

## Séance 6 : Statistique d'ordre des variables qualitatives

Une statistique ordinale fonctionne avec des variables qualitatives. La statistique ordinale sert à classer, à préciser un ordre. C'est ce que peut donner sur un sondage une question qui nécessitera de classer ses réponses. Par exemple « Rangez par fréquence d'usage le type de transport que vous sollicitez le plus : marche ; voiture ; transports en commun ; taxi ». Cette question nécessite forcément de trier et ranger ses réponses. A l'intérieur des variables qualitatives, la statistique ordinale s'oppose à la statistique nominale (qui sert à qualifier). En géographie, cela peut supposer une hiérarchie spatiale puisqu'il s'agit de ranger les éléments dans un ordre (souvent croissant) précis. Elle s'illustre par exemple en géographie physique par la classification de Strahler pour les cours d'eau d'un même bassin versant qui consiste à ordonner les différents fleuves et rivières en fonction de leur importance dans l'unité topographique qu'est le bassin versant. En géographie social, la statistique ordinale s'illustre par la loi rang-taille pour les villes, où l'on cherche à ordonner les villes en fonction de leur importance dans le réseau urbain. L'ordre à privilégier reste l'ordre croissant (ou ordre naturel) puisqu'il permet de détecter les valeurs aberrantes, comme celles excessivement petites ou excessivement grandes. La corrélation des rangs cherche à comprendre si les classements opérés sont identiques alors que la concordance de classements s'applique à l'ensemble des classements effectués, c'est une généralisation du coefficient de corrélation des rangs. La corrélation des rangs s'applique une fois qu'un rang est attribué à chaque objet selon deux variables différentes. En d'autres termes c'est répondre à la question « quand un élément est bien classé dans un classement, l'est-il aussi dans l'autre ? ». La concordance de classements adopte le même processus mais à plus petite échelle, en englobant plusieurs classements pour évaluer le nombre d'éléments concordants et discordants. Cette concordance répond plutôt à la question « dans quelle mesure les classements sont-ils cohérents entre eux ? » Le test de Spearman s'applique sur les rangs via le coefficient de corrélation. C'est un test non paramétrique de corrélation des rangs comme le test de Kendall mais ils changent par leur principe, leur interprétation et certaines propriétés statistiques. Spearman est plus sensible aux écarts importants de rang et mesure l'intensité d'une relation entre deux variables alors que Kendall s'appuie sur le nombre d'éléments concordants et discordants pour comparer l'ordre relatif des observations entre deux variables. Spearman peut s'appliquer sur un échantillon plus grand que Kendall. Le coefficient de Goodman-Kruskal utilise la différence entre les paires concordantes et les paires discordantes. Il calcule le « surplus » des paires concordantes vis-à-vis des paires discordantes. Le coefficient d'association de Yule a le même principe que celui de Goodman-Kruskal mais il s'applique pour des matrices 2x2 pour évaluer la fréquence des événements.

---

### Commentaires code Python

Les valeurs obtenues après le calcul des coefficient de Spearman et de Kendall indiquent qu'il y a une certaine relation entre le classement par population et le classement par densité. La corrélation des rangs de Spearman est de 0,0973 et la concordance de rangs de Kendall est de 0,0693 ; deux valeurs proches de zéro certes mais supérieur à zéro ce qui signifie une relation fine entre les deux classements.