

STATISTIQUES

I Nuage de points

- L'ensemble des points M_i de coordonnées (x_i, y_i) est appelé nuage de points associé à la série statistiques à deux variables.
- Le point G de coordonnées (\bar{x}, \bar{y}) , où \bar{x} et \bar{y} , sont les moyennes respectives des x_i et des y_i , est appelé le point moyen du nuage de points associé à la série.

Définition

Exemple

Le tableau suivant présente l'évolution du budget séminaires d'équipes et du chiffre d'affaire d'une société au cours des 6 dernières années :

Budget en milliers d'euros	8	10	12	14	16	18
Chiffre d'affaire en milliers d'euros	40	55	55	70	75	95

- Dans un repère, représenter le nuage de points (x_i, y_i) .
- Déterminer les coordonnées du point moyen G du nuage de points.

II Ajustement affine : interpolation et extrapolation

L'interpolation et l'extrapolation sont des méthodes qui consistent à estimer une valeur ou plusieurs valeurs inconnues dans une série statistique.

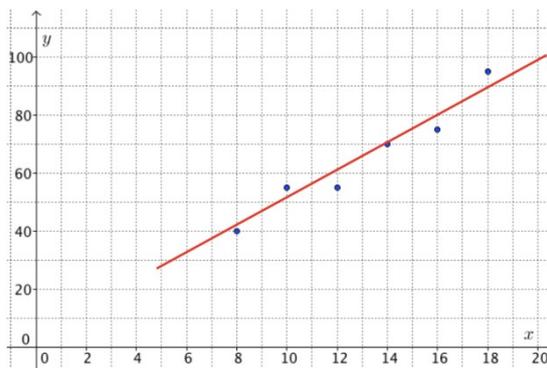
- Pour une interpolation, le calcul est réalisé dans le domaine d'étude fourni par les valeurs de la série.
- Pour une extrapolation, le calcul est réalisé en dehors du domaine d'étude.

Définition

Exemple

- On donne une série exprimant la population d'un pays en fonction des années et on souhaite faire des prévisions pour les années à venir. Les prévisions sortent du domaine d'étude de la série, on parle dans ce cas **d'extrapolation**.
- On donne une série exprimant la température extérieure et la consommation électrique correspondante. Les températures étudiées s'échelonnent entre -10°C et 10°C avec un pas de 4°C .
Sans faire de nouveaux relevés, on souhaite estimer la consommation électrique pour toutes les températures entières comprises entre -10°C et 10°C . Les calculs sont dans le domaine d'étude de la série, on parle dans ce cas **d'interpolation**.

Lorsque les points d'un nuage semblent alignés, on peut construire une droite, appelé droite d'ajustement (ou droite de régression), passant « au plus près » de ces points.



Exemple

La méthode des points moyens :

Cet ajustement consiste à déterminer la droite passant par deux points moyens du nuage de point.

On utilisera les données du premier exemple de ce cours.

- a. On note G_1 , le point moyen associé aux trois premiers points du nuage et G_2 le point moyen associé aux trois derniers points du nuage.
 - (a) Calculer les coordonnées de G_1 et G_2 .
 - (b) Tracer (G_1G_2) , la droite de régression que nous utiliserons pour la suite de l'exercice.
- b. A l'aide du graphique :
 - (a) Estimer le chiffre d'affaire à prévoir pour un budget publicitaire de 22 000 euros.
 - (b) Estimer le budget publicitaire qu'il faudrait prévoir pour obtenir un chiffre d'affaire de 100 000 euros.
 - (c) La méthode utilisée dans les questions 2a et 2b consiste-t-elle en une interpolation ou une extrapolation ?

Exemple

La méthode des moindres carrés :

Cet ajustement consiste à déterminer l'équation de la droite minimisant la somme des carrés des distances entre les points et la droite d'ajustement.

La droite d'ajustement de y en x a pour équation $y = ax + b$, avec :

$$a = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

$$b = \bar{y} - a\bar{x}$$

avec $\text{cov}(x, y) = \frac{1}{n} ((x_1 - \bar{x})(y_1 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y}))$

Et $\text{var}(x) = \frac{1}{n} ((x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)$

Propriété

Soit la série statistique suivante :

x_i	5	10	15	20	25	30	35	40
y_i	13	23	34	44	50	65	75	90

- Dans un repère, représenter le nuage de points (x_i, y_i) .
- (a) À l'aide de la calculatrice, déterminer une équation de la droite d'ajustement par la méthode des moindres carrés.
(b) Représenter la droite d'ajustement de y en x .
- Estimer graphiquement la valeur de x pour $y = 70$. Retrouver ce résultat par calcul. S'agit-il d'une interpolation ou d'une extrapolation ?

Exemple

On considère la série statistique à deux variables données dans le tableau suivant :

x_i	5	10	15	20	25	30	35	40
y_i	13	23	34	44	50	65	75	90

- Dans un repère, représenter le nuage de points.
- Déterminer une équation de la droite d'ajustement par la méthode des moindres carrés.
- Vérifier à l'aide de la calculatrice.
- Représenter la droite d'ajustement de y en x .
- Estimer graphiquement la valeur de x pour $y = 70$. Retrouver ce résultat par calcul. S'agit-il d'une interpolation ou d'une extrapolation ?

Le coefficient de corrélation de x et y est donné par :

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)\text{var}(y)}}$$

Le coefficient de corrélation ρ_{xy} est un nombre compris entre -1 et 1 qui mesure la relation entre les deux variables x et y . Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation linéaire entre les variables est forte.

- Si $\rho_{xy} > 0$ les valeurs prises par y ont tendance à croître quand les valeurs de x augmentent
- Si $\rho_{xy} < 0$ les valeurs prises par y ont tendance à décroître quand les valeurs de x augmentent
- Si $\rho_{xy} = 0$ les variations des variables x et y sont indépendantes.

Exemple

En reprenant les données de la méthode précédente, calculer le coefficient de corrélation et interpréter le résultat.

III Ajustement avec un changement de variable.

Exemple

Soit la série statistique suivante :

x_i	0	5	10	15	20	25	30	35	40	45	50
y_i	19,4	19,4	27,6	40,3	50	59	69	87	132	166	216

- a. Représenter le nuage de point dans un repère.
- b. (a) On effectue le changement de variable $z = \ln(y)$. Réaliser un nouveau tableau présentant les valeurs prises par les variables x et z .
 (b) Représenter un nouveau nuage de points à partir des données des variables x et z .
 (c) A l'aide la calculatrice, déterminer une équation de la droite d'ajustement de z en x par la méthode des moindres carrés. Représenter la droite d'ajustement.
- c. (a) En déduire la relation qui lie y et x puis tracer la courbe représentative de la fonction f définie par $y = f(x)$ dans le repère contenant le premier nuage de points.
 (b) En admettant que le modèle mathématique reste valable en dehors du domaine d'étude, extrapoler le nombre $f(55)$.