

Statistiques

Statistiques à deux variables

On appelle **série statistique à deux variables** est une série statistiques dans lesquelles on étudie simultanément deux caractères x et y observés sur une même population.

L'ajustement affine

IMAGE ICI

Méthode	Exemple	Formule
Classer les données par ordre croissant		
Classer les données en deux groupes égaux	1er groupe les 4 premières données, le 2em groupe : les quatres dernières données	
Calculer le point \$G_1\$	$(\bar{x}_1; \bar{y}_1) = (\frac{1+2+3+4}{4} ; \frac{1+3+5+4}{4}) = (2.5 ; 3.25)$	$(\bar{x}; \bar{y}) = (\frac{\sum_{i=1}^n x_i}{n} ; \frac{\sum_{i=1}^n y_i}{n})$
Calculer le point \$G_2\$	$(\bar{x}_2; \bar{y}_2) = (\frac{5+6+7+8}{4} ; \frac{6+1+2+3}{4}) = (6.5 ; 3)$	$(\bar{x}; \bar{y}) = (\frac{\sum_{i=1}^n x_i}{n} ; \frac{\sum_{i=1}^n y_i}{n})$
Calculer la pente de la droite	$m = \frac{3-3.25}{6.5-2.5} = -0.0625$	$m = \frac{y_2-y_1}{x_2-x_1}$
Calculer l'ordonnée à l'origine	$p = 1 - (-0.0625) * 1 = 1.0625$	$p = y_1 - m * x_1$
Droite finale	$y = -0.0625x + 1.0625$	$y = m * x + p$

Méthode des moindres carrés

Méthode courte

La méthode la plus courte pour calculer la droite des moindres carrés, c'est d'utiliser directement la calculatrice (CASIO fx-92B)

- 1. MENU > 2 > 2
- 2. Ensuite il faut intégrer toutes les valeurs \$x\$ et \$y\$ dans le tableau : NE SURTOUT PAS UTILISER LA TOUCHE AC
- 3. OPTN > 4
- 4. La calculatrice donne maintenant les valeurs \$a\$, \$b\$ et \$r\$ qu'il faut remplacer dans la formule suivante :

$Y = ax + b$

\$r\$ correspond au coefficient de régression

Méthode longue

VOIR DANS LE LIVRE A LA PAGE 106.

Calculer et interpreter le coefficient de corrélation

Le coefficient de corrélation correspond au \$r\$ que la calculatrice donne. Il suffit ensuite d'utiliser le barème suivant pour l'interpréter, sans même avoir besoin du graphique.

Qualité de la corrélation	Positive	Négative
Parfaite	$0.98 < r < 1$	$-1 < r < -0.98$
Forte	$0.80 < r < 0.98$	$-0.98 < r < -0.80$
Moyenne	$0.60 < r < 0.80$	$-0.80 < r < -0.60$
Faible	$0.35 < r < 0.60$	$-0.60 < r < -0.35$

Qualité de la corrélation	Positive	Négative
Nulle	$-0.35 < r < 0.35$	$-0.35 < r < 0.35$

Ce coefficient définit à quel point le lien entre la droite de régression et les points est proche. Si il n'y a aucun lien entre la droite et les points, c'est une corrélation nulle.

Différence entre corrélation et causalité

Une **corrélation** est un rapport entre deux valeurs (x et y)

Une **causalité** c'est quand x cause y (ou l'inverse).

Il est très important de comprendre que ces deux mots ne sont pas synonymes, c'est un piège dans lequel de nombreux journalistes tombent. Voici un exemple de corrélation :

Les pays qui ont le plus de prix nobels sont les pays avec la plus grande consommation de chocolat

Cette affirmation est factuelle, c'est en effet le cas, mais ce n'est pas pour autant que manger du chocolat va faire de vous un génie comme beaucoup de journeaux l'ont dit.