Lire et résumer des données

Seconde professionnelle · Chapitre 02

Organiser une série statistique

Définition · Comprendre la notion

Une étude statistique observe une population et un caractère. Le caractère est qualitatif s'il décrit une catégorie, quantitatif s'il est numérique. Un caractère quantitatif discret prend des valeurs séparées ; un caractère continu décrit notamment une mesure. Un tableau associe à chaque valeur ou classe son effectif. La somme des effectifs donne la taille totale N de la population. Les classes doivent couvrir les observations sans chevauchement : dans $[10;20[$, la valeur 20 est exclue et appartient à la classe suivante. Choisis une représentation cohérente : barres pour des catégories, bâtons pour des valeurs discrètes, histogramme pour des classes. Pour des classes d'amplitudes différentes, ce sont les aires des rectangles qui doivent représenter les effectifs.

Passer des effectifs aux fréquences

Définition · Comprendre la notion

La fréquence d'une catégorie d'effectif n_i est $f_i = n_i/N$. Les fréquences exactes ont pour somme 1, soit $100\ \%$. Un effectif cumulé croissant additionne les observations jusqu'à une valeur donnée. Une fréquence cumulée croissante rapporte ce cumul à N : elle aide à lire une proportion « au plus ». Un pourcentage arrondi peut faire légèrement varier la somme affichée. Il faut distinguer une variation d'effectif et une variation de proportion lorsque les tailles des populations diffèrent.

Calculer et interpréter une moyenne

Définition · Comprendre la notion

La moyenne d'une série de N valeurs est la somme des valeurs divisée par N. Si la valeur x_i apparaît n_i fois,

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i}.$$

Il s'agit d'une moyenne pondérée : chaque valeur compte autant de fois que son effectif. Pour des classes, le centre de classe fournit une approximation, pas la moyenne exacte des données perdues. La moyenne conserve l'unité du caractère. Elle est sensible aux valeurs extrêmes et n'est pas nécessairement une valeur effectivement observée. Pour fusionner deux groupes, pondère leurs moyennes par leurs effectifs.

Médiane et quartiles

Définition · Comprendre la notion

Commence par ranger les valeurs dans l'ordre croissant. La médiane partage la série en deux moitiés : au moins la moitié des valeurs sont inférieures ou égales à elle, et au moins la moitié supérieures ou égales. Pour un effectif impair, on prend la valeur centrale ; pour un effectif pair, on utilise ici la moyenne des deux valeurs centrales. Dans la convention scolaire française utilisée ici, Q_1 est la valeur au rang $\lceil N/4 \rceil$ et Q_3 au rang $\lceil 3N/4 \rceil$, les rangs commençant à 1. Certains logiciels utilisent une interpolation différente : annonce la convention quand elle change.

Comparer la dispersion de deux séries

> **Définition · Comprendre la notion**
>
> L'étendue est la différence entre maximum et minimum. L'écart interquartile est $Q_3 - Q_1$: il décrit la largeur de la partie centrale de la série. L'écart-type mesure les écarts à la moyenne : $\sigma = \sqrt{\sum n_i (x_i - \bar{x})^2 / N}$. Un outil numérique facilite ce calcul. On utilise ici l'écart-type de la population, avec le diviseur N ; ne le confonds pas avec l'estimation d'échantillon utilisant $N - 1$. Deux séries peuvent avoir la même moyenne mais des dispersions différentes. L'écart-type possède la même unité que les données ; la variance est son carré. Un diagramme en boîte permet de comparer médiane, quartiles et extrêmes.

Les repères essentiels

Organiser une série statistique.

1. Identifie la population, les individus et le caractère étudié.
2. Classe le caractère en qualitatif, quantitatif discret ou quantitatif continu.
3. Range les données dans des catégories ou classes sans chevauchement.
4. Vérifie que les effectifs sont positifs ou nuls et que leur somme retrouve le total.
5. Choisis un graphique adapté et indique clairement les unités et les échelles.

Passer des effectifs aux fréquences.

1. Identifie le total N de la population pertinente.
2. Compte ou additionne les observations qui satisfont exactement la condition.
3. Divise l'effectif obtenu par N, puis multiplie par 100 pour un pourcentage.
4. Compare des fréquences lorsque les populations n'ont pas la même taille.
5. Contrôle les bornes, les compléments et la somme des fréquences avant arrondi.

Calculer et interpréter une moyenne.

1. Calcule le total des valeurs, en multipliant chaque valeur par son effectif.
2. Calcule séparément le nombre total d'observations.
3. Divise les deux totaux ; pour fusionner des groupes, reconstitue d'abord leurs sommes.
4. Conserve l'unité et arrondis seulement le résultat final.
5. Contrôle que la moyenne est entre les extrêmes et qu'elle reconstitue le total avant arrondi.

Médiane et quartiles.

1. Trie les N valeurs dans l'ordre croissant en gardant les répétitions.
2. Pour la médiane, distingue N pair et impair puis repère les positions centrales.
3. Pour les quartiles, calcule $N/4$ et $3N/4$, puis prends les entiers supérieurs.
4. Lis les valeurs aux rangs trouvés ; ne donne pas les rangs comme résultat.
5. Contrôle le partage des observations et précise la convention employée.

Comparer la dispersion de deux séries.

1. Calcule ou relève le centre de chaque série avant de comparer sa dispersion.
2. Pour l'étendue, soustrais le minimum au maximum ; pour l'écart interquartile, calcule $Q_3 - Q_1$.
3. Pour l'écart-type de population, calcule les écarts à la moyenne, leurs carrés, leur moyenne puis sa racine.

4. Vérifie le diviseur N dans l'outil et conserve l'unité des données.
5. Sépare régularité statistique et conformité à des limites précises.

Un exemple pour se repérer

Exemple · Exemple commenté

La série ordonnée est 2, 4, 5, 7, 8, 9, 10, 13. Son effectif est 8. Les deux valeurs centrales sont au rang 4 et au rang 5 : 7 et 8. La médiane est donc $(7+8)/2 = 7{,}5$.

Pour le premier quartile, $N/4 = 2$: on lit la deuxième valeur, donc $Q_1 = 4$. Pour le troisième, $3N/4 = 6$: la sixième valeur est 9, donc $Q_3 = 9$. L'écart interquartile vaut $9 - 4 = 5$.

Quatre valeurs sont inférieures ou égales à 7,5 et quatre supérieures ou égales, ce qui vérifie le partage médian. Les quartiles sont des seuils de position, pas les moyennes de la première et de la dernière moitié.

Les erreurs à éviter

Attention · Comprendre une erreur fréquente

Une valeur et son effectif répondent à deux questions différentes : « quelle mesure ? » et « combien d'observations ? ». Écris les titres des lignes ou colonnes avant de calculer.