Statistiques

Tableau statistique

Soient X et Y deux variables statistiques sur une population de n individus .On désigne par $x_1, x_2, ..., x_p$ les valeurs prises par X et par $y_1, y_2, ..., y_q$ les valeurs prises par Y.

Y / X	y_1	y_2	…….	y_j	……..	y_q	Distribution marginale en X
x_1	n_{11}	n_{12}		n_{1j}		n_{1q}	n_1
x_2	n_{21}	n_{22}		n_{2j}		n_{2q}	n_2
⋮							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{iq}	n_i
⋮							
x_p	n_{p1}	n_{p2}		n_{pj}		n_{pq}	n_p
Distribution marginale en Y	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.q}$	n

n_i est l'effectif associé à la valeur x_i de X

$n_{.j}$ est l'effectif associé à la valeur y_j de Y

On rappelle :la moyenne de X est $\overline{X} = \frac{1}{n}\sum_{i=1}^{q} n_i x_i$;

la variance de X est $V(X)=\frac{1}{n}\sum_{i=1}^{q} n_i x^2_i - \left(\overline{X}\right)^2$ et l'écart type de X est $\sigma_X = \sqrt{V(X)}$

Covariance

1.Soit une série statistique double (X,Y) définie en données individuelles

On suppose que les valeurs prises par X sont $x_1, x_2, ..., x_n$ et les valeurs prises par Y sont $y_1, y_2, ...y_n$

On appelle covariance de (X,Y) le réel noté $cov(X,Y)=\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{X}\overline{Y}$

2. Dans le cas d'une série statistique(X,Y) définie par un tableau à double entrée

© najah.com

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

Ajustement Affine d'une série statistique

1.Soit (X,Y) une série statistique double et G(X,Y) son point moyen
Si on partage le nuage en deux nuages qui diffèrent d'au plus d'un élément et si G_1 et G_2 sont les points moyens de ces nuages alors (G_1G_2) est la droite de Mayer passant par G.

2. Soit (X,Y) une série statistique double

La droite d'équation :$y=\frac{\text{cov}(X,Y)}{V(X)}x+\left(\overline{Y}-\frac{\text{cov}(X,Y)}{V(X)}\overline{X}\right)$ est appelé droite des moindres carrés ou de régression de Y en X

La droite d'équation :$x=\frac{\text{cov}(X,Y)}{V(Y)}y+\left(\overline{X}-\frac{\text{cov}(X,Y)}{V(Y)}\overline{Y}\right)$ est appelé droite des moindres carrés ou de régression de X en Y.

Coefficient de corrélation linéaire

Soit (X,Y) une série statistique double

Le réel $\rho_{XY}=\frac{\text{cov}(X,Y)}{\sigma_X\sigma_Y}$ est appelé coefficient de corrélation linéaire

Si $|\rho_{XY}|\geq\frac{\sqrt{3}}{2}$ alors l'ajustement affine est justifié

© najah.com®