Atlasmaths

atlasmaths.com

1ÈRE BAC · SCIENCES EXPÉRIMENTALES · FICHE DE COURS

Statistique (séries à 2 variables)

الإحصاء

📚 Résumé du cours

I. Série statistique double

Définition

Une **série statistique double** est la donnée de n couples (x_i, y_i) représentant 2 caractères mesurés sur une même population.

Nuage de points

On représente les couples (x_i, y_i) dans un repère cartésien. L'ensemble des points est appelé **nuage de points.**

II. Point moyen

Le **point moyen** $G(\bar{x}, \bar{y})$ d'une série double a pour coordonnées :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

III. Variance et écart-type

- **Variance de x :** $V(x) = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \dfrac{1}{n}\sum x_i^2 - \bar{x}^2$
- **Écart-type de x :** $\sigma(x) = \sqrt{V(x)}$
- Idem pour $V(y)$ et $\sigma(y)$.

IV. Covariance

Définition

La **covariance** du couple (x, y) est :

$$\mathrm{Cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$$

V. Ajustement affine par les moindres carrés

Droite de régression de y en x

La meilleure droite ajustant le nuage de points est $y = ax + b$ avec :

$$a = \frac{\mathrm{Cov}(x, y)}{V(x)}, \qquad b = \bar{y} - a\bar{x}$$

Cette droite **passe toujours par le point moyen** $G(\bar{x}, \bar{y})$.

VI. Coefficient de corrélation linéaire

Définition

$$r = \frac{\mathrm{Cov}(x, y)}{\sigma(x) \cdot \sigma(y)}$$

Propriété : $-1 \leq r \leq 1$.

- $r \approx 1$: forte corrélation positive (croissante)
- $r \approx -1$: forte corrélation négative (décroissante)
- $r \approx 0$: pas de corrélation linéaire (le nuage n'est pas alignable)
- En pratique, on considère qu'il y a une bonne corrélation linéaire si $|r| > 0{,}9$.

VII. Méthode BAC type 2024

Énoncé : On a relevé les notes en maths (x) et physique (y) de 5 élèves :

x_i 10 12 14 15 18

y_i 11 13 13 16 17

Déterminer la droite de régression de y en x.

Solution :

$\bar{x} = (10 + 12 + 14 + 15 + 18)/5 = 13{,}8$. $\bar{y} = (11 + 13 + 13 + 16 + 17)/5 = 14$.

Calcul de $\mathrm{Cov}(x, y)$: $\frac{1}{5}(10 \cdot 11 + 12 \cdot 13 + 14 \cdot 13 + 15 \cdot 16 + 18 \cdot 17) - 13{,}8 \cdot 14$
$= \frac{110 + 156 + 182 + 240 + 306}{5} - 193{,}2 = 198{,}8 - 193{,}2 = 5{,}6$

Calcul de $V(x) = \frac{1}{5}(100 + 144 + 196 + 225 + 324) - 13{,}8^2 = 197{,}8 - 190{,}44 = 7{,}36$.

$a = 5{,}6/7{,}36 \approx 0{,}76$.

$b = 14 - 0{,}76 \times 13{,}8 \approx 3{,}5$.
Droite de régression : $\boxed{y \approx 0{,}76x + 3{,}5}$.

VIII. Top 4 pièges à éviter

1. **Confondre variance et écart-type.** $\sigma = \sqrt{V}$.
2. **Diviser par $n - 1$ au lieu de n.** Au lycée marocain, c'est n (variance brute).
3. **Oublier de vérifier la qualité de l'ajustement** avec r avant de conclure.
4. **Confondre a (pente) et b (ordonnée à l'origine)** dans la formule.

🎯 Formules clés

Moyennes : $\bar{x} = \frac{1}{n}\sum x_i$, $\bar{y} = \frac{1}{n}\sum y_i$ **Variance** : $V(x) = \frac{1}{n}\sum x_i^2 - \bar{x}^2$ **Covariance** : $\text{Cov}(x, y) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$ **Droite de régression** : $y = ax + b$ - $a = \text{Cov}(x, y)/V(x)$ - $b = \bar{y} - a\bar{x}$ - Passe par $G(\bar{x}, \bar{y})$ **Coefficient corrélation** : $r = \text{Cov}(x, y)/(\sigma(x)\sigma(y))$, avec $|r| \leq 1$

💡 Astuces & méthodes

- 🎯 Calcule $\bar{x}$ et $\bar{y}$ EN PREMIER. Toutes les autres formules en dépendent.
- 🎯 Vérifie ton résultat : la droite de régression DOIT passer par $G(\bar{x}, \bar{y})$.
- 🎯 Si $|r| < 0{,}9$, l'ajustement linéaire est peu fiable. Mentionne-le dans ta conclusion.