

Statistiques & échantillonnage

Un exemple pour tout le cours :

Salaire mensuel net (en euros)	1000	1200	1500	2500	3000
Nombre de personnes	5	8	24	13	2

I. Statistiques descriptives

I.1. 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**. Elle énumère des propriétés des individus de la population, appelés **caractères**.

Exemple: La population est :

Le caractère étudié est :

Les valeurs du caractère sont :

- Le nombre d'éléments de la population pour lesquels la variable prend une valeur donnée est l'**effectif** de cette valeur.

- La **fréquence** d'une valeur est égale à $\frac{\text{effectif de la valeur}}{\text{effectif total}}$

Exemple: L'effectif total est :

L'effectif de la valeur 1200 est :

La fréquence de la valeur 1200 est :

Définition 1

- Les **effectifs cumulés croissants** donnent les effectifs des valeurs inférieures à chaque valeur du caractère. De manière similaire, on peut étudier les **effectifs cumulés décroissants**.
- Les **fréquences cumulées croissantes** donnent les fréquences des valeurs inférieures à chaque valeur du caractère. De manière similaire, on peut étudier les **fréquences cumulées décroissantes**.

Exemple: Remplir le tableau suivant :

Salaire mensuel net (en euros)	1000	1200	1500	2500	3000
Effectif cumulé croissant					
Fréquence cumulée croissante (en %)					

I.2. Représentation d'une série statistique

1. Nuage de points

2. Diagrammes en bâtons

3. Histogramme

4. courbe des fréquences cumulées

Exemple: Quel est le pourcentage de salariés ayant un salaire mensuel inférieur ou égal à 1200 euros ?

I.3. Moyenne

Définition 2

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 vaut :

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

Exemple: La moyenne des salaires de l'entreprise est :

I.4. Médianes et quartiles

Définition 3

La **médiane** m est une valeur du caractère étudié telle que la moitié de l'effectif ait des valeurs inférieures à m et l'autre moitié des valeurs supérieures à m .

Pour déterminer la médiane de N valeurs, on range ces valeurs par ordre croissant.

- si N est impair, la médiane m est la valeur du caractère numéroté $\frac{N+1}{2}$
- si N est pair, la médiane m est le milieu entre les valeurs numérotées $\frac{N}{2}$ et $\frac{N}{2} + 1$

Définition 4

- Le **premier quartile** est le plus petit élément Q_1 des valeurs des termes de la série, tel qu'au moins 25% des données sont inférieures ou égales à Q_1 .
- Le **troisième quartile** est le plus petit élément Q_3 des valeurs des termes de la série, tel qu'au moins 75% des données sont inférieures ou égales à Q_3 .

Exemple: Pour l'entreprise, la médiane est égale à :

Le premier quartile vaut :

Le deuxième quartile vaut :

II. Echantillonnage

II.1. Echantillon

Définition 5

Un échantillon de taille n est obtenu à partir d'une population en répétant n fois de suite l'opération suivante : on prélève au hasard un de ses éléments, on note la valeur caractère prélevé et on remet l'élément prélevé dans la population.

Remarque: Souvent, pour des raisons pratiques, il n'y a pas de remise, mais les résultats restent vrais si l'effectif total est très grand par rapport à l'effectif de l'échantillon.

II.2. Intervalle de fluctuation

Soit une population dont on étudie un caractère pouvant prendre deux valeurs "vrai" ou "faux". On suppose connaître la proportion p (**proportion effective**) de la population pour laquelle la caractère est vrai, et on mesure la proportion f d'un échantillon (**fréquence observée**).

Définition 6

L'**intervalle de fluctuation au seuil de 95%** d'une fréquence d'un échantillon de taille n est l'intervalle I centré autour de la proportion effective p tel que la fréquence observée f se trouve dans I avec une probabilité égale à 0,95.

Propriété 1

Soit une population pour laquelle on connaît la proportion effective p d'un caractère, comprise entre 0,2 et 0,8. L'intervalle de fluctuation au seuil de 95% de f est l'intervalle $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}\right]$.