Statistique descriptive

Effectuer une étude statistique consiste à recueillir, présenter et exploiter des informations sur un caractère d'une population.

A. Effectifs et fréquences

Les 30 élèves d'une classe ont été répartis dans 5 groupes notés A, B, C, D et E.
Nous désirons effectuer une étude statistique de cette répartition.
La population étudiée est l'ensemble des 30 élèves. Le caractère étudié est le groupe auquel appartient chaque élève.

1- Recueil des données

Les données se présentent d'abord sous la forme d'une suite 30 lettres A, B, C, D ou E :
ACCDCEBBACCDEDDCBCCACDEBBACCDB.
Sous cette forme, elles sont difficilement exploitables.

2- Tableau des effectifs

Pour mieux comprendre la répartition des élèves dans les 5 groupes, nous allons créer un tableau des effectifs, c'est à dire un tableau indiquant le nombre d'élèves de chaque groupe.

Groupe	A	B	C	D	E
Effectif	4	6	11	6	3

Ce tableau nous permet de faire un certain nombre de constatations comme :
- le groupe contenant le plus d'élèves est le groupe C.
- le groupe contenant le moins d'élèves est le groupe E.
- …..

3- Tableau des fréquences

Pour mettre en évidence l'importance de chaque groupe par rapport à l'ensemble de la classe, nous faisons intervenir la notion de fréquence.
La fréquence associée à un groupe est la fraction que représente ce groupe dans la classe; ceci se traduit par :

$$\text{fréquence} = \frac{\text{effectif}}{\text{effectif total}}$$

La fréquence est donc un nombre compris entre 0 et 1. En multipliant ce nombre par 100 on obtient la fréquence en pourcentage.

Par exemple, le groupe B contient 6 élèves; sa fréquence est donc $\frac{6}{30}$ = 0,2; sa fréquence en pourcentage est $0{,}2 \times 100 = 20$; il y a 20% des élèves dans le groupe B.
En appliquant ceci à chaque groupe, on obtient le tableau des fréquences.

Groupe	A	B	C	D	E	Total
Effectif	4	6	11	6	3	30
Fréquence	0,13	0,2	0,37	0,2	0,1	1
Fréquence en %	13	20	37	20	10	100

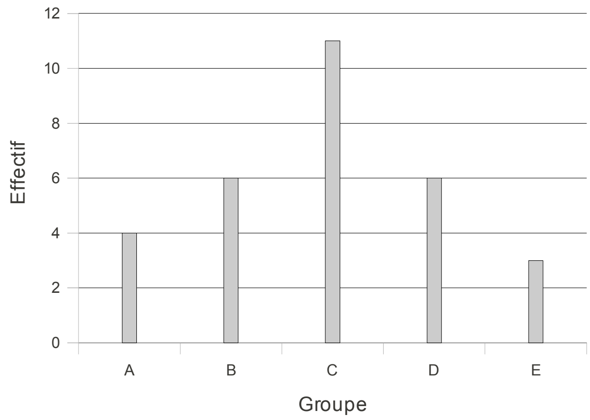
Remarques :
- Le total des fréquences est toujours égal à 1 et le total des fréquences en pourcentage est toujours égal à 100%.

- Les lignes donnant l'effectif, la fréquence et la fréquence en % sont proportionnelles.

4- Diagrammes en bâtons

Les diagrammes en bâtons donnent pour chaque valeur du caractère étudié un bâton dont la longueur est proportionnelle à l'effectif ou à la fréquence.

Exemple



Les diagrammes en bâtons permettent de comparer simplement les effectifs des différents groupes.

B. Paramètres statistiques

1- Moyenne

Considérons une série de notes. Comment trouver une série de même somme où toutes les notes seraient égales ?
On calcule la moyenne pour répondre à cette question.

a) Calcul de la moyenne

Considérons la série statistique suivante :

Valeurs	x_1	x_2	.....	x_p
Effectifs	n_1	n_2	.....	n_p

La somme des valeurs est $S = n_1x_1 + n_2x_2 + ... + n_px_p = \sum n_i x_i$.
L'effectif total est $n = n_1 + n_2 + ... + n_p = \sum n_i$
La moyenne de la série est $m = \dfrac{S}{n} = \dfrac{\sum n_i x_i}{\sum n_i}$.
Si on connait les fréquences f_i, alors $m = \sum f_i x_i$.

Exemple
Le tableau suivant donne la répartition des salaires dans une entreprise.

Salaire	900	1000	1200	1500	3000
Effectif	15	12	5	3	1

Le salaire moyen est :

$$S = \frac{15 \times 900 + 12 \times 1000 + 5 \times 1200 + 3 \times 1500 + 1 \times 3000}{15+12+5+3+1} = \frac{39000}{36} \simeq 1083{,}33 \text{ euros}$$

b) Propriétés de la moyenne

- Si on ajoute un même nombre b à toutes les valeurs d'une série statistique, la moyenne est augmentée du même nombre b.
- Si on multiplie par un même nombre a toutes les valeurs d'une série statistique, la moyenne est multipliée par le même nombre a.
- Si f est la fonction affine définie par f(x) = ax + b et si on remplace toutes les valeurs x_i d'une série statistique par $f(x_i)$, la moyenne m est transformée en f(m) = am + b.

c) Utilisation de moyennes partielles

Considérons deux séries statistiques d'effectifs n_1 et n_2 et de moyennes m_1 et m_2.

La moyenne de la série regroupant ces deux séries est

$$m = \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2}$$

On calcule une moyenne pondérée, c'est à dire une moyenne où chaque valeur est affectée d'un coefficient.

Exemple

La moyenne d'une classe de 21 élèves est 9,9. La moyenne d'une autre classe de 33 élèves est 10,7. Quelle est la moyenne pour l'ensemble des deux classes ?

$$m = \frac{21 \times 9{,}9 + 33 \times 10{,}7}{21 + 33} = \frac{561}{54} \approx 10{,}39$$

2- Médiane

La médiane Me d'une série partage la population en deux parties de sorte que :

- au moins 50% des individus correspondent à une valeur inférieure ou égale à la médiane.

- au moins 50% des individus correspondent à une valeur supérieure ou égale à la médiane.

a) Utilisation des effectifs

La médiane d'une série correspond à la valeur centrale lorsque les données ont été rangées dans l'ordre croissant.

Le tableau suivant donne les notes des 25 élèves d'une classe.

Notes	6	7	8	9	10	11	12	13	14	15	16	17
Effectifs	2	0	1	3	5	3	4	0	2	2	2	1

Ceci correspond à la série :

6-6-8-9-9-9-10-10-10-10-10-11-**11**-11-12-12-12-12-14-14-15-15-16-16-17

Comme il y a 25 élèves, la valeur centrale est la $13^{ème}$ note, soit 11.

La note médiane est donc Me = 11.

Remarque

Si l'effectif total était pair, on aurait deux valeurs centrales, la médiane serait alors la moyenne de ces deux valeurs.

b) Utilisation des effectifs cumulés

On peut aussi déterminer la médiane en cherchant les effectifs cumulés de la série, c'est à dire le nombre d'individus ayant une note inférieure ou égale à une note donnée.

Cela nous donne le tableau suivant :

Notes	6	7	8	9	10	11	12	13	14	15	16	17
Effectifs	2	0	1	3	5	3	4	0	2	2	2	1
Eff. cumulés	2	2	3	6	11	14	18	18	20	22	24	25

On constate que 14 élèves (donc au moins 50% des élèves) ont une note inférieure ou égale à 11.
D'autre part 11 élèves ont moins de 11, cela signifie que 14 élèves (donc au moins 50% des élèves) ont une note supérieure ou égale à 11.
La médiane est donc bien 11.

3- Quartiles

Lorsque les données sont rangées dans l'ordre croissant :

> Le 1er quartile est la plus petite valeur Q_1 telle qu'au moins 1/4 (ou 25%) des données de la liste sont inférieures à Q_1.
>
> Le 3ème quartile est la plus petite valeur Q_3 telle qu'au moins 3/4 (ou 75%) des données de la liste sont inférieures à Q_3.

Exemple
En reprenant les données du paragraphe précédent, on obtient :
1/4 de l'effectif total donne 25/4 = 6,25 ; le 1er quartile Q_1 est donc la 7ème note, c'est à dire la première note qui a un effectif cumulé supérieur ou égal à 7. Ainsi $Q_1 = 10$.
3/4 de l'effectif total donne $3\times25/4 = 18{,}75$; le 3ème quartile Q_3 est donc la 19ème note, c'est à dire la première note qui a un effectif cumulé supérieur ou égal à 19. Ainsi $Q_3 = 14$.

4- Paramètres de dispersion

Pour estimer la dispersion des valeurs d'une série statistique on peut utiliser son étendue ou son écart interquartile.

> L'étendue d'une série statistique est la différence entre la plus grande et la plus petite des valeurs du caractère étudié.
>
> L'écart interquartile d'une série statistique est la différence entre le 1er et le 3ème quartile.

C. Regroupements en classes

Lorsque le caractère étudié est de type numérique, on est souvent amené à regrouper les valeurs prises dans des intervalles.

1- Tableau des effectifs

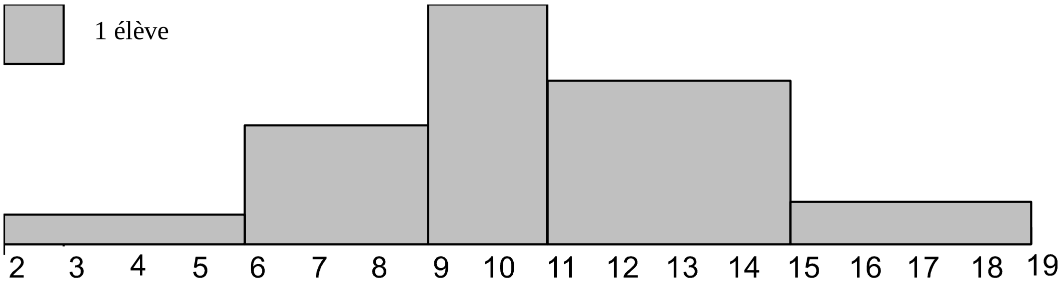
Considérons par exemple les notes obtenues à un devoir par les 30 élèves d'une classe.
Le relevé des notes est :
8 – 10 – 12 – 11,5 – 4 – 13 – 14,5 – 10,5 – 17 – 16,5 – 6 – 5,5 – 9 – 14 – 12 – 18 – 7,5 – 6 – 10,5 – 12 – 13 – 9 – 13,5 – 10 – 10,5 – 8 – 8,5 – 9,5 – 11 – 14,5
Indiquer l'effectif associé à chaque note serait peu lisible car il y a trop de notes différentes. On regroupe alors les notes dans des intervalles : [1 ; 6[, [6 ; 9[, [9 ; 11[, [11 ; 15[et [15 ; 19[.
Cela nous donne le tableau des effectifs suivants :

Classes de notes	[1 ; 6[	[6 ; 9[	[9 ; 11[	[11 ; 15[	[15 ; 19[
Effectifs	2	6	8	11	3

2- Histogrammes

Pour construire l'histogramme associé à ces données, nous devons graduer l'axe des abscisses qui contient les notes.

Attention
Ce sont les aires des différents rectangles qui doivent être proportionnelles aux effectifs; la légende indique l'aire correspondant à 1 élève.



Comme les largeurs des rectangles ne sont pas égales, leurs hauteurs ne sont pas proportionnelles aux effectifs. Ainsi, le groupe qui a l'effectif le plus élevé ne correspond pas au rectangle le plus haut, mais à celui qui a la plus grande aire.

3- Calcul de la moyenne et de la médiane

Le tableau suivant donne la répartition des employés d'une entreprise en fonction des salaires mensuels exprimés en euros.

Classes	[800;900[	[900;1000[	[1000;1050[	[1050;1150[	[1150;1300[
Effectifs	25	35	68	12	10

a) Moyenne

On calcule la moyenne de cette série en faisant comme si tous les individus d'une classe étaient associés au centre de la classe.
Cela revient donc à chercher la moyenne de la série suivante :

Valeurs	850	950	1025	1100	1225
Effectifs	25	35	68	12	10

La moyenne est donc :

$$m = \frac{25\times 850+35\times 950+68\times 1025+12\times 1100+10\times 1225}{25+35+68+12+10} = \frac{149650}{150} \simeq 997{,}67$$

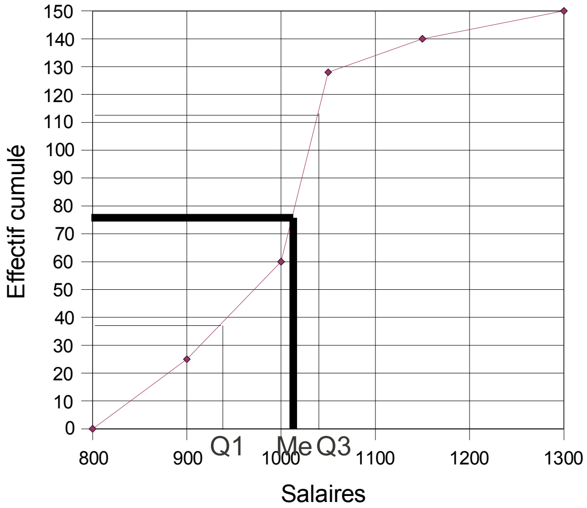
Le salaire moyen dans l'entreprise est donc 997,67 euros.

b) Médiane et quartiles

La médiane et les quartiles peuvent être déterminés graphiquement en utilisant le polygone des effectifs cumulés.
Le tableau des effectifs cumulés est :

Valeurs	800	900	1000	1050	1150	1300
Effectifs cumulés	0	25	60	128	140	150

Ce qui signifie :
- personne n'a un salaire inférieur à 800 euros
- 25 employés ont un salaire inférieur à 900 euros
- 60 employés ont un salaire inférieur à 1000 euros
- ….



Cela nous donne graphiquement le polygone des effectifs cumulés.

L'effectif total est de 150.
La médiane est la valeur correspondant à un effectif cumulé de 150/2 = 75; ainsi Me $\approx$ 1010 euros.
Le 1er quartile est la valeur correspondant à un effectif cumulé de 150/4 = 37,5 ; ainsi $Q_1 \approx 940$ euros.
Le 3ème quartile est la valeur correspondant à un effectif cumulé de $3\times150/4 = 112{,}5$; ainsi $Q_3 \approx 1040$ euros.