y_m - \mu) \right) \end{aligned}$$

Nun ist

$$\left(\sum_{j=1}^N (y_j - \mu) \right)^2 = \sum_{j=1}^N \sum_{\substack{m=1 \\ j \neq m}}^N (y_j - \mu)(y_m - \mu) + \sum_{j=1}^N (y_j - \mu)^2$$

wobei

$$\sum_{j=1}^N (y_j - \mu) = 0,$$

d.h. der linke Term der Gleichung ist 0. Damit gilt:

$$\sum_{j=1}^N \sum_{\substack{m=1 \\ j \neq m}}^N (y_j - \mu)(y_m - \mu) = - \sum_{j=1}^N (y_j - \mu)^2$$

und man erhält:

$$\begin{aligned} V(\bar{Y}) &= \frac{1}{nN} \left(\sum_{j=1}^N (y_j - \mu)^2 + \frac{(n-1)}{(N-1)} \sum_{j=1}^N \sum_{\substack{m=1 \\ j \neq m}}^N (y_j - \mu)(y_m - \mu) \right) \\ &= \frac{1}{nN} \left(\sum_{j=1}^N (y_j - \mu)^2 - \frac{(n-1)}{(N-1)} \sum_{j=1}^N (y_j - \mu)^2 \right) \\ &= \frac{1}{nN} \left(1 - \frac{(n-1)}{(N-1)} \right) \sum_{j=1}^N (y_j - \mu)^2 \end{aligned}$$

Mit

$$\begin{aligned} \frac{1}{nN} \left(1 - \frac{(n-1)}{(N-1)} \right) &= \frac{N-1}{nN(N-1)} - \frac{n-1}{nN(N-1)} \\ &= \left(\frac{N-1}{nN} - \frac{n-1}{nN} \right) \frac{1}{N-1} \\ &= \frac{N-n}{nN} \frac{1}{N-1} \end{aligned}$$

gilt dann:

$$\begin{aligned} V(\bar{Y}) &= \frac{N-n}{nN} \frac{1}{N-1} \sum_{j=1}^N (y_j - \mu)^2 \\ &= \frac{N-n}{nN} \sigma^2 \end{aligned}$$

Beweis von Satz 31.1.10

Es gilt laut der Definition des Erwartungswertes:

$$E(S^2) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \frac{1}{n-1} \sum_{y_j \in \mathfrak{S}_i} (y_j - \bar{y}_i)^2$$

Zudem gilt:

$$\begin{aligned} \sum_{y_j \in \mathfrak{S}_i} (y_j - \bar{y}_i)^2 &= \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu - (\bar{y}_i - \mu))^2 \\ &= \sum_{y_j \in \mathfrak{S}_i} \left((y_j - \mu)^2 - 2(y_j - \mu)(\bar{y}_i - \mu) + (\bar{y}_i - \mu)^2 \right) \end{aligned} \tag{31.6.2}$$

und durch Verteilung des Summenzeichens:

$$\begin{aligned}
 \sum_{y_j \in \mathfrak{S}_i} (y_j - \bar{y}_i)^2 &= \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 - 2(\bar{y}_i - \mu) \frac{n}{n} \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu) + n(\bar{y}_i - \mu)^2 \\
 &= \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 - 2(\bar{y}_i - \mu)n(\bar{y}_i - \mu) + n(\bar{y}_i - \mu)^2 \\
 &= \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 - 2n(\bar{y}_i - \mu)^2 + n(\bar{y}_i - \mu)^2 \\
 &= \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 - n(\bar{y}_i - \mu)^2
 \end{aligned}$$

Damit sowie mit der Definition von $E(Q)$, $V(\bar{Y})$ und Satz 31.1.8 gilt dann:

$$\begin{aligned}
 E(S^2) &= \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \frac{1}{n-1} \left(\sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 - n(\bar{y}_i - \mu)^2 \right) \\
 &= \frac{1}{n-1} \left(\sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \sum_{y_j \in \mathfrak{S}_i} (y_j - \mu)^2 - n \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) (\bar{y}_i - \mu)^2 \right) \\
 &= \frac{1}{n-1} (E(Q) - nV(\bar{Y})) \\
 &= \frac{1}{n-1} \left(E(Q) - n \frac{N-n}{nN} \sigma^2 \right) \\
 &= \frac{1}{n-1} \left(E(Q) - \frac{N-n}{N} \sigma^2 \right)
 \end{aligned}$$

Da nun

$$E(Q) = \frac{N-1}{N-1} E(Q) = \frac{N-1}{N-1} \frac{n}{N} \sum_{j=1}^N (y_j - \mu)^2 = \frac{n(N-1)}{N} \sigma^2$$

gilt

$$E(S^2) = \frac{1}{n-1} \left(\frac{n(N-1)}{N} \sigma^2 - \frac{N-n}{N} \sigma^2 \right)$$

und da

$$\begin{aligned}
 \frac{n(N-1)}{N} \sigma^2 - \frac{N-n}{N} \sigma^2 &= \sigma^2 \left(\frac{n(N-1)}{N} - \frac{N-n}{N} \right) \\
 &= \sigma^2 \frac{n(N-1) - N + n}{N} \\
 &= \sigma^2 \frac{nN - n - N + n}{N} \\
 &= \sigma^2 \frac{nN - N}{N} = \sigma^2 (n-1)
 \end{aligned}$$

gilt

$$E(S^2) = \frac{1}{n-1} \sigma^2 (n-1) = \sigma^2$$

Beweis von Satz 31.1.11

$$E\left(\frac{1}{n}\left(1 - \frac{n}{N}\right) S^2\right) = \frac{1}{n}\left(1 - \frac{n}{N}\right) E(S^2) \stackrel{E(S^2)=\sigma^2}{=} \frac{1}{n}\left(1 - \frac{n}{N}\right) \sigma^2$$

Beweis von Satz 31.1.15

$$V(N\bar{Y}) = N^2 V(\bar{Y}) = \frac{N^2}{n}\left(1 - \frac{n}{N}\right) \sigma^2$$

Beweis von Satz 31.2.2

Sei $y_j = 1$, wenn $y_j \in A_{\mathfrak{S}_i}$ und $y_j = 0$ sonst. Dann gilt:

$$E(P_A) = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) p_i^A = \sum_{i=1}^{\binom{N}{n}} P(\mathfrak{S}_i) \bar{y}_i = E(\bar{Y}) = \mu = P_A$$

Beweis von Satz 31.2.3

Mit $y_j = 1$, wenn $y_j \in A_{\mathfrak{S}_i}$ und $y_j = 0$ sonst, gilt

(i)

$$\sum_{i=1}^N y_i^2 = \sum_{i=1}^N y_i = N_A = N\pi_A$$

(ii)

$$\begin{aligned} \sigma^2 &= \frac{1}{N-1} \sum_{i=1}^N (y_i - \pi_A)^2 = \frac{1}{N-1} \left(\sum_{i=1}^N y_i^2 - 2\pi_A \sum_{i=1}^N y_i + N\pi_A^2 \right) \\ &= \frac{1}{N-1} \left(\sum_{i=1}^N y_i^2 - 2\pi_A N \frac{1}{N} \sum_{i=1}^N y_i + N\pi_A^2 \right) \\ &= \frac{1}{N-1} \left(\sum_{i=1}^N y_i^2 - 2N\pi_A^2 + N\pi_A^2 \right) \\ &= \frac{1}{N-1} \left(\sum_{i=1}^N y_i^2 - N\pi_A^2 \right) \end{aligned}$$

daraus mit (i)

$$(iii) \quad \sigma^2 = \frac{1}{N-1} \left(\sum_{i=1}^N y_i^2 - N(P_A)^2 \right) = \frac{1}{N-1} (N\pi_A - N\pi_A^2) = \frac{N}{N-1} \pi_A(1 - \pi_A)$$

$$(iv) \quad \text{Da mit Satz 31.1.8 } V(P_A) = V(\bar{Y}) = \frac{N-n}{nN} \sigma^2, \text{ gilt mit (iii): } V(P_A) = \frac{N-n}{nN} \frac{N}{N-1} \pi_A(1 - \pi_A) = \frac{N-n}{N-1} \frac{\pi_A(1-\pi_A)}{n}$$

31.6.1 Übungen

1. Im Wallis versucht ein Marktforschungsunternehmen die durchschnittlichen Ausgaben der Touristen für typische Walliser Lebensmittel (Bindenfleisch, Roggenbrot, Käse, etc.) zu ermitteln. Dabei wird unter den Touristen (1 Million) eine Zufallsstichprobe von 500 Personen gezogen. Es ergeben sich folgende Ausgaben für solche Produkte pro Person (siehe Excel-Datei Stichproben.xls). Berechnen Sie die durchschnittlichen Ausgaben und ein 95%-Vertrauensintervall für den erhaltenen Wert. (Berechnen Sie auch ein 98%- Vertrauensintervall)
2. Im Auftrag der Tourismusbranche im Oberwallis soll die „Studien- und Marktforschungs GmbH Oberwallis“ das Total der Ausgaben für Ferien pro erwachsene Person im Oberwallis erheben (in Franken). Dabei werden 600 erwachsene Oberwalliserinnen und Oberwalliser befragt (siehe Excel-Datei Stichproben.xls). Berechnen Sie das geschätzte Total der Ausgaben für die erwachsene Oberwalliser Bevölkerung ($N = 45'000$) und ein 95%- Vertrauensintervall (98%-Vertrauensintervall).

3. Eine Studie soll den Anteil der Befürworter höherer Militärausgaben unter der stimmberechtigten Bevölkerung der Schweiz ermitteln. Es wird eine Zufallsstichprobe von 700 Personen gezogen. Es ergeben sich die folgenden Resultate (siehe Excel-Datei Stichproben.xls). Ermitteln Sie den Anteil und ein 95%-Vertrauensintervall. Wie gross muss n sein, um eine Vertrauensintervallbreite von 0.02 (entspricht $\pm 1\%$) zu erhalten? ($N = 4500000$) Bei Meinungsumfragen ist Schichtung die Regel, wir verzichten aber im Beispiel darauf).
4. Im Oberwallis soll das Mobilitätsverhalten von Erwachsenen untersucht werden. Dazu werden anteilmässig insgesamt 500 Personen befragt, unter anderem zur Frage, wie viele Kilometer sie ungefähr pro Jahr im Auto sitzen (Resultate siehe Excel-Datei Stichproben.xls). $N = 45'000$. Der Anteil der Personen, die im Beruf Auto fahren, betrage 0.33 (bezüglich der Menge aller Autofahrer). Versuchen Sie bei der Stichprobe eine Schichtung vorzunehmen. Bestimmen Sie die Vertrauensintervall mit und ohne Schichtung.
5. Im Oberwallis soll das Mobilitätsverhalten von Erwachsenen untersucht werden. Dazu werden 500 Personen befragt, unter anderem zur Frage, wieviele Kilometer sie ungefähr pro Jahr im Auto sitzen (Resultate siehe Excel-Datei Stichproben.xls). $N = 45'000$. Der Anteil der Personen, die im Beruf Auto fahren, betrage $\frac{1}{3}$ (bezüglich der Menge aller Autofahrer). Nehmen Sie bei der Stichprobe eine Nachschichtung vor. Bestimmen Sie die Vertrauensintervalle mit und ohne Nachschichtung.

31.6.2 Lösungen

1. Ein erhebliches Problem wird sich bei der Ziehung einer wirklichen Zufallsstichprobe ergeben. Man müsste dazu eigentlich eine Liste aller Touristen haben und dann daraus mit Hilfe von Zufallszahlen eine Stichprobe ziehen. Insbesondere dürfte man nicht z.B. in Brig im Sommer auf der Strasse einfach Touristen befragen.
Wir erhalten: Mittelwert: 49.348
Standardabweichung: 28.9928278

$$\bar{Y}_{gi} = \bar{Y} \pm \Phi^{-1}\left(\beta + \frac{1-\beta}{2}\right) \sqrt{\frac{1}{n} \left(1 - \frac{n}{N}\right) S^2}$$

$$49.348 + 1.96 \sqrt{\frac{1}{500} \left(1 - \frac{500}{1000000}\right)} 28.9928278 = 51.889$$

$$49.348 - 1.96 \sqrt{\frac{1}{500} \left(1 - \frac{500}{1000000}\right)} 28.9928278 = 46.807$$

$$\Rightarrow]46.807, 51.889[$$
98%-Vertrauensintervall:
$$49.348 + 2.326341928 \sqrt{\frac{1}{500} \left(1 - \frac{500}{1000000}\right)} 28.9928278 = 52.364$$

$$49.348 - 2.326341928 \sqrt{\frac{1}{500} \left(1 - \frac{500}{1000000}\right)} 28.9928278 = 46.332$$

$$\Rightarrow]46.332, 52.364[$$
2. Arithmetisches Mittel der Stichprobe: $\bar{y}_i = 611.1883333$
Varianz der Stichprobe: 21836.5196

$$Y \pm 1.96 \sqrt{\left(1 - \frac{600}{45000}\right) \frac{45000^2 \cdot 21836.5196}{600}} \Rightarrow$$

$$2.7503 \times 10^7 \pm 5.2853 \times 10^5$$

$$Y \pm 2.326341928 \sqrt{\left(1 - \frac{600}{45000}\right) \frac{45000^2 \cdot 21836.5196}{600}} \Rightarrow 2.7503 \times 10^7 \pm 6.2732 \times 10^5$$
3. Wir erhalten 400 Befürworter und 300 Gegner (mit Excel: zählenwenn). Damit beträgt der Anteil der Befürworter $\frac{400}{700} = 0.57143$.

$$P_A \pm 1.96 \sqrt{\left(1 - \frac{700}{4500000}\right) \frac{0.57143 \cdot (1 - 0.57143)}{700 - 1}} \Rightarrow$$

$$0.57143 \pm 3.6684 \times 10^{-2}$$
4. Wir berechnen mit Excel die Kennzahlen Mittelwert und Varianz für die Gesamtstichprobe und die Schichten:

Gesamtstichprobe: Mittelwert: 90888.902; Varianz: 81370215186
 Führt im Beruf: Mittelwert: 263203.9222; Varianz: 1.99747E+11
 Führt nicht im Beruf: Mittelwert: 4472.660661; Varianz: 1074205.466

Mit Schichtung erhalten wir den Schätzwert: $\frac{1}{45000}(15000 \cdot 263203.922 + 30000 \cdot 4472.660661) = 90716$.

Wir erhalten ohne Schichtung das 95%-Vertrauensintervall: $90888.902 \pm 1.96 \sqrt{\frac{81370215186}{500}} = 90888.902 \pm 24864.37517$.

Wir erhalten mit Schichtung das 95%-Vertrauensintervall: Die Varianz des Mittelwertes ist:

$$V(\bar{Y}) = \sum_{h=1}^L \left(W_h^2 \cdot \frac{1}{n_h} \cdot \left(1 - \frac{n_h}{N_h}\right) \cdot \hat{D}_h^2 \right) =$$

$$\left[\left(\frac{1}{3}\right)^2 \frac{1}{167} \left(1 - \frac{167}{\frac{1}{3} \cdot 45000}\right) 1.99747 \cdot 10^{11} \right] +$$

$$\left[\left(\frac{2}{3}\right)^2 \frac{1}{500-167} \left(1 - \frac{501-167}{\frac{2}{3} \cdot 45000}\right) 1074205.466 \right] = 131420562.6$$

$$90716 \pm 1.96 \sqrt{131420562.6} \Rightarrow 90716 \pm 22469.20634.$$

Weil die Mittelwerte der beiden Schichten stark auseinander liegen, wäre eigentlich eine starke Verminderung des Vertrauensintervalls zu erwarten. Dies tritt im Beispiel nicht ein - wegen der sehr grossen Varianz bei den Fahrern im Beruf (es gibt Verfahren, um bei solchen Fällen die Varianz der Mittelwerte zu vermindern, indem man bei der Schicht mit grosser Varianz eine grössere Stichprobe zieht).

5. Wir führen eine Nachschichtung durch: Personen, die beruflich im Auto unterwegs sind und Personen, die nicht im Auto unterwegs sind. Wir erhalten für die beiden Gruppen die folgenden Anteile: $\frac{378}{500} = 0.756$ für Privatfahrer und $\frac{122}{500} = 0.244$ für „Führt im Beruf“. Wir erhalten die folgenden Varianzen und Mittelwerte.

Privatfahrer: Mittelwert 4505.0344; Varianz: 1066213
 Führt im Beruf: Mittelwert 46784.69; Varianz: 430414.9
 Insgesamt ohne Schichtung: Mittelwert 14821.27; Varianz: $3.3 \cdot 10^8$
 Gewichtetes Gesamtmittel: $(4505.0344 \cdot 30000 + 46784.69 \cdot 15000) / 45000 = 18598$.

Die Gesamtmittel unterscheiden sich relativ stark, da die Stichprobenschichtgrössen nicht den Anteilen in der Grundgesamtheit entsprechen (in der Stichprobe tauchen nur 24.4% Profifahrer auf!). Der gewichtete Schätzer ist besser, da ohne Berücksichtigung der Schichten der Beitrag der Profifahrer unterschätzt wird.

Wir vergleichen die Vertrauensintervalle mit und ohne Schichtung:

Mit Schichtung: $\left(\frac{2}{3}\right)^2 \cdot \left(1 - \frac{378}{30000}\right) \cdot \frac{1}{378} \cdot 1066213 + \left(\frac{1}{3}\right)^2 \cdot \left(1 - \frac{122}{15000}\right) \cdot \frac{1}{122} \cdot 430414.9 = 1626.645$
 $\bar{Y}_S \pm 1.96 \sqrt{1626.645} \Rightarrow y_S \pm 79.05$
 $1.96 \sqrt{1626.645} = 79.05$ Ohne Schichtung: $\bar{Y}_{oS} \pm 1.96 \sqrt{\frac{3.3 \cdot 10^8}{500}} \Rightarrow \bar{y}_{oS} \pm 1592.3$

Da die Mittelwerte der Schichten sehr weit auseinander liegen, ergibt sich ein markanter Unterschied in den Vertrauensintervallen. (Beispiel 5 weist nicht dieselben Daten auf wie Beispiel 4. Somit sind die Unterschiede in den Resultaten der beiden Aufgaben nicht Effekten der Methoden - Schichtung oder Nachschichtung - zuzuschreiben).

31.7 Lernziele

- Die Problemlage beim Schätzen von Parametern einer Grundgesamtheit schildern können. Insbesondere den Unterschied zwischen Schätztheorie und Stichprobentheorie erläutern können (Worum geht es in der Schätztheorie im Gegensatz zur Stichprobentheorie).

- Die Bedeutung von Vertrauensintervallen bei der Angabe von Resultaten von Meinungsumfragen erläutern können.
- Den Wert der Schätzstatistiken Mittelwert, Total und Anteil sowie deren Vertrauensintervalle für einfache und geschichtete Stichproben berechnen können (Übungen der obigen Art machen können).

Teil VIII

Datenerhebung

Kapitel 32

Einige Bemerkungen zu Datenerhebung und -Bearbeitung

32.1 Planung der Datenerhebung

Bei der Planung der Datenerhebung ist wichtig, die Ziele der Untersuchung vor Augen zu haben. Geht es darum,

- Hypothesen zu testen oder
- möglichst präzise Parameter einer Grundgesamtheit zu schätzen.

32.1.1 Testen von Hypothesen

Beim Testen von Hypothesen sind im Allgemeinen relativ kleine Stichprobengrößen erforderlich, wobei der Stichprobenumfang von den verwendeten Testverfahren abhängen (von z.B. 16 nötigen Daten bis zu einigen Hundert bei multivariater ordinaler Regression). Entsprechend sind bei Datenerhebung die verwendeten Testverfahren im Voraus zu bestimmen, damit man die Stichprobengröße entsprechend anpassen kann. Zu beachten sind die verschiedenen Einschränkungen bei den Tests ($n_i > 50$; > 30 ; erwartete Häufigkeiten bei Chi-Quadrat-Test grösser als 5, etc.). Es ist bei der Berechnung der Stichprobengröße der Rückfluss zu beachten. Kennt man die Anteile in der Grundgesamtheit, so kann man z.B. bezüglich des Chi-Quadrat-Tests zum vornherein bestimmen, wie viele Untersuchungsobjekte man in etwa haben muss, damit die erwarteten Häufigkeiten genügend gross sind).

Bei Testverfahren ist zu beachten, dass bezüglich der Stichprobengröße „mehr nicht unbedingt besser ist“. Wie erwähnt (s. S. 381) führen sehr grosse Stichproben zu signifikanten Ergebnissen, auch wenn die entsprechenden Unterschiede minimal und inhaltlich bedeutungslos sind. In solchen Fällen geht es darum, die richtige Stichprobengröße zu bestimmen. Die entsprechenden Stichprobengrößen lassen sich aus den jeweiligen Formeln für die Teststatistik berechnen (s. für ein Beispiel S. 381). Bei verschiedenen Tests und nur einer Datenerhebung muss man sich nach der Notwendigkeit des Tests richten, der am meisten Daten verlangt. Bei der Berechnung der anderen Tests ist auf die inhaltliche Bedeutsamkeit zu achten, wenn Unterschiede statistisch signifikant sind.

Bezüglich Unabhängigkeit der Zufallsvariablen, als deren Werte wir die Daten betrachten, werden oft nur inhaltliche Überlegungen angestellt - Zufallsstichproben sind dazu nicht unbedingt erforderlich. Man spricht in diesem Fall von „willkürlichen Stichproben“. Untersucht man z.B. ob zwei Abfüllmaschinen die gleiche Varianz aufweisen, so werden wir z.B. die ersten 50 von diesen Maschinen produzierten Abfüllergebnisse berücksichtigen und nicht eine wirkliche Zufallsstichprobe aus ein paar Tausend produzierten Abfüllergebnisse wählen. Wir gehen stillschweigend davon aus, dass die Abfüllergebnisse pro spezifische Maschine unabhängig sind. Oder bei der

Qualitätskontrolle wird man nicht alle zu untersuchenden Objekte durchnummerieren, um eine Zufallsstichprobe zu ziehen. Kann man davon ausgehen, dass durch den Transport sich z.B. nicht die fehlerhaften Teile irgendwo ansammeln, so wird man eine willkürliche Stichprobe ziehen.

Bei Personenumfragen kann ein solches Verfahren schwieriger zu rechtfertigen sein. Befragt man Personen auf dem Bahnhof in Visp zu Ihrer Zufriedenheit mit dem touristischen Angebot im Oberwallis, so hat man in der Stichprobe Bahnfahrer und nicht Autofahrer. Es kann aber durchaus sein, dass sich Bahnfahrer bezüglich Zufriedenheit von Autofahrern unterscheiden. Möchte man aber geschlechtsspezifische Unterschiede bezüglich der touristischen Angebote überprüfen, ergibt sich mit der Umfrage am Bahnhof eventuell kein Problem, da vermutlich kein Zusammenhang zwischen den geschlechtsspezifischen Einstellungen und der Mobilitätsart besteht. Im Notfall kann man immer die Menge der zu untersuchenden Objekte einschränken: Menge der bahnfahrenden Touristen in Visp an einem schönen Frühjahrstag um Mittag.

Ähnliches gilt bezüglich Internet-Umfragen. Man kann dabei Umfragen unterscheiden, die sich an eine spezifische Gruppe von Personen richten, die z.B. nach Zufallsprinzip ausgewählt wurden (problemlos, wenn man kontrolliert, dass alle antworten), oder Umfragen, die ohne Kontrolle der Zielpersonen von diesen ausgefüllt werden. Ob solche unkontrollierten Umfragen angemessen sind, hängt von der Zielrichtung ab. Es ist sicher nicht möglich, Rückschlüsse auf die Grundgesamtheit zu ziehen (z.B. alle Schweizer oder alle Weinkonsumenten), da bei einer solche Umfrage spezifische Gruppen von Menschen (altermässig, soziale Schicht, Internet-Gewohnheiten, etc.) übervertreten sind und das Antwortverhalten gewöhnlich mit den untersuchten Merkmalen korreliert. Geht es aber z.B. um geschlechtsspezifische Einstellungsunterschiede, kann die Analyse durchaus instruktiv sein. Im Test überprüft man dann, ob z.B. eine Kreuztabelle mit der Hypothese vereinbar ist, dass die faktischen Häufigkeiten sich signifikant von den erwarteten Häufigkeiten unterscheiden. Dazu werden alle Informationen aus den vorliegenden Daten gezogen.

Bezüglich Internet-Umfrage ist ein besonderes Gewicht auf die inhaltliche Relevanz zu richten, da bei solchen Umfragen oft grosse Stichprobenumfänge zu Stande kommen, die zu irrelevanten, statistisch signifikanten Ergebnisse führen.

32.1.2 Schätzen von Parametern einer Grundgesamtheit

Geht es um die möglichst präzise Schätzung der Parameter einer Grundgesamtheit (Anteile, Mittel, Total), so sind fast unabhängig von der Grösse der Grundgesamtheit grosse Stichproben erforderlich. Es ist jeweils die Grundmenge der Untersuchungsobjekte sauber einzugrenzen (Menge der Walliser; der Jugendlichen im Oberwallis; der Touristen im Oberwallis im Jahr 2009; Menge der Objekte einer Lieferung, etc.) und aus dieser eine Zufallsstichprobe zu ziehen. Bei Schichtung, die eine Zerlegung der Grundgesamtheit liefern muss, ist aus jeder Schicht eine Zufallsstichprobe zu ziehen. Gut gewählte Schichtung führt zu besseren Ergebnissen, d.h. zu kleineren Standardfehlern der Schätzer (s. entsprechendes Kapitel auf Seite 561).

Welche Parameter welcher Grundgesamtheit möchte ich kennen? Parameter und Grundgesamtheit aufschreiben. Wie exakt sollten die Ergebnisse sein? Gewünschte Vertrauensintervalle aufschreiben und Stichprobengrösse berechnen. Achtung: je nach Grundgesamtheit ergeben sich unterschiedliche Rückflüsse. Der schätzungsweise Rückfluss muss berücksichtigt werden. Dieser kann durch eine Miniuntersuchung geschätzt werden. Man kann sich auch beim Bundesamt für Statistik für entsprechende Erfahrungszahlen erkundigen. Die Stichprobengrössen müssen entsprechend angepasst werden (Achtung: systematische Verfälschung der Resultate durch Zusammenhang des Antwortverhaltens mit zu untersuchenden Variablen!).

32.2 Methoden der Stichprobenerhebung

Je nach den Daten die man erhebt, sind spezifische Methoden nötig:

- Daten aus Datenbanken
- Daten per Umfrage: Umfrage auf der Strasse, per Telefon, per Internet, an der Haustüre

- Daten aus empirischen Untersuchungen an Objekten (z.B. Krankheitsbilder, Qualitätsprüfung von Lieferungen, etc.).

Die Methoden werden zudem unterschiedlich sein, ob es

- darum geht, Parameter einer Grundgesamtheit zu bestimmen und je nach dabei verwendetem Verfahren (Zufallsstichprobe, geschichtete Zufallsstichprobe, gemischte Stichprobenpläne, s. Kapitel „Stichprobentheorie“, Kapitel 31, S. 561 ff.)
- oder ob es um Tests geht (s. oben bezüglich Stichprobengrößen und Stichprobenverfahren).

Bei Umfragen ist zu beachten:

1. Nur Daten erheben, welche in einem Zusammenhang schriftlich festgelegten Zielen stehen (Lange Fragebogen ermüden; die Menschen werden der beständigen Fragerei langsam überdrüssig!).
2. Bei der Fragebogenerstellung:
 - (a) Sich fragen, wie man die Antworten in eine Tabelle von Urdaten überträgt (Genau die Variablen festlegen, die verwendet werden. Ausprägungen festlegen).
 - (b) Festhalten, welche Tests man zur Überprüfung welcher Hypothesen auf welche Variablen anwenden wird.
 - (c) Fragen mit Mehrfachantworten vermeiden. Typische Fragen mit Mehrfachantworten: „Kreuzen Sie zutreffendes an“, „Wählen Sie aus den folgenden Antwortmöglichkeiten die drei zutreffendsten“. Bei der Erfassung der Antworten solche Fragen muss für jede Antwortmöglichkeit eine dichotome Variable erzeugt werden. Es ergeben sich viele fehlende Antworten. Man sollte entsprechend Mehrfachantworten nur ermöglichen, wenn man die z.B. Anzahl der Antworten zählt oder eine Häufigkeitsauszählung der gegebenen Antworten vornimmt.
3. Fragebogen müssen an Personen der festgelegten Grundgesamtheit getestet werden. Dazu gibt es recht anspruchsvolle Verfahren, die wir hier nicht behandeln können. Statt dieser Verfahren kann man aber auch pragmatisch vorgehen: Wir testen die Fragebogen an Versuchspersonen, indem wir mit ihnen diskutieren - können sie antworten, wie sie möchten? sind die Frage klar? etc. Ist das Antwortverhalten unterschiedlich genug - oder geben alle oder fast alle Personen dieselbe Antwort? Wenn nötig, Änderungen am Fragebogen vornehmen (s. für weitere Einstiegsinformationen (Diekman, 2010)).

32.3 Behandlung fehlender Daten

Fehlende Daten können die Schlussfolgerungen aus den Daten verfälschen. Dies ist der Fall, wenn zwischen dem Antwortverhalten von Personen und den untersuchten Eigenschaften ein Zusammenhang besteht. Zudem können fehlende Daten dazu führen, dass man nicht alle Information aus einem Datensatz nutzen kann. Fehlt bei einer multiplen Regression z.B. der Wert der unabhängigen Variable und wird der Datensatz für die Analyse aus den Daten gestrichen, so wird nicht alle zur Verfügung stehende Information verwendet. Entsprechend wurden Methoden entwickelt, um mit fehlenden Daten umzugehen (Little & Rubin, 2002).

Analyse der vollständigen Fälle

Eine Möglichkeit, mit unvollständigen Datensätzen umzugehen, besteht darin, die unvollständigen Datensätze zu streichen. Der Vorteil des Verfahrens besteht in seiner Einfachheit. Zudem liegt für unterschiedliche Tests nachher jeweils dieselbe Datenbasis vor. Der Nachteil ist der Informationsverlust, eine erhöhte Varianz (kleineres n) und entsprechend eine kleinere Macht der Tests. Zudem kann eine systematische Verzerrung vorliegen, wenn das Antwortverhalten mit den untersuchten Eigenschaften zusammenhängt.

Analyse der verfügbaren Fälle

Für jeden Test werden aus der Datentabelle alle nutzbaren Daten verwendet, selbst wenn der Datensatz sonst nicht vollständig ist. Das Verfahren ist einfach und verwendet mehr Informationen als die Verwendung nur der vollständigen Fälle, kann aber zu inkonsistenten Resultaten führen, da nicht für jede Berechnungen eine identische Menge von Datensätzen verwendet wird.

Einfach Imputation

Für die fehlenden Werte werden plausible Werte eingesetzt (imputare). Man ersetzt z.B. fehlende Werte der Variable X durch den Mittelwert der vorliegenden Daten der Variable X . Dieses Vorgehen ist nicht empfehlenswert, da dadurch die Varianz vermindert wird und Hypothesentests nicht mehr gültig sind - die ja jeweils auf der geschätzten Varianz ruhen. Statt des Mittelwertes kann man z.B. mit den vorliegenden Daten eine Regressionsgerade schätzen und fehlende Werte auf Grund der Regressionsgerade schätzen. Die Varianz wird hier nicht so stark unterschätzt wie beim Mittelwert, da die erklärte Varianz nicht wegfällt. Der Varianz der Residuen wird aber immer noch nicht Rechnung getragen, so dass sich dieselben Probleme stellen wie bei der Mittelwertmethode.

Um die Varianz nicht zu verkleinern, kann man Daten aus einer Verteilung nach Zufallsprinzip wählen. Man würde ein Verteilungsmodell für die Residuen schätzen (s. entsprechendes Kapitel „Schätzen von stetigen Wahrscheinlichkeitsverteilungen durch Daten“, S. 295) und gemäss diesem Verteilungsmodell ein Residuum wählen. Normalverteilte Daten kann man mit R z.B. erzeugen mit `rnorm`(Anzahl, Erwartungswert, Standardabweichung). Ist man also mit Hilfe eines P-P-Plots zum Schluss gekommen, dass eine Normalverteilung der Residuen vorliegt, kann man die Parameter der Normalverteilung durch den Erwartungswert (sollte bei Residuen 0 sein) und die Standardabweichung der Residuen schätzen und diese dann für die Erzeugung von Residuen verwenden. Das geschätzte Datum wird aus dem Schätzwert auf der Regressionsgerade und dem Residuum zusammengesetzt (Addition).

Eine weitere Imputations-Methode ist die Substitution. Man ersetzt einen Datensatz durch einen solchen aus einer anderen Erhebung. Nimmt z.B. jemand das Telefon nicht ab, so wählt man eine Person, die nicht auf der Liste der zu befragenden Personen vorkommt. Bei strukturidentischen Datensätzen (gleiche Variablen) werden Datensätze auch aus anderen Erhebungen ersetzt (Cold-Deck-Imputation) oder aus einem Pool von ähnlichen Datensätzen der vorliegenden Erhebung nach Zufallsprinzip gewählt (Hot-Deck-Imputation). Beim letzten Verfahren wird die Varianz wiederum vermindert. Bei der Cold-Deck-Imputation sind die Auswirkungen theoretisch kaum abschätzbar.

Es gibt weitere, komplexere Imputationsmethoden, für deren Darlegung uns die Grundlagen fehlen (EM-Algorithmus und Maximum Likelihood-Methoden).

Übung 32.3.1. *Datensatz mit zwei fehlenden Daten (zwei metrisch skalierte Variablen, s. Excel-Datei 228Imputation.xls, Tabelle 1). Schätzen der Regressionsgerade. Schätzen der Verteilung der Residuen. Wahl aus der Verteilung, um fehlendes Datum zu imputieren.*

Multiple Imputation

Die Methode der Einfachen Imputation (z.B. Regression mit W-Modell für Residuen) wird mehrfach (D-fach) angewendet, um mehrere komplette Datensätze zu erhalten. Aus den D Datensätze werden die Kennwerte berechnet und zusammengefasst (z.B. Mittel der erhaltenden Kennwerte). Mit der Methode sieht man, wie abhängig die Ergebnisse von den Imputationen sind.

Übung 32.3.2. *Datensatz mit zwei fehlenden Daten (zwei metrisch skalierte Variablen; s. oben). Schätzen der Regressionsgerade. Schätzen der Verteilung der Residuen. Wahl aus der Verteilung, um fehlendes Datum zu imputieren. 20-fache Wiederholung des Vorgangs. Bestimmung der Varianz der Kennwerte (z.B. des Mittelwertes der Variable mit den zwei fehlenden Daten).*

32.4 Auswertung

1. Nachdem die Daten erhoben und in eine Tabelle der Urdaten übertragen sind, zuerst eine univariate Analyse vornehmen (Je nach Skalenniveau und Anzahl der Ausprägungen):
 - (a) Häufigkeitsverteilungen - Diagramme; eindimensionale Streudiagramme; empirische Verteilungsfunktion.
 - (b) Kennzahlen der zentralen Lage (Mittelwert, Median, Modus) und der Streuung (Standardabweichung, Varianz, Quartile).
 - (c) Sind durch die zu testenden Hypothesen bivariate Tests vorgesehen, eine bivariate beschreibende Analyse der Daten vornehmen (Gamma, bivariate Streudiagramme, etc. - je nach Skalenniveau).
 - (d) Die Ergebnisse genau studieren und diskutieren. Die dabei gewonnenen Einsichten schriftlich festhalten.
2.
 - (a) Beim Schätzen von Parametern der Grundgesamtheit: Parameter berechnen, Vertrauensintervalle kontrollieren (empirische Stichprobengrösse!). Diskutieren, ob sich eventuell durch das Rücklaufverhalten systematische Verzerrungen ergeben. Anteile an der Stichprobe mit den Anteilen in der Grundgesamtheit vergleichen (sind die Anteil der Frauen an der Stichprobe signifikant von den Anteilen der Frauen in der Grundgesamtheit verschieden?).
 - (b) Die Tests durchführen. Achtung: man sollte nur Tests durchführen, für die man Hypothesen aufgestellt hat. Führt man andere Tests durch, sollten die Ergebnisse nur als Hinweise für künftige Untersuchungen betrachtet werden (Siehe entsprechende Überlegungen beim bivariaten Chiquadrattest - Bemerkung in den Lösungen zu den Übungen). Die strikte Beachtung dieser Anforderung ist für seriöse Untersuchungen unerlässlich!

Anhang A

Liste der Symbole

k	absolute Häufigkeitsverteilung, S. 23
h	relative Häufigkeitsverteilung, S. 34
$\bar{x}$	das arithmetische Mittel des Datenvektors $x = (x_1, \dots, x_n)$, S. 44
$\tilde{x}$	der Median des Datenvektors $x = (x_1, \dots, x_n)$, S. 49
s^2	Varianz eines Datenvektors, S. 60
s	Standardabweichung eines Datenvektors, S. 61
k_{ij}	absolute gemeinsame Häufigkeit zweier Variablen der Ausprägungen i und j , S. 92
k_{+j}	Spaltensumme (der j -ten Spalte) der absoluten gemeinsamen Häufigkeitsverteilung, S. 92
k_{i+}	Zeilensumme (der i -ten Zeile) der absoluten gemeinsamen Häufigkeitsverteilung, S. 92
OR	Odds Ratio (Chancenverhältnis), S. 97
K	Kontingenzkoeffizient, S. 99
K^*	korrigierter Kontingenzkoeffizient, S. 100
τ	Goodman und Kruskal's tau, S. 100
γ	Gamma - Kennzahl für den Zusammenhang klassierter ordinalskalierter Daten, S. 102
r_{xy}	Pearson-Korrelation der Variablen X und Y , S. 104
r_S	Spearman-Rang-Korrelation zweier Variablen, S. 110
$r_{xy/z}$	partielle Korrelation der Variablen X und Y unter Kontrolle der Variable Z , S. 123
$n!$	n-Fakultät, S. 145
$\binom{n}{m}$	n tief m (Binomialkoeffizient), S. 147
ω	Ergebnis eines Zufallsexperimentes, S. 161
Ω	Ergebnismenge (Menge von Ergebnissen eines Zufallsexperimentes), S. 161
$\mathcal{P}(\Omega)$	Menge aller Ereignisse (= Potenzmenge von Ω), S. 162
A^c	Gegenereignis des Ereignisses A , S. 163
$f(\omega)$	Wahrscheinlichkeit des Eintreffens des Ergebnisses ω ; Wert der Wahrscheinlichkeitsfunktion in ω , S. 165
$P(A)$	Wahrscheinlichkeit des Ereignisses A , Wert der Wahrscheinlichkeitsverteilung P in A , S. 166
$P(A B)$	Wahrscheinlichkeit von A unter der Bedingung B , S. 166
X	Zufallsvariable, S. 197
$\mathfrak{X}$	Bild der Zufallsvariable X , S. 197
$f_X(x)$	Wahrscheinlichkeit des Eintreffens von x , Wert der Bildwahrscheinlichkeitsfunktion in x bei diskret verteilten Zufallsvariablen, S. 203, Wert der Dichtefunktion der Zufallsvariable X bei stetig verteilten Zufallsvariablen, S. 274
$P_X(A)$	Wahrscheinlichkeit des Eintreffens des Ereignisses A , Wert der

	Bildwahrscheinlichkeitsverteilung P_X in A , 199
$P(X = x)$	die Wahrscheinlichkeit, dass die Zufallsvariable X den Wert x annimmt, S. 200
$P(a < X < b)$	die Wahrscheinlichkeit, dass die Zufallsvariable X einen Wert grösser als a und kleiner als b annimmt, S. 200
F_X	Verteilungsfunktion der Zufallsvariable X , S. 201
$E(X), E_P(X)$	Erwartungswert der Zufallsvariable X unter der Wahrscheinlichkeitsverteilung P , S. 203
$V(X), V_P(X)$	Varianz der Zufallsvariable X unter der Wahrscheinlichkeitsverteilung P , S. 206
$P_{(X_1, X_2)}$	gemeinsame Wahrscheinlichkeitsverteilung auf das kartesische Produkt der Bilder $\mathfrak{X}_1$ und $\mathfrak{X}_2$ der Zufallsvariablen X_1 und X_2 , S. 210
$f_{(X_1, X_2)}$	gemeinsame Wahrscheinlichkeitsfunktion auf das kartesische Produkt der Bilder $\mathfrak{X}_1$ und $\mathfrak{X}_2$ der Zufallsvariablen X_1 und X_2 , S. 211
$\sum_{i=1}^n X_i$	Statistik Summe der Werte der Zufallsvariablen X_i , S. 216
$\bar{X}_n$	Statistik Mittelwert (Mittelwert der Werte von n Zufallsvariablen), S. 218
$\prod_{i=1}^n X_i$	Statistik Produkt der Werte der Zufallsvariablen X_i , S. 219
$KOV(X_1, X_2)$	Kovarianz der Zufallsvariablen X_1 und X_2 , S. 220
$KOR(X_1, X_2)$	Korrelation der Zufallsvariablen X_1 und X_2 , S. 221
S^2	Statistik Varianz der Werte von Zufallsvariablen, S. 224
$X \sim B(1, p)$	die Zufallsvariable X ist bernoulliverteilt mit dem Parameter p , S. 237
$X \sim B(n, p)$	die Zufallsvariable X ist binomialverteilt mit den Parametern n und p , S. 239
$X \sim P(\lambda)$	die Zufallsvariable X ist poissonverteilt mit dem Parameter λ , S. 243
$X \sim H(N, n, M)$	die Zufallsvariable X ist hypergeometrisch verteilt mit den Parametern N (Grösse Grundgesamtheit), n (Stichprobengrösse) und M (Grösse der ausgezeichneten Ergebnisse in der Grundgesamtheit), S. 251
$X \sim U(a, b)$	die Zufallsvariable X ist stetig uniform verteilt mit den Parametern a und b , S. 283
$X \sim Exp(\lambda)$	die Zufallsvariable X ist exponentialverteilt mit dem Parameter λ , S. 288
$X \sim N(\mu, \sigma^2)$	die Zufallsvariable X ist normalverteilt mit den Parametern μ und σ^2 , S. 305
Φ	Verteilungsfunktion der Standardnormalverteilung, S. 312
χ_n^2	Chi-Quadrat-Verteilung mit dem Parameter n , S. 347
t_n	t-Verteilung mit dem Parameter n , S. 352
$F_{m,n}$	F-Verteilung mit den Parametern m und n , S. 355

Literatur

- Agresti, A. (1990). *Categorical Data Analysis*. New York: John Wiley.
- Agresti, A. (2002). *Categorical Data Analysis* (2. Aufl.). New York: John Wiley.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden: eine anwendungsorientierte Einführung*. Berlin: Springer.
- Baclawski, K. (2008). *Introduction to Probability with R*. London: Chapman and Hall.
- Bauer, T. K., Gigerenzer, G. & Krämer, W. (2014). *Warum dick nicht doof macht und Genmais nicht tötet: Über Risiken und Nebenwirkungen der Unstatistik*. Frankfurt: Campus Verlag.
- Beck-Bornholdt, H. & Dubben, H. (2001). *Der Hund, der Eier legt: Erkennen von Fehlinformation durch Querdenken*. Reinbek: Rowohlt.
- Bleymüller, J., Gehlert, G. & Gülicher, H. (2000). *Statistik für Wirtschaftswissenschaftler*. München: Vahlen.
- Bortz, J. (1984). *Lehrbuch der Statistik für Sozialwissenschaftler*. Berlin: Springer.
- Bortz, J. (2008). *Verteilungsfreie Methoden in der Biostatistik*. Berlin: Springer.
- Büning, H. & Trenkler, G. (1994). *Nichtparametrische statistische Methoden*. Berlin: Walter de Gruyter.
- Chambers, J. M. (2008). *Software for Data Analysis: Programming with R*. Berlin: Springer.
- Cochran, W. G. (1977). *Sampling Techniques*. New York: John Wiley and Sons.
- Dalgaard, P. (2008). *Introductory Statistics with R*. Berlin: Springer.
- Diekman, A. (2010). *Empirische Sozialforschung: Grundlagen, Methoden, Anwendung*. Reinbek: Rowohlt.
- Fahrmeir, L., Künstler, R., Pigeat, I. & Tutz, G. (2004). *Statistik: der Weg zur Datenanalyse*. Berlin: Springer.
- Hartung, J. (1999). *Statistik: Lehr- und Handbuch der angewandten Statistik*. München: Oldenbourg.
- Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.
- Hippmann, H.-D. (1997). *Statistik für Wirtschafts- und Sozialwissenschaftler* (2. Aufl.). Stuttgart: UTB Schäffer-Poeschel.
- Hosmer, D. W. & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons.
- Hunziker, A. & Scheerer, F. (1975). *Statistik: Instrument der Betriebsführung*. Stuttgart: Poeschel.
- Hüsler, J. (Hrsg.). (2009). *Angewandte Statistik: Skripten des Nachdiplomkurses der Universität Bern*. Bern: Universität Bern: Institut für mathematische Statistik und Versicherungslehre.
- Krämer, W. (1997). *So lügt man mit Statistik*. Frankfurt a. M.: Campus.
- Krenger, U. (1998). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Braunschweig: Vieweg.
- Little, R. & Rubin, D. (2002). *Statistical analysis with missing data*. New York: John Wiley and Sons.
- Lumley, T. (2010). *Complex Surveys: A Guide to Analysis Using R*. New York: John Wiley and Sons.
- Müller, P. H. (1991). *Lexikon der Stochastik*. Berlin: Akademie Verlag.
- Moeschlin, O. (1996). *Diskrete Stochastik*. Hagen: Fernuniversität Hagen.
- Moeschlin, O. (1997). *Testtheorie*. Hagen: Fernuniversität Hagen.
- Moeschlin, O. (1998a). *Schätztheorie*. Hagen: Fernuniversität Hagen.
- Moeschlin, O. (1998b). *Wahrscheinlichkeitstheorie I: Mass- und Wahrscheinlichkeitstheorie*. Hagen: Fernuniversität Hagen.
- Moeschlin, O. (2001). *Angewandte Statistik*. Hagen: Fernuniversität Hagen.
- Moeschlin, O., Pohl, C. & Steinert, F. (1998). *Experimental Stochastics*. Berlin: Springer.
- Riedwyl, H. (1980). *Regressionsgerade und Verwandtes*. Bern: UTB Haupt.
- Schlittgen, R. (1997). *Einführung in die Statistik*. München: Oldenbourg.
- Särndal, C.-E., Swensson, B. & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.

- Stahel, W. (2013, September). *Lineare regression*. Internet. Verfügbar unter <http://stat.ethz.ch/~stahel/courses/regression/reg1-script.pdf> (Seminar für Statistik der ETH Zürich, Unterlagen zum Teil 1 des Kurses Angewandte Regression)
- Stahel, W. A. (1995). *Statistische Datenanalyse: Eine Einführung für Naturwissenschaftler*. Braunschweig: Vieweg.
- Thompson, S. K. (2012). *Sampling* (3. Aufl.). Hoboken: Wiley.
- Wewel, M. C. (2006). *Statistik im Bachelor-Studium der BWL und VWL: Methoden, Anwendung, Interpretation*. München: Pearson.