Statistique à deux caractères : tableaux croisés, nuage de points, ajustement affine

Fiche méthode

(Maths 1re (tronc commun) · Information chiffrée · Épreuve anticipée)

La démarche pour une question de statistique

1. Je repère les données : une série (indicateurs, graphique), un tableau croisé (population de référence) ou une série à deux variables $(x_i ; y_i)$.

2. Je choisis le calcul : moyenne, médiane, quartiles ; fréquence marginale ou conditionnelle (« parmi . . . » donne le dénominateur) ; point moyen, droite de Mayer ; équation de la droite d'ajustement.

3. Je calcule à la main, sans arrondir trop tôt.

4. Je contrôle et je conclus : une fréquence est entre 0 et 1, la droite passe par G, et je précise si une estimation est une interpolation ou une extrapolation.

Le réflexe — savoir sur quelle population, ou sur quelle plage de valeurs, porte chaque calcul.

Cas 1 · Moyenne, médiane, quartiles

Réflexe — ranger la série, puis chercher les rangs N/2, N/4 et 3N/4.

Exemple. Exemple : série 7 ; 12 ; 3 ; 9 ; 16 ; 10 ; 4 ; 8. Moyenne, médiane, Q_1 et Q_3 ?

Somme : 69, donc $\overline{x} = \dfrac{69}{8} = 8{,}625$. Série rangée : 3 ; 4 ; 7 ; 8 ; 9 ; 10 ; 12 ; 16. $N = 8$: $Me = \dfrac{8+9}{2} = 8{,}5$. $\dfrac{8}{4} = 2$: $Q_1 = 4$; $\dfrac{24}{4} = 6$: $Q_3 = 10$.

Attention — Pour un quartile, on arrondit le rang à l'entier supérieur ; on ne fait pas de moyenne.

Cas 2 · Fréquences d'un tableau croisé

Réflexe — « parmi les . . . » donne le dénominateur.

Exemple. Exemple : sur 250 clients d'une boulangerie, 150 viennent le matin. 90 clients du matin et 40 clients de l'après-midi achètent une viennoiserie. Fréquence des acheteurs de viennoiserie ? parmi les clients du matin ? Fréquence des clients du matin parmi les acheteurs de viennoiserie ?

Acheteurs : $90 + 40 = 130$, et $\dfrac{130}{250} = 0{,}52$, soit 52 % (fréquence marginale). Parmi les 150 clients du matin : $\dfrac{90}{150} = 0{,}6$, soit 60 %. Parmi les 130 acheteurs : $\dfrac{90}{130} = \dfrac{9}{13} \approx 0{,}692$, soit environ 69,2 %.

Attention — Diviser par l'effectif total donne la fréquence de la case, pas une fréquence conditionnelle.

Cas 3 · Point moyen et droite de Mayer

Réflexe — deux moyennes pour G ; Mayer : ranger, couper en deux, G_1 et G_2.

Exemple. Exemple : points $(1;3)$, $(2;6)$, $(4;8)$ et $(5;13)$. Point moyen et droite de Mayer ?

$\overline{x} = \dfrac{12}{4} = 3$ et $\overline{y} = \dfrac{30}{4} = 7{,}5$: $G(3;7{,}5)$. $G_1(1{,}5;4{,}5)$ et $G_2(4{,}5;10{,}5)$, donc $a = \dfrac{10{,}5 - 4{,}5}{4{,}5 - 1{,}5} = 2$ et $b = 4{,}5 - 2 \times 1{,}5 = 1{,}5$: $y = 2x + 1{,}5$. Contrôle : $2 \times 3 + 1{,}5 = 7{,}5$.

Attention — La droite de Mayer passe par G : si le contrôle échoue, un calcul est faux.

Cas 4 · Estimer avec une droite d'ajustement

Réflexe — remplacer x, ou résoudre ax + b = y, puis dire interpolation ou extrapolation.

Exemple. Exemple : pour des températures de 10 °C à 30 °C, le nombre de bouteilles d'eau vendues par un kiosque est ajusté par $y = 15x + 60$. Ventes estimées à 24 °C ? à 38 °C ? Température pour 480 bouteilles ?

$15 \times 24 + 60 = 420$ bouteilles (interpolation). $15 \times 38 + 60 = 630$ bouteilles (extrapolation, peu sûre).
$15x + 60 = 480$ donne $15x = 420$, soit $x = 28$ °C (interpolation).

Attention — Une extrapolation lointaine n'est qu'un ordre de grandeur, souvent faux.

Les pièges

- Médiane ou quartiles cherchés dans une série non rangée.
- Axe vertical qui ne part pas de 0 : les écarts entre barres paraissent plus grands.
- Population de référence inversée : « parmi les A » et « parmi les B » n'ont pas le même dénominateur.
- Moyenne simple des fréquences de groupes de tailles différentes : il faut pondérer par les effectifs.
- Point moyen confondu avec le milieu du premier et du dernier point.
- Extrapolation sans réserve, ou ordonnée à l'origine interprétée alors que x = 0 est loin des données.
- Causalité tirée d'un nuage allongé : un lien statistique ne prouve pas que x agit sur y.

Ce document est la propriété intellectuelle de Ben, professeur particulier (benprofparticulier.fr), protégée par le Code de la propriété intellectuelle (art. L. 111-1 et L. 122-4). Il est mis à disposition pour un usage personnel de révision. Toute reproduction, diffusion, mise en ligne ou revente, totale ou partielle, sans autorisation écrite est interdite et constitue une contrefaçon (art. L. 335-2).