

Objectifs : Caractéristiques de position et de dispersion : moyenne, médiane, quartiles.

Passage effectif / fréquence. Calculer des effectifs ou des fréquences cumulées.

Représenter une série statistique : nuage de points, histogramme, courbe des fréquences cumulées.

1) Définition : L'ensemble sur lequel on travaille en statistique est appelé **population**.

Si cet ensemble est trop vaste, on en restreint l'étude à une partie appelée **échantillon**.

Un élément de cet ensemble est appelé **individu**.

Définition : La particularité commune que l'on étudie est appelée **caractère**.

L'effectif d'une « valeur » d'un caractère est le nombre d'individus ayant cette valeur.

Les valeurs prises par le caractère peuvent être de 2 types

(1) si celles-ci s'expriment par un nombre, il s'agit d'un **caractère quantitatif** (dans ce cas, le nombre se note en général x_i)

Si les valeurs du nombre exprimé en (1) sont isolées, il s'agit d'un **caractère discret**.

Par contre, si ces valeurs sont prises dans tout un intervalle de \mathbb{R} , il s'agit d'un **caractère continu**.

(2) si la particularité étudiée ne s'exprime pas par un nombre, il s'agit d'un **caractère qualitatif**.

Exemple :

Population	Elèves d'un lycée		
Individu	Un élève		
Caractère	Couleur des yeux	taille	Nombre de frères et sœurs
valeurs du caractère	Bleu, vert, marron	On regroupe les valeurs en classes (intervalles) [1,60;1,65[; [1,65;1,70[0, 1, 2, 3...
	Caractère qualitatif	Caractère quantitatif continu	Caractère quantitatif discret

Définition : Une série statistique est l'ensemble des résultats d'une étude : valeurs du caractère et effectifs correspondants. On représente souvent une série statistique sous forme d'un tableau.

2) Fréquence

Définition : Le nombre d'individus (n_i) d'un caractère d'une série statistique est appelé effectif.

Le nombre total d'individus (N) de la population est appelé effectif total.

Le rapport $f_i = \frac{n_i}{N}$ est appelé **fréquence**.

Propriétés : f_i est un nombre toujours compris entre 0 et 1. Souvent, les nombres f_i s'expriment par un pourcentage. La somme des nombres f_i est toujours égale à 1.

Définitions : On note x_i , une valeur prise par un caractère quantitatif.

L'effectif cumulés croissant (respectivement décroissant) de x_i est la somme des effectifs des valeurs inférieures (respectivement supérieures) ou égales à x_i .

La fréquence cumulée croissante (respectivement décroissante) de x_i est la somme des fréquences des valeurs inférieures (respectivement supérieures) ou égales à x_i .

Exercice 1 : Pour 121 portées de souris blanches, on a dénombré les petits souriceaux :

Nombre de petits	1	2	3	4	5	6	7	8	9
Effectif	7	11	16	17	26	31	11	1	1

a) Quelle est la population étudiée ? Quel est le caractère étudié ?

b) Calculer, à 10^{-2} près, les fréquences puis les fréquences cumulées croissantes.

3) Représentations graphiques possibles d'une série statistique :

- diagramme en bâtons** : Dans un repère orthogonal (O ; I, J), dont vous préciserez l'origine et les unités sur chaque axe, placer les valeurs du caractère x_i en abscisses et les effectifs n_i ou les fréquences f_i en ordonnées. Tracer des segments dont la longueur est proportionnelle à l'effectif, ou la fréquence.
- diagramme circulaire** : Compléter les données statistiques par les fréquences ; Puis en multipliant ces fréquences par 360, vous pourrez tracer des secteurs angulaires dont l'angle est proportionnel à l'effectif de la valeur.
- histogramme** : Dans le cas d'un histogramme, les effectifs (ou les fréquences) et les aires des rectangles sont proportionnels. Si tous les rectangles ont la même largeur, alors la longueur est proportionnelle à l'effectif.

Exercice 2 : Effectuer un diagramme en bâtons des données de l'exercice précédent, puis une courbe des fréquences cumulées croissantes.

Exercice 3 : En 2009, d'après une étude réalisée par le ministère de la Culture et de la Communication, la répartition des visiteurs dans les musées par groupe sociaux a été la suivante :

Etudiants	Classe populaire	Classe moyenne inférieure	Classe moyenne supérieure	Classe supérieure
9%	32%	16%	22%	21%

Construire un diagramme à secteurs représentant cette répartition.

4) Analyse de données

a) Mesures de position

On se donne une série statistique :

Valeur	x_1	x_2	...	x_p
Effectif	n_1	n_2	...	n_p
Fréquences	f_1	f_2	...	f_p

La moyenne

Définition : La **moyenne** de cette série est le nombre réel, noté \bar{x} , tel que :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_p x_p}{N} \quad \text{où } N \text{ est l'effectif total ; } N = n_1 + n_2 + \dots + n_p.$$

On note souvent $N = \sum_{i=1}^p n_i$ (somme des n_i de $i = 1$ à p) ; $\bar{x} = \frac{1}{N} \sum_{i=1}^p n_i x_i = \sum_{i=1}^p f_i x_i$

Remarque : si les effectifs représentent des coefficients, on l'appelle moyenne pondérée

La médiane

Définition : Pour une série ordonnée, la **médiane** d'une série statistique est la valeur du caractère qui partage cette série en deux groupes de même effectif.

Méthode : Si la série contient n valeurs rangées dans l'ordre croissant :

- si n est impair, on prend la $\frac{n+1}{2}$ ème valeur pour médiane.

- si n est pair, on prend pour médiane la moyenne entre la $\frac{n}{2}$ ème et la $\frac{n}{2} + 1$ ème valeur.

Exemples : Avec un effectif total **impair** : $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5$, la médiane est x_3 .

Avec un effectif total **pair** : $x_1 \leq x_2 \leq x_3 \leq x_4 \leq x_5 \leq x_6$, la médiane est $\frac{x_3 + x_4}{2}$

Mode

On appelle **mode** d'une série statistique une valeur du caractère qui correspond au plus grand effectif.

Remarque: il peut y avoir plusieurs modes dans une même série statistique

Les quartiles

Définitions : La liste des N données est rangée par ordre croissant.

Le premier quartile Q_1 est la plus petite donnée de la liste telle qu'au moins 25% des données soient inférieures ou égales à Q_1 .

Le troisième quartile Q_3 est la plus petite donnée de la liste telle qu'au moins 75% des données soient inférieures ou égales à Q_3 .

Méthode : Pour Q_1 , on calcule $N/4$, puis on détermine le premier entier p supérieur ou égal à $N/4$. Cet entier p est le rang de Q_1 . Pour Q_3 , on fait de même avec $3N/4$

Exemple : Pour $N=15$, on a $N/4=3,75$ et $3N/4 = 11,25$. Donc Q_1 est la quatrième valeur de la série et Q_3 est la douzième valeur.

b) Mesures de dispersion

On appelle **étendue** d'une série statistique la différence entre la plus grande valeur du caractère et la plus petite.

Exemple : considérons les séries statistiques suivantes:

Valeur	2	2,1	2,3	2,5	2,8	3
Effectif	5	7	8	10	8	2

L'étendue est : $3-2 = 1$

Écart interquartile : c'est la différence Q_3-Q_1

Exercice 4 : Trois concurrents au tir à l'arc ont obtenu les scores suivants (notés sur 20) :

Alizée	2	18	19	10	11
Benoit	12	12	11	12	13
Maéva	14	2	16	10	14

Pour chaque tireur, calculer le score moyen, le score médian et l'étendue. Commenter ces résultats.

Exercice 5 : Ci-dessous la répartition des salaires mensuels des employés d'une petite entreprise :

198 000 ; 198 000 ; 144 000 ; 252 000 ; 420 000 ; 198 000 ; 144 000 ; 252 000 ; 288 000 ; 252 000 ; 198 000 ; 252 000 ; 198 000 ; 288 000 ; 252 000 ; 198 000 ; 288 000 ; 288 000 ; 420 000 ; 198 000 et 144 000.

- Construire un tableau donnant les salaires, les effectifs et les effectifs cumulés croissants.
- Déterminer le mode de cette série et le salaire moyen d'un employé.
- Déterminer la médiane de cette série. Quelle est la signification de ce nombre ?
- Déterminer le premier quartile de cette série, puis l'interpréter.