

Statistiques descriptives

I Série statistique. Tableau d'effectifs. Fréquences d'apparition.

Définition : Série statistique

Lors d'un relevé de mesures effectué sur les individus d'une population, l'ensemble des données collectées constitue une **série statistique**.

☞ Une série statistique à caractère quantitatif est dite **ordonnée** après que les valeurs collectées ont été rangées dans l'ordre croissant (ou décroissant).

☞ L'**étendue** désigne l'écart entre la plus grande et la plus petite des valeurs prises par le caractère.

Définition : Fréquence d'apparition

On considère une série statistique comportant p modalités (ou p classes) d'effectifs $n_1, \dots, n_p$ et d'effectif total N .

La **fréquence d'apparition** de la modalité (ou de la classe) correspond à la proportion d'individus dont le caractère est égal à cette modalité (ou appartenant à cette classe).

Ainsi, pour tout entier i compris entre 1 et p :

$$f_i = \frac{n_i}{N} \quad \text{et} \quad f_1 + f_2 + \dots + f_p = 1$$

II Médiane et quartiles

Dans cette partie et la suivante, on considère uniquement des séries statistiques à caractère quantitatif.

Définition : Médiane

Dans une série statistique **ordonnée** :

une **médiane** partage les valeurs de la série en deux groupes de même effectif.

Définition : Quartiles

Le **premier quartile** d'une série statistique numérique est la plus petite valeur de la série telle qu'au moins 25% des valeurs lui soient inférieures ou égales.

Le **troisième quartile** d'une série statistique numérique est la plus petite valeur de la série telle qu'au moins 75% des valeurs lui soient inférieures ou égales.

III Moyenne

Définition : Moyenne

La **moyenne** d'une série statistique se note $\bar{x}$.

$$\bar{x} = \frac{\text{somme totale des valeurs prises par le caractère}}{\text{nombre de valeurs}}$$

Si $x_1, x_2, \dots, x_p$ désignent les p modalités du caractère d'une série statistique et $n_1, n_2, \dots, n_p$ désignent les effectifs correspondants, alors

$$\bar{x} = \frac{n_1 \times x_1 + n_2 \times x_2 + n_3 \times x_3 + \dots + n_p \times x_p}{n_1 + n_2 + n_3 + \dots + n_p}$$

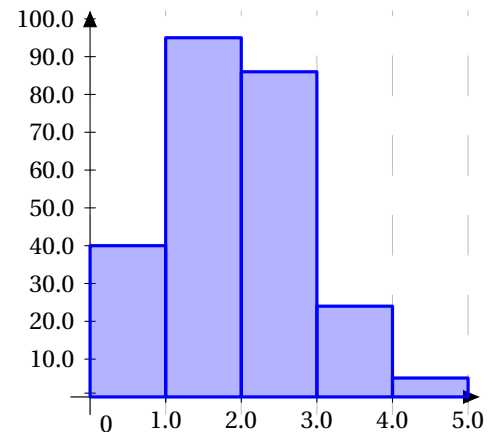
Exemple 1 : Exemple d'étude statistique

Une enquête sur le temps (en heures) de travail personnel quotidien des élèves de Seconde d'un lycée a donné les résultats suivants :

Les réponses ont été rangées dans 5 classes d'amplitude une heure.

Temps	[0; 1[	[1; 2[	[2; 3[	[3; 4[	[4; 5[
Effectif	40	95	86	24	5

Il est alors possible de représenter ces données sous forme d'un **histogramme** :



Moyenne :

Même s'il n'est pas possible ici de connaître exactement la réponse de chaque élève, il est possible d'estimer la moyenne par calcul. L'effectif total est de 250 et on suppose que les réponses des élèves sont uniformément réparties dans chaque intervalle d'heure. On prend donc le **centre de chaque classe** :

Estimation de la moyenne : $\frac{40 \times 0.5 + 95 \times 1.5 + 86 \times 2.5 + 24 \times 3.5 + 5 \times 4.5}{250} = 1,936$ soit presque 2 h.

Médiane, quartiles et fréquences cumulées :

Une estimation de la médiane et des quartiles peut être calculée à partir du tableau (en supposant toujours une répartition uniforme des réponses à l'intérieur de chaque classe).

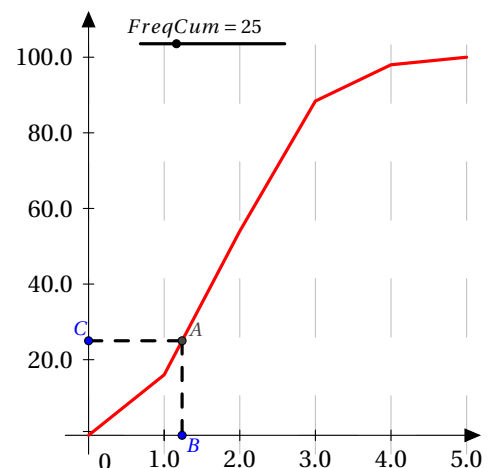
Cependant, la lecture sur une **courbe de fréquences cumulées** croissantes permet d'estimer rapidement ces indicateurs.

Il faut donc compléter le tableau avec :

- une ligne des **Effectifs Cumulés Croissants** (ECC)
- une ligne des **Fréquences Cumulées Croissantes** (FCC)

Temps	[0; 1[	[1; 2[	[2; 3[	[3; 4[	[4; 5[
Effectif	40	95	86	24	5
ECC	40	135	221	...	250
FCC	16%	54%	88,4%	...	100%

La ligne brisée ci-contre représente la **courbe des fréquences cumulées croissantes** :



En se positionnant à une hauteur de 25 (soit 25% en fréquences cumulées), on peut lire l'abscisse du point B : 1,24.

On estime donc le premier quartile à : 1,24 h.



Compléter le dernier tableau et lire graphiquement la **médiane** et le **troisième quartile**.