

Chapitre 6

Statistique bivariée et géographie

Il est très rare que les variables représentant des facteurs géographiques ne soient pas liées entre elles. La statistique a par conséquent inventé un certain nombre d'outils afin de mesurer de telles liaisons.

6.1 Relation statistique et interprétation

Au sens statistique du terme, la relation entre deux caractères X et Y consiste à confronter des données empiriques, observées ou mesurées sur le terrain, à des données théoriques, produites par un modèle et à déterminer l'importance de l'écart entre données empiriques et théoriques. Plus l'écart de cette relation entre les deux caractères est petit et plus leur relation est forte. Plus cet écart est grand, moins le modèle ajuste bien la relation entre deux caractères, soit parce que le modèle n'est pas adéquat, soit parce que leur relation est faible.

6.2 Démarche générale pour étudier une relation

Il faut commencer par émettre l'hypothèse qu'il existe une relation logique entre deux caractères.

Vérifier et quantifier la relation implique de disposer des données correspondantes. Qu'il soit réalisé informatiquement ou manuellement, le traitement des données permet de quantifier la relation globale entre les deux caractères.

1. Il faut constituer un tableau de données.
2. Si et seulement si les données sont entachées d'erreurs aléatoires, on peut tester la significativité statistique de la relation par un test statistique qui consiste à comparer l'intensité calculée à une intensité fournie par une loi de distribution des erreurs aléatoires. Si l'intensité calculée lui est supérieure,

la relation est très probablement significative, car elle est trop forte pour être seulement due au hasard. Si l'intensité de la relation est suffisamment forte, il reste à l'expliquer et, le plus souvent, à la nuancer.

3. Si l'intensité de la relation est significative, il faut examiner les écarts au modèle.
4. La dernière étape consiste à expliquer la relation.

6.3 Modèle et nature des variables

Il existe trois types de caractères :

1. le caractère de nature quelconque connu par classes (type A) ;
2. le caractère continu quantitatif (ou quasi continu) et le caractère qualitatif ordinal connu par individu (type B) ;
3. la variable catégorielle de nature quelconque, mais possédant peu de modalités connues par individu (type C).

6.3.1 Analyse de deux variables qualitatives : la correspondance statistique

Le modèle du χ^2 est adapté au traitement de deux caractères de nature quelconque, pourvu qu'ils soient connus par classes. Néanmoins, il s'applique essentiellement pour les variables qualitatives, mais elle fonctionne également avec les variables quantitatives. Dans tous les cas, une **correspondance statistique** est établie.

Pour représenter graphiquement deux variables qualitatives, on utilise une **analyse multidimensionnelle**.

6.3.2 Analyse de deux variables quantitatives : la corrélation statistique

Croiser deux variables de type B s'étudie par une corrélation ou une régression statistique. Il existe une extension du modèle aux variables ordinales.

6.3.3 Analyse d'un couple unissant une variable qualitative et une variable quantitative

L'**analyse de variance** est adaptée à la mise en relation entre une variable catégorielle et une variable quantitative, toutes deux connues individu par individu.

6.4 Intensité et « signification » d'une relation entre deux variables

Quelle que soit la nature des caractères X et Y à mettre en relation et le modèle à employer pour ce faire, on calcule l'intensité de la relation entre eux. Si X et Y sont relatifs à une population, et non à un échantillon de celle-ci, on commentera directement l'intensité de la relation entre X et Y , sans procéder à un test de sa significativité.

Dans le cas d'un échantillon, le test de signification à deux variables suit les mêmes règles que celui à une variable, à savoir :

1. comparer la relation calculée à une loi de distribution des erreurs aléatoires d'échantillonnage du modèle utilisé ;
2. poser l'hypothèse nulle H_0 d'indépendance entre X et Y ;
3. choisir le risque d'erreur α du test établissant s'il existe, ou non, une indépendance entre X et Y ; α est l'erreur commise en rejetant H_0 alors qu'elle est vraie ;
4. déterminer ν le(s) nombre(s) de degré de liberté du test, c'est-à-dire le(s) nombre(s) de valeurs de X et de Y qui varient librement) ;
5. lire dans la table de la loi de probabilité du modèle utilisée la valeur plafond en fonction de α et de ν . Si l'intensité calculée de la relation entre X et Y est inférieure à la valeur plafond lue dans la table, on conclura que X et Y sont très probablement indépendants. Si elle lui est supérieure ou égale, on rejettera l'hypothèse H_0 d'indépendance entre X et Y .

La géographie analyse bien souvent, non pas deux variables, mais une multitude. Ainsi, des méthodes d'analyse descriptive généralisées ont été inventées et rendues faciles dans leur usage par l'ordinateur. L'ensemble de cette partie propose une série de chapitres présentant tous les aspects des relations entre deux variables (type A, type B et type C). Il se conclura par un chapitre de transition vers le multivarié en présentant une analyse multidimensionnelle à partir des relations entre deux variables.

Bibliographie