

STATISTIQUES DESCRIPTIVES

Les illustrations des notions s'appuient toutes sur cette série :

Une entreprise emploie 52 personnes. On souhaite étudier le montant en euros de leur salaire mensuel net. Le service comptable fournit le tableau suivant :

Salaire mensuel net en €	1000	1200	1500	2500	3000
Nombres de personnes	5	8	24	13	2

I. Vocabulaire

- La **population** désigne l'ensemble des personnes ou objets sur lesquels porte l'étude statistique.
- L'ensemble des données recueillies s'appelle **la série statistique**. La grandeur étudiée s'appelle **le caractère** étudié dans cette série statistique. Il prend différentes **valeurs** (numériques ou non)

Ex : *population -> 52 personnes* *Caractère -> montant du salaire mensuel*

- L'**effectif** d'une valeur du caractère est le nombre d'éléments de la population ayant cette valeur.
- L'**effectif total** est le nombre total de personnes ou d'objets étudiés dans une série.
- La **fréquence** d'une valeur du caractère est le quotient de l'effectif de la valeur par l'effectif total.

Ex : *valeur du caractère -> 1000, 1200, 1500, 2500, 3000*

Effectif total -> 5+8+24+13+2=52

Effectif de la valeur 1200 -> 8

Fréquence de la valeur 1200 -> $\frac{8}{52} = 0,154$ cela signifie que 15,4% des personnes gagnent 1200€.

- Pour calculer **les effectifs cumulés** croissants, on additionne les effectifs des valeurs inférieures à chaque valeur du caractère.
- Pour calculer **les fréquences cumulées croissantes**, on additionne les fréquences des valeurs inférieures à chaque valeur du caractère.

Ex :

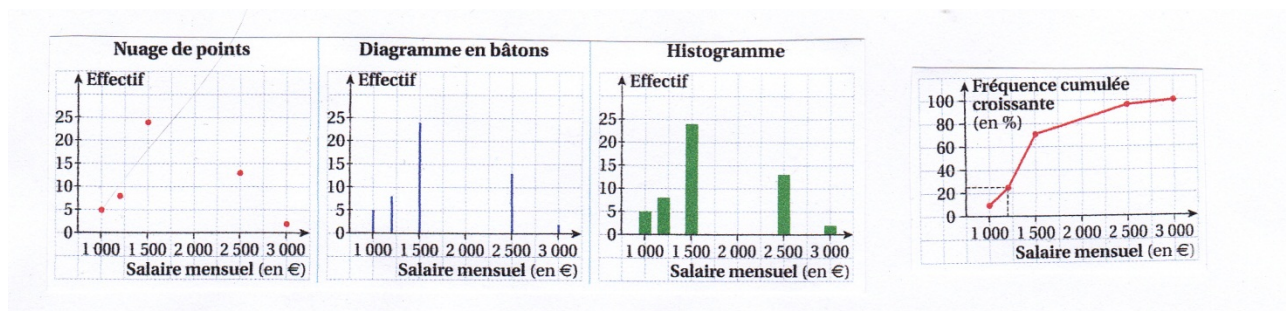
Salaire mensuel net en €	1000	1200	1500	2500	3000
Effectif (nbre de personnes)	5	8	24	13	2
Effectif cumulé croissant	5	13	37	50	52
Fréquence (arr. au centième)	0,1	0,15	0,46	0,25	0,04
Fréquence cumulée croissante	0,1	0,25	0,71	0,96	1

Cela signifie, par ex, que 37 salariés ont un salaire mensuel inférieur ou égal à 1500€, et qu'il représente 71% des salariés de l'entreprise.

II. Représentation d'une série statistique

- Un **nuage de points** est l'ensemble des points ayant pour abscisse une valeur du caractère (1^{ère} ligne du tableau) et pour ordonnée l'effectif correspondant (2^{ème} ligne du tableau).
- Un **histogramme** indique sous forme de rectangles les effectifs ou les fréquences correspondants à chaque valeur du caractère. Ils sont *principalement utilisés lorsque les valeurs sont regroupées en « classe »*. Dans ce cas, les rectangles sont adjacents. L'aire de chaque rectangle est proportionnelle la fréquence.
- Une courbe de fréquences cumulées relie le nuage de points qui correspondent aux fréquences cumulées croissantes.
- Les diagrammes circulaires représente chaque valeur du caractère par un secteur angulaire dont l'angle est proportionnel à l'effectif de la valeur.

Remarque : Les histogrammes



III. Résumer une série

Pour résumer une série, on utilise des caractéristiques de **position** (moyenne, médiane, quartiles...) et des caractéristiques de **dispersion** (étendue, écart interquartile ...).

1) Caractéristiques de position

Définition : La moyenne d'une série statistique dont les valeurs du caractère sont $x_1, x_2 \dots x_k$ et les effectifs correspondants $n_1, n_2 \dots n_k$ est notée $\bar{x}$ et se calcule ainsi :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n_1 + n_2 + \dots + n_k}$$

Si $f_1, f_2 \dots f_k$ sont les fréquences respectives des valeurs $x_1, x_2 \dots x_k$, on peut aussi calculer la moyenne d'une série en utilisant ces fréquences :

$$\bar{x} = f_1x_1 + f_2x_2 + \dots + f_kx_k$$

Ex :
$$\bar{x} = \frac{5 \times 1000 + 8 \times 1200 + 24 \times 1500 + 13 \times 2500 + 2 \times 3000}{5 + 8 + 24 + 13 + 2} = \frac{89100}{52} = 1713$$

ou
$$\bar{x} \approx 0,1 \times 1000 + 0,15 \times 1200 + 0,46 \times 1500 + 0,25 \times 2500 + 0,04 \times 3000 \approx 1715$$

Le salaire moyen de cette entreprise est 1713€.

Définition : La médiane notée m d'une série statistique est la valeur du caractère qui partage la série en deux groupes de même effectif : l'un des groupe contient les valeurs inférieures ou égales à m et l'autre groupe contient les valeurs supérieures ou égales à m .
Pour un effectif impair $2n+1$ c'est la $(n+1)^{\text{ième}}$
Pour un effectif pair $2n$ c'est la moyenne de la $n^{\text{ième}}$ et $(n+1)^{\text{ième}}$ valeur.

Ex : La fréquence cumulée 0,5 est atteinte et dépassée pour un salaire de 1500€.

La médiane de cette série est $m=1500\text{€}$.

Définition : Le **premier quartile** noté Q_1 est la plus petite des valeurs du caractère telle qu'au moins 25% des valeurs de la série sont inférieures ou égales à Q_1 : Q_1 est la 1^{ère} valeur de la série pour laquelle la fréquence cumulée croissante 0,25 est atteinte ou dépassée.

Définition : Le troisième quartile noté Q_3 est la plus petite des valeurs du caractère telle qu'au moins 75% des valeurs sont inférieures ou égales à Q_3 : Q_3 est la 1^{ère} valeur de la série pour laquelle la fréquence cumulée croissante 0,75 est atteinte ou dépassée.

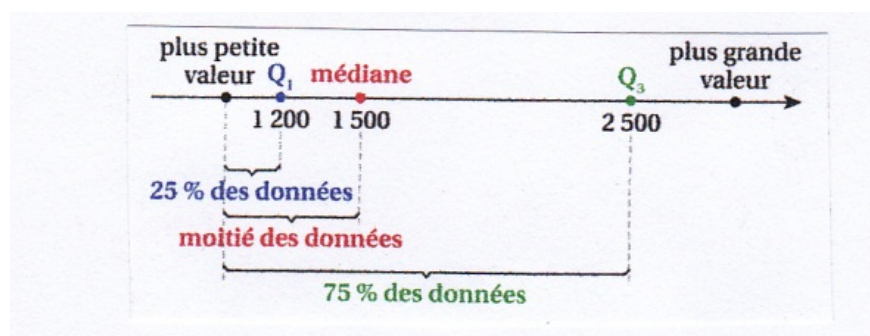
Remarques : → On peut aussi utiliser la fréquence cumulée croissante 0,5 pour calculer la médiane mais attention uniquement si la valeur 0,5 est dépassée mais pas atteinte, sinon la méthode n'est plus correcte

→ Quartiles et médiane sont les valeurs d'une série statistiques qui la sépare en 4 groupes de même effectif.

Ex : L'effectif total est de 52.

La fréquence cumulée 0,25 est atteinte pour un salaire de 1200€. $Q_1 = 1200\text{€}$

La fréquence cumulée 0,75 est atteinte et dépassée pour un salaire de 2500€. $Q_3 = 2500\text{€}$



Concrètement, cela veut dire que, dans cette entreprise, 25% des salaires sont inf ou égaux à 1200€, au moins 50% des salaires sont inf ou égaux à 1500€ et au moins 75% des salaires sont inf ou égaux à 2500€.

2) Caractéristiques de dispersion

Définition : L'**étendue** d'une série statistique est la différence entre la valeur minimale et la maximale d'une série.

Ex : $3000 - 1000 = 2000$. L'étendue de la série des salaires est 2000€ .

Définition : L'intervalle inter-quartile d'une série statistique est $[Q_1; Q_3]$. Il contient au moins 50% des valeurs de la série. Son amplitude $Q_3 - Q_1$ est appelée écart interquartile.

Ex : $Q_3 - Q_1 = 2500 - 1200 = 1300$. L'écart interquartile de cette série de salaires est 1300€ .

Cela signifie que au moins 50% des salariés ont un salaire compris entre 1200 et 1500€.