

ÉTUDE DE LA RELATION ENTRE DEUX VARIABLES (le coefficient de corrélation)

C'est une technique qui permet d'étudier la relation qui pourrait exister entre deux variables quantitatives X et Y:

- Corrélation positive, c'est-à-dire à toute augmentation au niveau de X correspond une augmentation au niveau de Y. Les deux variables varient dans le même sens et avec une intensité similaire. Exemple: la taille et le poids
- Corrélation négative, c'est-à-dire à toute augmentation au niveau de X correspond une diminution au niveau de Y. Les deux variables varient dans deux sens opposés et avec une intensité similaire.

Le coefficient de corrélation de Bravais-Pearson

Le coefficient de corrélation de Bravais-Pearson est un indice statistique qui exprime l'intensité et le sens (positif ou négatif) de la relation linéaire entre deux variables quantitatives. C'est une mesure de la liaison linéaire, c'est à dire de la capacité de prédire une variable x par une autre y à l'aide d'un modèle linéaire.

Il permet de mesurer l'intensité de la liaison entre deux caractères quantitatifs. C'est donc un paramètre important dans l'analyse des régressions linéaires (simples ou multiples). En revanche, ce coefficient est nul ($r = 0$) lorsqu'il n'y a pas de relation linéaire entre les variables (ce qui n'exclut pas l'existence d'une relation autre que linéaire). Par ailleurs, le coefficient est de signe positif si la relation est positive (directe, croissante) et de signe négatif si la relation est négative (inverse, décroissante).

Ce coefficient varie entre -1 et +1 ; l'intensité de la relation linéaire sera donc d'autant plus forte que la valeur du coefficient est proche de +1 ou de - 1, et d'autant plus faible qu'elle est proche de 0.

- une valeur proche de +1 montre une forte liaison entre les deux caractères. La relation linéaire est ici croissante (c'est-à-dire que les variables varient dans le même sens);

- une valeur proche de -1 montre également une forte liaison mais la relation linéaire entre les deux caractères est décroissante (les variables varient dans le sens contraire);
- une valeur proche de 0 montre une absence de relation linéaire entre les deux caractères.

RQ : L'existence d'une corrélation élevée entre deux variables x et y ne conduit pas à l'existence d'une relation cause de à effet.

On utilise la connaissance de x pour prédire des valeurs de y ; cela n'implique pas qu'un changement de x cause un changement de y.

Le calcul :

Le coefficient r de Bravais-Pearson entre deux variables X et Y se calcule à partir de la covariance et des écart-types en appliquant la formule suivante :

$$r_{x,y} = \frac{COV_{x,y}}{s_x \times s_y}$$

Avec

$$COV_{x,y} = \frac{\sum_i (x_i - m_x)(y_i - m_y)}{n}$$

Lorsque r est calculé sur un échantillon, on est amené à tester sa significativité. Pour un risque d'erreur alpha fixé, le degré de signification varie en fonction du nombre d'individus de la distribution. Quand ce nombre augmente, le seuil inférieur de signification diminue. La table fournit les valeurs r critiques au-delà desquelles les coefficients de corrélation de Pearson obtenus sont significatifs. Pour des échantillons de taille n, le degré de liberté d.d.l = n-2.