

I/ Vocabulaire

Les premières études statistiques étaient démographiques : on en a conservé le vocabulaire.

- L'ensemble sur lequel porte l'étude statistique s'appelle la **population**.
- Si cet ensemble est trop vaste, on en restreint l'étude à une partie appelée **échantillon**.
- Un **individu** est un élément de la population étudiée.
- Le **caractère** (ou la **variable**) est une propriété que l'on souhaite étudier de la population.

Deux types de caractères :

- Les **caractères quantitatifs**, c'est à dire des caractères auxquels on peut associer des nombres.
 - Le caractère est **discret** lorsqu'il ne peut prendre qu'un nombre fini de valeurs.
 - Le caractère est **continu** lorsqu'il peut prendre toutes les valeurs d'un intervalle.
- Les caractères **qualitatifs** sont les caractères qui ne sont pas quantitatifs.

Effectuer une étude statistique consiste à ;

- 1/ Recueillir des informations (enquête)
- 2/ Les représenter et les comparer (au moyen de tableaux, de calculs, de graphiques, de schémas, ...)
- 3/ En tirer des conclusions sur un caractère d'une population, c'est à dire une propriété que possèdent les individus de cette population.

- L'**effectif total**, souvent noté N, est le nombre d'individus de la population.
- L'**effectif** d'une valeur du caractère est le nombre d'individus ayant cette valeur.
- Les **effectifs cumulés croissants** s'obtiennent en ajoutant au fur et à mesure les effectifs des valeurs précédentes, lorsque les valeurs sont rangées par ordre croissant.
- La **fréquence** d'une valeur du caractère est le quotient $\frac{\text{effectif de la valeur}}{\text{effectif total}}$, c'est donc un nombre compris entre **0 et 1**.
- Les **fréquences cumulées croissantes** s'obtiennent en ajoutant au fur et à mesure les fréquences des valeurs précédentes, lorsque les valeurs sont rangées par ordre croissant.

Exemple utilisé tout au long de ce chapitre :

Dans un service, on a répertorié le nombre d'interventions par jour sur un mois et leur durée :

Exemple explicatif : il y a eu 2 jours dans le mois où le service est intervenu 3 fois dans la journée.

1/ Préciser dans les deux premières lignes celle qui correspond aux *valeurs* et aux *effectifs*.

2/ Quelle est la fréquence du nombre de jour où il y a eu 3 interventions ?

$$\text{fréquence} = \frac{\text{effectif de la valeur}}{\text{effectif total}} = \frac{2}{25} = 0.08.$$

3/ Compléter le tableau ci-dessous :

Nombre d'interventions x_i (Valeurs : caractères)	3	5	6	7	8	9
Nombre de jours n_i (effectifs)	2	5	6	8	3	1
Effectif cumulé croissant (nombre de jours où il y a eu au plus x_i interventions)	2	$2+5$ = 7	$2+5+6$ = 6+7 = 13	$2+5+6+8$ = 13+8 = 21	$2+5+6+8+3$ = 21+3 = 24	somme totale = 24+1 = 25
Fréquence f_i du nombre de jour où il y a eu le même nombre d'intervention x_i	$\frac{2}{25}$	$\frac{1}{5}$	$\frac{6}{25}$	$\frac{8}{25}$	$\frac{3}{25}$	$\frac{1}{25}$
Fréquences cumulées croissantes (en %) (fréquence du nombre de jours où il y a eu au plus x_i interventions)	$\frac{2}{25}$	$\frac{2}{25} + \frac{1}{5}$ = $\frac{7}{25}$	$\frac{2}{25} + \frac{1}{5} + \frac{6}{25}$ = $\frac{7}{25} + \frac{6}{25}$ = $\frac{13}{25}$	$\frac{2}{25} + \frac{1}{5} + \frac{6}{25} + \frac{8}{25}$ = $\frac{13}{25} + \frac{8}{25}$ = $\frac{21}{25}$	$\frac{2}{25} + \dots + \frac{8}{25} + \frac{3}{25}$ = $\frac{21}{25} + \frac{3}{25}$ = $\frac{24}{25}$	somme totale = $\frac{24}{25} + \frac{1}{25}$ = $\frac{25}{25}$

4/ Quelle est la fréquence des jours où il y a eu au plus 6 interventions ?

C'est la **fréquences cumulées croissantes** qui renseigne de cette fréquence ; réponse : $\frac{13}{25} = 0.52$.

II/ Indicateurs de position centrale

On résume en général une étude statistique à caractère quantitatif discret dans un tableau :

Valeur	x_1	x_2	...	x_i	...	x_p	Total
Effectif	n_1	n_2	...	n_i	...	n_p	N
Fréquence	f_1	f_2	...	f_i	...	f_p	1

Exemple obtenu au I/ :

Valeur x_i	3	5	6	7	8	9	Total
Effectif n_i	2	5	6	8	3	1	25
Fréquence f_i	$\frac{2}{25}$	$\frac{1}{5}$	$\frac{6}{25}$	$\frac{8}{25}$	$\frac{3}{25}$	$\frac{1}{25}$	1

1/ La moyenne

Définition :

La **moyenne** notée $\bar{x}$ de cette série est le quotient de la somme des valeurs par l'effectif total.

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \dots + n_px_p}{n_1 + n_2 + \dots + n_p},$$

Propriété :

La moyenne peut être obtenue directement avec les fréquences : $\bar{x} = f_1 \times x_1 + f_2 \times x_2 + \dots + f_p \times x_p$.

Exemple : La moyenne de la série 1 est $\bar{x} = \frac{3 \times 2 + 5 \times 5 + 6 \times 6 + 7 \times 8 + 8 \times 3 + 9 \times 1}{2 + 5 + 6 + 8 + 3 + 1} = 6.24$.

Propriété :

Lorsque toutes les valeurs d'une série statistique sont multipliées par le même réel m et sont toutes augmentées d'un même réel p , alors la **moyenne** de cette série est aussi multipliée par m et est augmentée par p , c'est-à-dire est $m\bar{x} + p$.

Exemple :

Un enseignant remarque qu'une de ses classes a 8 de moyenne à un devoir et que personne n'y dépasse 10.

- S'il multiplie par 2 chacune des notes obtenues, la moyenne de la classe devient 16 ($=8 \times 2$).
- S'il augmente chaque note de 3, la moyenne de classe devient 11 ($=8+3$).

2/ La médiane

Définition :

La **médiane Me** d'une série statistique partage cette série en deux parties de telle sorte que :

- **au moins la moitié** des valeurs sont inférieures ou égales à la médiane ;
- **au moins la moitié** des valeurs sont supérieures ou égales à la médiane.

Méthode : Si la série contient n valeurs rangées dans l'ordre croissant :

- si n est impair, on prend la $\frac{n+1}{2}$ ième valeur pour médiane.
- si n est pair, on prend pour médiane la moyenne entre la $\frac{n}{2}$ ième et la $\frac{n}{2} + 1$ ième valeur.

Exemple :

- La médiane des notes 6 ; 10 ; 15 ; 14 ; 9 ; 13 est $\frac{10+13}{2}$, soit 11.5 car :

6 9 10 || 13 14 15
 Me

- La médiane de la série statistique de l'exemple 2 est 6 car **52% des valeurs sont inférieures ou égales à 6** tandis que **72% des valeurs sont supérieures ou égales à 6** ; on a bien **au moins 50%** des valeurs plus petites que la médiane et **au moins 50%** des valeurs plus grandes que cette médiane.

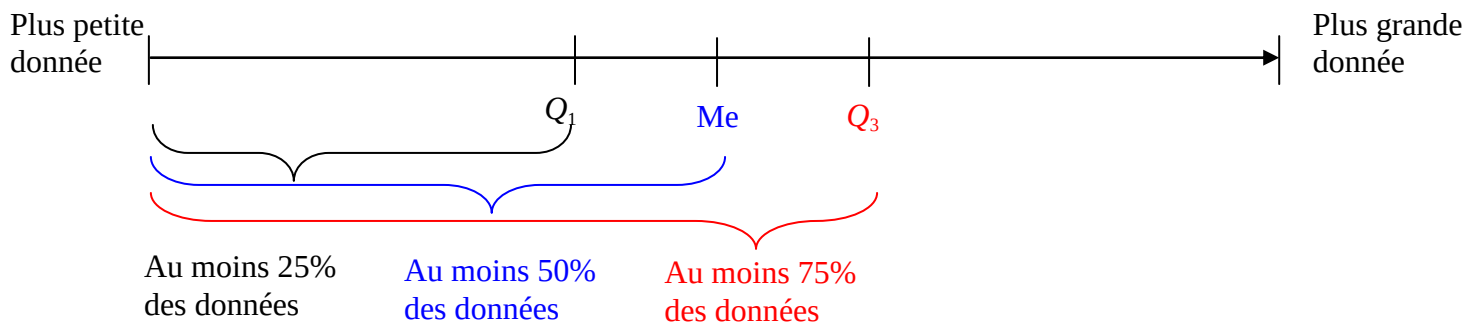
Valeur	3	5	6	7	8	9
Effectif	2	5	6	8	3	1
ECC	2	7	13	21	24	25
Fréquence	$\frac{2}{25}$	$\frac{1}{5}$	$\frac{6}{25}$	$\frac{8}{25}$	$\frac{3}{25}$	$\frac{1}{25}$
FCC	0.04	0.28	0.52	0.84	0.96	1

3/ Les quartiles

Définition :

La liste des N données d'une série statistique est rangée par ordre croissant.

- Le premier quartile est la plus petite donnée Q_1 de la liste telle **qu'au moins un quart** des données de la liste sont inférieures ou égales à Q_1 .
- Le troisième quartile est la plus petite donnée Q_3 de la liste telle **qu'au moins les trois quarts** des données de la liste sont inférieures ou égales à Q_3 .



Calcul pratique des quartiles :

- Pour Q_1 , on calcule $\frac{N}{4}$, puis on détermine le premier entier supérieur ou égal à $\frac{N}{4}$; cet entier est le rang de Q_1 que l'on peut alors déterminer.
- Pour Q_3 , on fait de même avec $\frac{3N}{4}$.

Exemples : Dans l'exemple discret du I/ :

Pour Q_1 :

$N = 25$ donc $\frac{N}{4} = 6.25$. Q_1 est la 7ème valeur.

Ainsi, $Q_1 = 5$.

Remarque :

28% des valeurs inférieures ou égales à Q_1 : on a bien **au moins 25%** des valeurs inférieures à Q_1 .

Pour Q_3 :

$N = 25$ donc $\frac{3N}{4} = 18.75$. Q_3 est la 19ème valeur.

Ainsi, $Q_3 = 7$.

Remarque : 84% des valeurs inférieures ou égales à Q_3 : on a bien **au moins 75%** des valeurs inférieures à Q_3 .

Valeur	3	5	6	7	8	9
Effectif	2	5	6	8	3	1
ECC	2	7	13	21	24	25
Fréquence	2/25	1/5	6/25	8/25	3/25	1/25
FCC	0.04	0.28	0.52	0.84	0.96	1

III/ Indicateurs de dispersion

1/ Écart interquartile

Définition :

On considère une série statistique de premier quartile Q_1 et de troisième quartile Q_3 .

- l'intervalle interquartile** est l'intervalle $[Q_1 ; Q_3]$.

Intérêt : la proportion de valeurs du caractère se trouvant dans l'intervalle interquartile est d'environ 50% : celles les plus centrales.

- Une caractéristique de dispersion **l'écart interquartile** $Q_3 - Q_1$.

Exemple :

L'intervalle interquartile de la série statistique du I/ est : **[5;7]**, l'écart interquartile est **7-5 = 2**.

2/ Écart-type

Définition :

L'**écart-type** σ d'une série statistique est la racine carrée de la moyenne des carrés des écarts entre les valeurs de

la série et la moyenne $\bar{x}$. $\sigma = \sqrt{\frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}}$.

Exemple :

L'écart-type de la série statistique du I/ est proche de **1.619**.

$$\text{En effet, } \sigma = \sqrt{\frac{n_1(x_1 - \bar{x})^2 + n_2(x_2 - \bar{x})^2 + \dots + n_p(x_p - \bar{x})^2}{n_1 + n_2 + \dots + n_p}}$$
$$= \sqrt{\frac{2 \times (2 - 6.24)^2 + 5 \times (5 - 6.24)^2 + 6 \times (6 - 6.24)^2 + 8 \times (7 - 6.24)^2 + 3 \times (8 - 6.24)^2 + 1 \times (9 - 6.24)^2}{2 + 5 + 6 + 8 + 3 + 1}} = \sqrt{\frac{65.52}{25}} \approx 1.618889.$$

Remarque :

L'écart-type est souvent calculé à la calculatrice.

Remarque :

En pratique, une série statistique sera souvent résumée par un indicateur de position centrale (moyenne ou médiane, ...) et d'un indicateur de dispersion (écart interquartile ou écart-type, ...)

Exemple :

On peut résumer la série statistique du I/ donnant le nombre d'intervention par :

- le couple (moyenne ; écart – type) : **(6.24 ; 1.619)**.
- le couple (médiane ; écart interquartile) : **(6 ; 2)**.