

Statistiques à une et deux variables

SECONDE

1 Séries statistiques à une variable

Définition

Une étude statistique porte sur une **population**, qui contient des **individus**. On étudie dans la population un **caractère**, soit **quantitatif** (valeurs numériques), soit **qualitatif**.

À retenir

Pour un caractère quantitatif, on peut regrouper les valeurs dans des **classes** (intervalles).

Méthode

Exemple : Dans une étude sur la taille des élèves d'un lycée, la population est l'ensemble des élèves, chaque élève est un individu, et le caractère étudié (la taille, en cm) est **quantitatif**. On peut regrouper les tailles en classes, par exemple $[150 ; 160[$, $[160 ; 170[$, etc.

2 Fréquences, fréquences cumulées

Définition

Les valeurs d'un caractère quantitatif étudié sont rangées ci-dessous dans l'ordre croissant.

Valeur	x_1	x_2	x_3	$\dots$	x_k
Effectif	n_1	n_2	n_3	$\dots$	n_k
E.C.C.	n_1	$n_1 + n_2$	$n_1 + n_2 + n_3$	$\dots$	N
Fréquence	f_1	f_2	f_3	$\dots$	f_k
F.C.C.	f_1	$f_1 + f_2$	$f_1 + f_2 + f_3$	$\dots$	1

À retenir

L'**effectif total** est $N = n_1 + n_2 + \dots + n_k$. La **fréquence** de la valeur x_i est $f_i = \frac{n_i}{N}$.

L'**effectif cumulé croissant** (E.C.C., resp. la **fréquence cumulée croissante**, F.C.C.) de la valeur x_i est la somme des effectifs (resp. des fréquences) des valeurs inférieures ou égales à x_i dans la série.

3 Moyenne pondérée

Propriété

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{N} \quad \text{ou} \quad \bar{x} = f_1x_1 + f_2x_2 + \dots + f_kx_k$$

À retenir

Si les données sont regroupées en **classes**, on prend pour valeurs les **centres des classes**.
De façon générale, lorsqu'on transforme toutes les valeurs d'une série statistique de moyenne $\bar{x}$ par une fonction affine $x \mapsto mx + p$, la moyenne de la nouvelle série est $m\bar{x} + p$.

Méthode

Exemple : Une série a pour valeurs 2, 5 et 9 avec les effectifs respectifs 3, 4 et 3 ($N = 10$).

$$\bar{x} = \frac{3 \times 2 + 4 \times 5 + 3 \times 9}{10} = \frac{6 + 20 + 27}{10} = \frac{53}{10} = 5,3$$

4 Médiane

Définition

La **médiane** sépare une série ordonnée en deux sous-séries de même effectif.

Méthode

Exemple : Pour la série ordonnée 2 ; 3 ; 5 ; 7 ; 9 (effectif impair), la médiane est la valeur centrale : $Me = 5$.

Exemple : Pour la série ordonnée 2 ; 4 ; 6 ; 8 (effectif pair), la médiane est la moyenne des deux valeurs centrales : $Me = \frac{4 + 6}{2} = 5$.

5 Quartiles

Définition

Le 1^{er} (resp. 3^e) quartile, noté Q_1 (resp. Q_3), est la **plus petite valeur** de la série telle qu'au moins 25 % (resp. 75 %) des valeurs lui soient inférieures ou égales.

Méthode

Exemple : Pour une série de 20 valeurs ordonnées, 25 % de 20 vaut 5 : Q_1 est donc la 5^e valeur de la série. De même, 75 % de 20 vaut 15 : Q_3 est la 15^e valeur.

6 Écart interquartile

Définition

$$\text{Écart interquartile} = Q_3 - Q_1$$

À retenir

Plus $Q_3 - Q_1$ est grand, plus les valeurs de la série sont **dispersées autour de la médiane**.

7 Étendue et écart type

Définition

Étendue : Valeur maximale – Valeur minimale.

Propriété

$$\sigma = \sqrt{\frac{n_1(x_1 - \bar{x})^2 + \dots + n_k(x_k - \bar{x})^2}{N}} \quad \text{ou} \quad \sigma = \sqrt{\frac{f_1(x_1 - \bar{x})^2 + \dots + f_k(x_k - \bar{x})^2}{N}}$$

À retenir

Plus σ est grand, plus les valeurs de la série sont **dispersées autour de la moyenne**.

Méthode

Exemple : Reprenons la série 2 ; 5 ; 9 d'effectifs 3 ; 4 ; 3, de moyenne $\bar{x} = 5,3$.

$$\sigma = \sqrt{\frac{3(2 - 5,3)^2 + 4(5 - 5,3)^2 + 3(9 - 5,3)^2}{10}} = \sqrt{\frac{32,67 + 0,36 + 41,07}{10}} \approx \sqrt{7,41} \approx 2,72$$

8 Séries statistiques à deux variables qualitatives

Définition

Dans une population, on étudie **deux caractères qualitatifs**, présentés dans un **tableau croisé** (effectifs à double entrée).

Méthode

Exemple : On a répertorié les animaux d'un refuge selon leur espèce et l'état de leur stérilisation. La population est l'ensemble des animaux ; les caractères étudiés sont l'**espèce** et l'**état de stérilisation**, tous deux qualitatifs.

	Chien	Chat	Total
Stérilisé	12	8	20
Non stérilisé	18	42	60
Total	30	50	80

À retenir

Fréquence : proportion d'individus vérifiant deux critères à la fois (une case du tableau).

Fréquence marginale : proportion d'individus vérifiant un seul critère (un total de ligne ou de colonne).

Fréquence conditionnelle : proportion d'individus vérifiant un critère, **parmi** ceux qui vérifient déjà un autre critère.

Méthode

Exemple : La **fréquence** des chats stérilisés dans la population est $\frac{8}{80} = 0,1 = 10\%$.

Exemple : La **fréquence marginale** des animaux stérilisés est $\frac{20}{80} = \frac{1}{4} = 25\%$.

Exemple : La **fréquence marginale** des chats est $\frac{50}{80} = 0,625 = 62,5\%$.

Exemple : La **fréquence conditionnelle** des chiens parmi les animaux stérilisés est $\frac{12}{20} = \frac{60}{100} = 60\%$.