

Le Gamma de Goodman et Kruskal

Le Gamma de Goodman et Kruskal exprime la corrélation entre deux variables ordinales (en général avec peu de valeurs des variables). Il s'utilise par conséquent pour mesurer la corrélation entre des variables souvent appelées de type Likert. Le Gamma peut cependant aussi être utilisé pour mesurer une différence entre deux groupes par rapport à une variable ordinale de ce type. Pour le comprendre il faut savoir comment le Gamma se calcule.

On considère une tableau croisé $I \times J$ (qui contient I lignes et J colonnes) Soit a_{ij} la fréquence d'une cellule de ce tableau croisé (i pour la ligne, j pour la colonne, alors a_{23} est la fréquence de la cellule qui se trouve dans la deuxième ligne et la troisième colonne).

Exemple 1. *tableau croisé 4×5 , la cellule $a_{3,2} = 7$.*

5	6	1	2	3
4	18	5	3	8
2	7	39	7	2
0	9	4	3	40

Le calcul de Gamma

1. On calcule pour chaque cellule a_{ij} le produit $a_{ij} \cdot a_{i+k,j+r}$ pour $i < k \leq I$ et $j < r \leq J$. Ensuite on calcule la somme de ces produits. C'est le nombre des couples dits concordants Π_c

Exemple 2. *Si on a sur les deux axes un ordre croissant, p.ex. de plus en plus riche (lignes) et de plus en plus content (colonnes), Π_c exprime le nombre des couples en concordance avec l'hypothèse d'une corrélation positive entre richesse et satisfaction. Si on considère la cellule a_{11} et a_{22} du tableau croisé suivant, il y a $5 \cdot 7$ couples tel que les personnes de la cellule a_{22} sont à la fois plus riches et plus contentes que les personnes de la cellule a_{11} .*

	<i>très pauvre</i>	<i>pauvre</i>	<i>moyen</i>	<i>riche</i>
<i>mécontent</i>	5	6	1	2
<i>assez content</i>	2	7	39	7
<i>content</i>	0	9	4	3

Dans l'exemple on calcule

$$\begin{aligned} \Pi_c = & 5 \cdot 7 + 5 \cdot 39 + 5 \cdot 7 + 5 \cdot 9 + 5 \cdot 4 + 5 \cdot 3 + \\ & 2 \cdot 9 + 2 \cdot 4 + 2 \cdot 3 + \\ & 6 \cdot 39 + 6 \cdot 7 + 6 \cdot 4 + 6 \cdot 3 + \\ & 1 \cdot 7 + 1 \cdot 3 + \\ & 7 \cdot 4 + 7 \cdot 3 + \\ & 39 \cdot 3 = 871 \end{aligned}$$

2. Enfin, on calcule pour chaque cellule a_{ij} le produit $a_{ij} \cdot a_{i-k, j-r}$ pour $i < k \leq I$ et $j < r \leq J$. Ensuite on calcule la somme de ces produits. C'est le nombre des couples dits discordants Π_d .

Exemple 3. Si on a sur les deux axes un ordre croissant, p.ex de plus en plus riche et de plus en plus content, Π_d exprime le nombre des couples qui contredisent l'hypothèse d'une corrélation positive entre richesse et satisfaction. Si on considère la cellule a_{24} et a_{32} du tableau croisé de l'exemple 2, il y a $7 \cdot 9$ couples tel que les personnes de la cellule a_{24} sont à la fois plus riches et moins contentes que les personnes de la cellule a_{32} . Dans l'exemple 2 on calcule

$$\begin{aligned}\Pi_d = & 2 \cdot 39 + 2 \cdot 7 + 2 \cdot 2 + 2 \cdot 4 + 2 \cdot 9 + 2 \cdot 0 + \\ & 1 \cdot 7 + 1 \cdot 2 + 1 \cdot 9 + 1 \cdot 0 + \\ & 6 \cdot 2 + 6 \cdot 0 + \\ & 7 \cdot 4 + 7 \cdot 9 + 7 \cdot 0 + \\ & 39 \cdot 9 + 39 \cdot 0 + \\ & 7 \cdot 0 = 594\end{aligned}$$

3. Le gamma est ensuite défini par

$$\frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d}$$

Dans l'exemple 2 on obtient :

$$\frac{871 - 594}{871 + 594} = 0.18908$$

Gamma est un nombre entre -1 et 1. Si les fréquences se concentrent sur la diagonale principale, Gamma est positif, si les fréquences se concentrent sur la contre-diagonale (de haut à droite vers le bas à gauche), Gamma est négatif. Pour interpréter le signe de Gamma, il faut toujours tenir compte des noms des valeurs des variables et de leur ordre.

Comparer deux groupes par Gamma

Si on applique Gamma à un tableau croisé $2 \times J$, Gamma ne mesure plus une corrélation, mais une différence entre les groupes de la variable dichotome par rapport à la variable ordinale. Considérons un exemple :

	mécontent	peu content	content	très content
hommes	40	10	4	2
femmes	5	7	12	30

Le nombre des couples concordants exprime le nombre de fois qu'une femme est plus contente qu'un homme et le nombre des couples discordants exprime le nombre de fois qu'une femme est moins contente qu'un homme. Cette réflexion montre que Gamma est

utile pour comparer deux groupes par rapport à une variable ordinale. Traditionnellement on utilisait en sciences sociales un test t dans ce contexte qui compare des moyennes. C'est critiquable : premièrement le calcul des moyennes n'a pas de sens pour des variables ordinales. Le test t suppose en comparant des moyennes des variables métriques. Deuxièmement ce test suppose la distribution normale pour les données des deux groupes. Il n'y a cependant pas de sens de parler d'une distribution normale par rapport à des données ordinales.

Finalement on peut utiliser Gamma pour des tableaux 2×2 . Gamma exprime, si c'est positif, que le produit $a_{11}a_{22}$ est plus grand que le produit $a_{21}a_{12}$. Il y a alors différence entre les variables des lignes : les fréquences se concentrent dans les cellules a_{11} et a_{22} . Prenons l'exemple du genre (homme ou femme) et du tabagisme (fumeur oui ou non)

	fumeur	non-fumeur
homme	60	30
femme	20	50

On voit tout de suite, que proportionnellement les hommes fument plus que les femmes. Gamma donne :

$$\frac{60 \cdot 30 - 30 \cdot 20}{60 \cdot 30 + 30 \cdot 20} = 0.5 > 0$$

Le quotient des cotes (OR = odds ratio) donne :

$$\frac{\frac{60}{30}}{\frac{20}{50}} = \frac{60 \cdot 50}{20 \cdot 30} = 5 > 1$$

Le quotient des cotes est plus grand que 1 si et seulement si Gamma est plus grand que 0. Le quotient des cotes se situe entre 0 et 1 si et seulement si Gamma est inférieur à 0.

Tests statistiques

Les tests d'inférence permettent de voir si le gamma dévie significativement de 0 ou pas — pour un test bilatéral (p.ex. il y a différence entre les hommes et les femmes par rapport à la variable ordinale).

Un test unilatéral à droite vérifie, si le Gamma est plus grand que 0 (p.ex. si les femmes sont à la première ligne et les hommes à la deuxième et que la variable ordinale contient dans l'ordre suivant les valeurs "peu content", "assez content", "content" et "très content", un test unilatéral à droite suppose l'hypothèse que les hommes sont plus contents que les femmes).

Un test unilatéral à gauche vérifie si le Gamma est inférieur à 0. (p.ex. si les femmes sont à la première ligne et les hommes à la deuxième et que la variable ordinale contient dans l'ordre suivant les valeurs "peu content", "assez content", "content" et "très content", un test unilatéral à gauche suppose l'hypothèse que les hommes sont moins contents que les femmes).

Pour des échantillons assez grands on utilise une approximation normale (voir Agresti (1990, p. 58)). Des logiciels comme SAS et SPSS offrent pour des échantillons petits des

tests exacts basés sur la combinatoire. Finalement on peut calculer des tests basés sur la méthode Monte Carlo.

(rédigé par Paul Ruppen, support statistique de la HEP-VS, www.hep-vs.ch, 19 septembre 2021)

Références

Agresti, A. (1990). *Categorical Data Analysis*. New York : John Wiley.

Goodman, L. A., & Kruskal, W. H. (1954). Measures of Association for Cross Classifications. *Journal of the American Statistical Association*(49), 732–764.