

STATISTIQUES

Résultats à retenir :

1°/ Soit (X, Y) une série statistique double sur un échantillon de taille n et soit $(x_i, y_i)_{1 \leq i \leq n}$ les valeurs numériques prises respectivement par les variables X et Y .

La distribution marginale de la variable X est la distribution des valeurs $(x_i)_{1 \leq i \leq n}$ prises par la variable X .

La distribution marginale de la variable Y est la distribution des valeurs $(y_i)_{1 \leq i \leq n}$ prises par la variable Y .

2°/ Soit X une série statistique sur un échantillon de taille n .

Si $\bar{X}$, $V(X)$ et σ_X désignent respectivement la moyenne, la variance

et l'écart-type de la série, alors $\bar{X} = \frac{1}{n} \sum_{i=1}^p n_i x_i$,

$V(X) = \frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{X})^2$, $\sigma_X = \sqrt{V(X)}$, où les valeurs $x_1, x_2, \dots, x_p$

désignent les valeurs distinctes prise par la variable X si elle est discrète, ou les centres des classes si la variable X est continue.

L'entier n_i désigne l'effectif de la valeur x_i .

3°/ Soit (X, Y) une série statistique double sur un échantillon de taille n

On appelle covariance de (X, Y) le réel, noté $\text{cov}(X, Y)$ défini par

$$\text{cov}(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{X} \bar{Y}, \text{ où } (x_i, y_i) \text{ est la valeur observée pour}$$

l'individu i si X et Y sont discrètes, ou le centre de la classe si l'une des variables est continue.

3°/ soit (X, Y) une série statistique double de taille n .

soit n_{ij} le nombre de fois qu'apparaît le couple (x_i, y_j) .

$$\text{alors } \text{cov}(X, Y) = \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^p n_{ij} x_i y_j - \bar{X} \bar{Y}.$$

4°/ soit (X, Y) une série statistique double de valeurs $(x_i, y_i)_{1 \leq i \leq n}$.

l'ensemble des points M_i de coordonnées (x_i, y_i) dans un repère orthogonal est appelé nuage de points représentant la série statistique.

le point moyen du nuage est le point dont les coordonnées sont les moyennes $\bar{X}$ et $\bar{Y}$.

5°/ Principe de la méthode de Mayer

Soit un nuage de points représentant une série statistique double (X, Y) et G son point moyen.

On scinde le nuage de points de (X, Y) en deux parties contenant à peu près le même nombre de points.

On considère alors les points moyens G_1 et G_2 des deux nuages obtenus.

La droite $(G_1 G_2)$ définit un ajustement affine du nuage de points représentant la série statistique double (X, Y) .

La droite $(G_1 G_2)$ est appelée droite de Mayer et passe par le point moyen G du nuage global.

6°/ Soit (X, Y) une série statistique double sur un échantillon de taille n et telle que $\sigma_X \neq 0$.

Soit (x_i, y_i) $1 \leq i \leq n$ les valeurs observées de la série. Alors la

somme $\sum_{i=1}^n (ax_i + b - y_i)^2$ est minimale pour le couple (a_0, b_0) tel que

$$a_0 = \frac{\text{cov}(X, Y)}{\sigma_X^2} \text{ et } b_0 = \left(\bar{Y} - \frac{\text{cov}(X, Y)}{\sigma_X^2} \bar{X} \right)$$

7°/ Soit (X, Y) une série statistique double sur un échantillon de taille n .

La droite d'équation $y = \frac{\text{cov}(X, Y)}{\sigma_X^2} (x - \bar{X}) + \bar{Y}$ est appelée droite

des moindres carrés de Y en X , ou droite de régression de Y en X .

La droite d'équation $x = \frac{\text{cov}(X, Y)}{\sigma_X^2} (y - \bar{Y}) + \bar{X}$ est appelée droite des

moindres carrées de X en Y , ou droite de régression de X en Y .

8°/ Les droites des moindres carrés de Y en X et de X en Y passent par le point moyen G du nuage associé à la série (X, Y) .

* Soit (X, Y) une série statistique double. On appelle coefficient de corrélation linéaire le réel noté ρ_{XY} défini par $\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$.

* Soit (X, Y) une série statistique double. Alors $-1 \leq \rho_{XY} \leq 1$.

Le coefficient de corrélation linéaire est invariant par changement d'unité ou d'origine.

Les statisticiens conviennent que lorsque $|\rho_{XY}| > \frac{\sqrt{3}}{2}$, l'ajustement affine est justifié et les prédictions faites au moyen de cet ajustement sont raisonnables.