

# Coefficients d'association (Regard symétrique ou asymétrique)

(Variable indépendante)

Nominal	Ordinal	Intervalles
$\chi^2$ Coeff. de contingence C $\Phi$ (phi) V de Cramer $\lambda$ de Goodman et Kruskal $\lambda_x \lambda_y \lambda_{xy}$		
	Coefficients ordinaux Q de Yule $\gamma$ (Gamma) $\tau_b$ (tau b) $\tau_c$ (tau c) D de Somer	
$\eta$ (Eta)		$r$ (de Pearson)

**Nominal**

**Ordinal**

**Intervalles**

(Variable dépendante)

**Examinez les tableaux croisés soigneusement avant de résumer la relation à l'aide d'un coefficient.**  
 (si vous souhaitez utiliser un coefficient!)

... et dans le doute retournez aux tableaux croisés....

## Tests statistiques et coefficients d'association (tableaux croisés)

La colonne de gauche indique le niveau de mesure minimal pour l'utilisation du coefficient correspondant: *n*= nominal, *o*=ordinal, *o+* =ordinal métrique, *i*=intervalles, *d*=dichotomiques

- n**  $\chi^2$  **Chi-deux (khi-deux) [(Pearson) Chi-square]** Statistique (test) utilisée pour tester l'hypothèse que deux variables sont indépendantes. A ne pas utiliser si une des cellules a une valeur théorique inférieure à 1, ou si plus de 20% des cellules ont des valeurs théoriques inférieures à 5. Le  $\chi^2$  **de Pearson** est la forme la plus courante.
- n** • La *correction de Yates* [Yates' correction for continuity] est appliquée au calcul du  $\chi^2$  pour les tableaux 2x 2. Cette correction vise à améliorer l'approximation. Les valeurs du  $\chi^2$  corrigées sont toujours plus petites que les valeurs non corrigées. Si le nombre d'observations d'un tableau 2x2 est inférieur à 20, le test de Fisher (Fisher's exact test) est calculé.
- o+** •  $\chi^2$  **de Mantel-Haenszel: (Linear-by-linear association)** Mesure d'intensité de la liaison linéaire. Le  $\chi^2$  de Mantel-Haenszel se calcule en multipliant le coefficient de corrélation de Pearson (*r*) par le nombre d'observations moins 1. Il a un seul degré de liberté. S'applique aux données au moins ordinales.
- n** •  $\chi^2$  **du rapport de vraisemblance** [likelihood ratio  $\chi^2$ ] Statistique de qualité de l'ajustement similaire au  $\chi^2$  de Pearson. Pour les échantillons de grande taille, les deux statistiques se valent.
- d** **Test de McNemar** pour variables dichotomiques, basé sur la distribution du chi-2.
- n**  $\phi = \sqrt{\frac{\chi^2}{n}}$  (**Phi**) Mesure de liaison basée sur le  $\chi^2$ .  $\Phi$  prend le signe du coefficient de corrélation de Pearson. Pour les tableaux 2x2 ;  $\Phi$  varie entre 0 et 1. Pour les tableaux plus grands que 2x2,  $\Phi$  peut dépasser 1.
- n**  $V = \frac{\phi^2}{\sqrt{\min(c-1, l-1)}}$  **V de Cramer**: Mesure de liaison basée sur le  $\chi^2$ . Le V de Cramer est toujours compris entre 0 et 1. Le V est une extension du  $\Phi$  aux tableaux plus grands que 2x2. (*l*=nombre de lignes, *c*=nombre de colonnes)
- n**  $C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$  **Coefficient de contingence**: Mesure de liaison sur la base du  $\chi^2$ . Ce coefficient est toujours compris entre 0, mais la valeur maximale (toujours inférieures à 1) dépend des dimensions du tableau.
- n**  $\lambda$  (**Lambda**) Mesure de liaison de la réduction proportionnelle de l'erreur (PRE) de prédiction lorsque une variable indépendante est introduite pour prévoir/expliciter une variable dépendante. Une valeur de 1 signifie que la variable explicative prédit parfaitement la variable à expliquer. Il existe également une forme symétrique du  $\lambda$ .
- n**  $\tau$  (**tau**) **de Goodman et Kruskal**: Mesure de la réduction proportionnelle de l'erreur comparant la prévision calculée sur les proportions marginales seules à celle où les valeurs d'une variable supplémentaire sont connues. A la différence de  $\lambda$ , où la catégorie modale est utilisée pour les prévisions,  $\tau$  utilise les proportions marginales.
- n** **Coefficient d'incertitude**: [uncertainty coefficient] Mesure de la réduction proportionnelle de l'erreur basée sur des critères d'entropie. Plus le coefficient d'incertitude se rapproche de sa borne supérieure (1), plus on obtient d'informations sur la valeur de la seconde variable d'une observation à partir de la valeur de la première. Une version symétrique est également proposée.
- o**  $\tau_b$  (**tau-b**) **de Kendall** (souvent appelé coefficient de corrélation de Kendall): Mesure de liaison non paramétrique. Fonction de la proportion de couples de rangs concordants. Une valeur de -1 ou +1 ne peut être obtenue que pour les tableaux carrés.
- o**  $\tau_c$  (**tau-c**) **de Kendall**: Mesure de corrélation non paramétrique qui ne tient pas compte des ex aequo. Elle peut atteindre, ou quasiment atteindre, une valeur de +1 ou -1 pour un tableau de quelque dimension qu'il soit.
- o**  $\gamma$  (**gamma**): Mesure d'intensité de la liaison entre deux variables mesurées à un niveau ordinal. On peut dire que c'est la probabilité que deux observations tirées au hasard sont concordantes moins la probabilité qu'elles sont discordantes, en supposant qu'il n'y a pas d'ex aequo.
- o** **Q de Yule** Identique au  $\gamma$  pour tableau 2x2 (rapport des produits croisés). Varie entre -1 et +1.

- **D de Somers:** Mesure d'intensité de la liaison entre deux variables ordinales. C'est une variante asymétrique de gamma qui ne diffère de celui-ci que par l'inclusion du nombre de couples non ex aequo à la variable explicative. Le D de Somers indique l'excédent proportionnel des couples concordants par rapport aux discordants parmi les couples non ex aequo dans la variable explicative. Une version symétrique est également proposée.
- + **r de Pearson:** [Pearson's r, product-Moment correlation] Mesure de liaison *linéaire* entre deux variables. r varie entre -1 (relation négative parfaite; tous les points se situent sur une droite de pente négative) et +1 (relation positive parfaite). 0 indique l'absence de relation linéaire.
- + **Corrélation de Spearman:** La corrélation de Spearman est un coefficient de corrélation de Pearson basée sur les rangs des données. (version non paramétrique). Il convient aux données ordinales ou aux données d'intervalles qui ne satisfont pas à l'hypothèse de normalité
- in  $\eta$  (**Eta**) Mesure de liaison appropriée pour une variable dépendante mesurée sur une échelle intervalles et une variable explicative ayant un nombre limité de modalités. Eta est asymétrique et ne suppose pas l'existence d'une relation linéaire entre les variables.  $\eta^2$  est interprété comme la proportion de variance dans la variable dépendante expliquée par les différences entre les groupes.
- $\kappa$  (**Kappa**) mesure la concordance entre deux indicateurs lorsque les deux servent à évaluer le même objet. La différence entre la proportion observée d'observations sur lesquelles les indicateurs concordent et la proportion attendue du fait du hasard est divisée par la différence maximum envisageable entre les proportions observées et attendues, les marges étant données. Une valeur de 1 indique une concordance parfaite. Une valeur de 0 indique que la concordance ne dépasse pas celle due au hasard.
- **Risque relatif** [relative risk] Le risque relatif mesure l'intensité de la liaison entre la présence d'un facteur et la réalisation d'un événement. L'estimation de la cohorte est le rapport entre les taux d'incidence de l'événement lorsque le facteur est présent et lorsqu'il ne l'est pas. Une valeur de 1 indique qu'il n'y a pas de liaison entre le facteur et l'événement. Dans les études par groupe-témoin, le risque relatif est estimé à partir d'un rapport de probabilités.

#### **Terminologie:**

**Coefficients basés sur les rangs:** Couples de rangs concordants (Si l'observation A a un rang supérieur à l'observation B à l'issue du premier classement, alors A doit également avoir un rang supérieur à B dans le second classement.), discordants (A rang supérieur, B rang inférieur) et ex-aequo [ties](A et B ont le même rang).

**Coefficient symétrique:** Pas de distinction entre variable dépendante et variable indépendante.

**Coefficient asymétrique:** Une variable indépendante est utilisée pour expliquer/prédire une variable dépendante.

#### **Test statistiques (statistiques en option de la procédure CROSSTABS)**

- **DF** [Degrees of freedom] Degrés de liberté
- **Significance** [p-value]: Probabilité d'erreur (de commettre une erreur de type  $\alpha$ ) observé permettant de décider de rejeter ou pas l'hypothèse nulle. Si cette valeur est inférieure au seuil fixé a priori (par exemple 0.05, 0.01, 0.001), l'hypothèse nulle est rejetée.
- **ASE1** erreur-type asymptotique ne supposant pas l'hypothèse nulle [Asymptotic standard error not assuming null hypothesis] Estimation asymptotique de l'erreur-type d'une statistique, ne supposant pas que l'hypothèse nulle est vraie. Cette estimation se rapproche de la valeur de l'erreur-type dans la population au fur et à mesure que la taille de l'échantillon augmente. Elle est appropriée au calcul des intervalles de confiance.
- **T-value:** obtenue en divisant la valeur du coefficient par son erreur-type (ASE1): sert à déterminer la probabilité d'erreur. (Certains textes statistiques parlent de "valeurs" [Z-values])