

STATISTIQUES

La statistique a pour objet de recueillir des observations portant sur des sujets présentant une certaine propriété et de traduire ces observations par des nombres qui permettent d'avoir des renseignements sur cette propriété. Le but de la statistique descriptive est de structurer et de représenter l'information contenue dans les données

1) Vocabulaire :

La **population** est l'ensemble des sujets étudiée
 Individu : c'est un élément de la population. Effectif total : c'est le nombre total d'individus. Caractère : c'est la propriété étudiée. On distingue les caractères discrets qui ne peuvent prendre qu'un nombre fini de valeurs (notes à un devoir...) et les caractères continus dont on regroupe les valeurs par intervalles (taille, durée d'écoute...)

2) Séries statistiques associées à un caractère discret :

On appelle série statistique la donnée simultanée (dans un tableau) des valeurs du caractère étudié (noté x_i), rangées dans l'ordre croissant, et des effectifs (notés n_i) de ces valeurs. et à la place des effectifs (n_i), on peut aussi utiliser les fréquences : $f_i = \frac{n_i}{N}$ où N représente

l'effectif total) ou les fréquences en pourcentages

$$P_i = \frac{n_i}{N} \times 100$$

L'effectif cumulé croissant d'une valeur x est la somme des effectifs des valeurs y tels que $x \leq y$. L'effectif cumulé décroissant d'une valeur x est la somme des effectifs des valeurs y tels que $y \geq x$

Exemple 1 : Les notes sur 20 obtenues lors d'un devoir de mathématiques dans une classe de seconde sont les suivantes : 10, 8, 11, 9, 12, 10, 8, 10, 7, 9, 10, 11, 12, 10, 8, 9, 10, 9, 10, 11. • La population étudiée est la classe et les individus sont les élèves. L'effectif total est égal à 20 et la note obtenue au devoir est le caractère discret que l'on étudie.

La série statistique définie par les effectifs est la suivante :

Valeurs x_i	7	8	9	10	11	12
Effectif	1	3	4	7	3	2
Effectif cumulé croissant	1	4*	8	15	18	20
Effectif cumulé décroissant	19	10	12	5	2	0

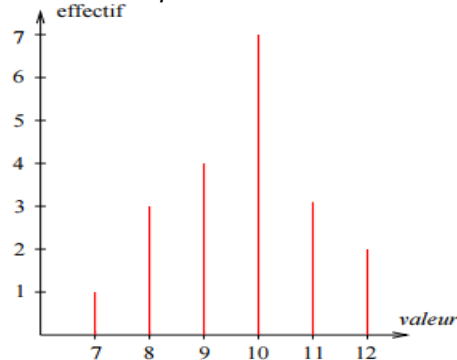
La série statistique définie par les fréquences en pourcentage est la suivante :

Valeurs du caractère (notes)	7	8	9	10	11	12
Fréquences	5%	15%	20%	35%	15%	10%

3) Représentation graphique

Pour les caractères quantitatifs discrets, on utilise le diagramme en bâton : dans un repère orthogonal, pour chaque valeur de la série statistique on trace un trait vertical dont la hauteur est proportionnelle

Avec l'exemple des notes :



4) les Paramètres de position d'une série statistique (le mode ; la Moyenne ; la Médiane)

le mode : c'est la valeur du caractère correspondant au plus fort effectif. Dans l'exemple 1 c'est : la note : 10

Médiane : on appelle médiane d'une série statistique discrète toute valeur M du caractère telle qu'au moins 50% des individus aient une valeur du caractère inférieure ou égale à M et au moins 50% des individus aient une valeur du caractère supérieure ou égale à M .

Recherche pratique de la médiane :

On range les valeurs du caractère une par une dans l'ordre croissant (chaque valeur du caractère doit apparaître un nombre de fois égal à l'effectif correspondant).

Si l'effectif total est impair, la médiane M est la valeur du caractère située au milieu.

Si l'effectif total est pair, la médiane M est la demi-somme des 2 valeurs situées au milieu.

Dans l'exemple 1 : méthode 1 :

On écrit les valeurs de la série dans l'ordre croissant :

7 8 8 8 9 9 9 9 10 10 **10** 10 10 10 10 11 11 11 12 12

L'effectif total est pair : la médiane M est la demi-somme des 2 valeurs situées au milieu d'où la médiane est égale à 10

Méthode 2 : le demie L'effectif total est : $\frac{20}{2} = 10$

Le plus petit effectif cumulé supérieur à 10 est 15

La note associée est 10 donc la médiane est 10

La moyenne : est égale à :

$$m = \frac{7 \times 1 + 8 \times 3 + 9 \times 4 + 10 \times 7 + 11 \times 3 + 12 \times 2}{N} = \frac{194}{20} = 9.7$$

5) les paramètres de dispersions (L'écart-moyen ; la Variance ; L'écart-type)

Exemple2 : on considère la série statistique suivante :

7	2	1	caractères
1	4	5	Effectifs

Calculer les Paramètres de dispersions de cette série statistique (L'écart-moyen ; la Variance ; L'écart-type)

Solution :

On calcul d'abord la moyenne :

$$m = \frac{5 \times 1 + 4 \times 2 + 1 \times 7}{10} = \frac{20}{10} = 2$$

L'écart-moyen : e

$$e = \frac{5 \times |1-2| + 4 \times |2-2| + 1 \times |7-2|}{10} = \frac{5 \times |-1| + 4 \times |0| + 1 \times |5|}{10}$$

$$e = \frac{5 \times 1 + 4 \times 0 + 1 \times 5}{10} = \frac{10}{10} = 1$$

la Variance : V

$$V = \frac{5 \times |1-2|^2 + 4 \times |2-2|^2 + 1 \times |7-2|^2}{10} = \frac{5 \times |-1|^2 + 4 \times |0|^2 + 1 \times |5|^2}{10}$$

6) Séries statistiques associées à un caractère continu

6-1 Classement des données

La seule différence par rapport aux caractères discrets, c'est que les valeurs du caractère sont regroupées dans des intervalles (Appelés classes du caractère).

Exemple3 : Temps passé devant la télévision par 34 élèves pendant une certaine journée.

temps en minutes	[0, 15[[15, 30[[30, 60[[60, 120[[120, 180[
nombre d'élèves	7	5	8	10	4

6-2 Représentation graphique

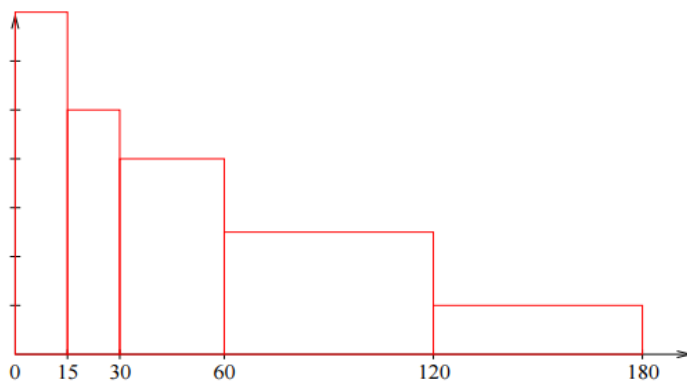
Pour la représentation graphique d'un caractère continu, on utilise généralement un histogramme : dans un repère orthogonal on porte en abscisse les valeurs des bornes des intervalles (selon l'unité choisie), puis pour chaque intervalle on trace un rectangle dont l'aire est proportionnelle à l'effectif (selon l'unité choisie).

I Remarque : En pratique, il est conseillé de commencer par construire un tableau donnant la largeur et l'aire de chaque rectangle (selon les unités choisies). On peut alors facilement en déduire la hauteur de chaque rectangle ce qui facilite la construction graphique de l'histogramme.

I Pour l'exemple3 : proposé ci-dessus :

(Unités : en abscisse 1 cm représente 15 min et 1 cm² Représente 1 élève)

temps en minutes	[0, 15[[15, 30[[30, 60[[60, 120[[120, 180[
aire du rectangle en cm ²	7	5	8	10	4
largeur du rectangle en cm	1	1	2	4	4
hauteur du rectangle en cm = $\frac{\text{aire}}{\text{largeur}}$	7	5	4	2,5	1



6-3 Calcul des paramètres de position et de dispersion

Pour calculer les différents paramètres d'une série statistique associée à un caractère continu, on prend comme valeur du caractère le milieu de chaque classe.

Pour l'exemple3, la série devient :

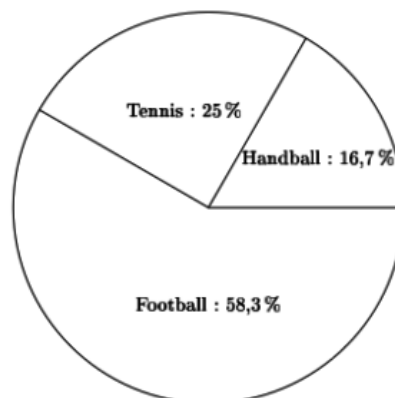
valeur (milieu de chaque intervalle) x_i	7,5	22,5	45	90	150
effectif n_i	7	5	8	10	4

On en déduit que :

$$\bar{x} = \frac{7 \times 7,5 + 5 \times 22,5 + 8 \times 45 + 10 \times 90 + 4 \times 150}{34} \approx 60$$

7) Autre Représentation graphique : diagramme circulaire

Exemple4 : voici un diagramme circulaire représentant la répartition des adhérents à un club sportif.



Sachant que le club compte 240 adhérents, combien d'adhérents jouent ...

- Au football ?
- Au tennis ?
- Au handball ?

Solution : On multiplie l'effectif total (240) par la fréquence de chaque caractère indiquée dans le camembert pour obtenir l'effectif du caractère. Ainsi :

- Football : $240 \times 0,583 = 140$
- Tennis : $240 \times 0,25 = 60$
- Handball : $240 \times 0,167 = 40$

« C'est en forgeant que l'on devient forgeron » Dit un proverbe.
C'est en s'entraînant régulièrement aux calculs et exercices

Que l'on devient un mathématicien

