

Statistiques

1 Vocabulaire

La **population** est l'ensemble des individus sur lesquels portent une étude statistique. (Par exemple classe de seconde, habitants de la France ...)

Le **caractère** (ou variable) d'une série statistique est une propriété étudiée sur chaque individu :

- Lorsque le caractère ne prend que des valeurs (ou modalités) numériques, il est **quantitatif** :
 - **discret** s'il ne peut prendre que des valeurs isolées (notes, âge ...)
 - **continu** dans le cas contraire (poids, taille ...). On peut alors effectuer un regroupement des valeurs par **classes**.
- Sinon, on dit qu'il est **qualitatif** (couleur des yeux, sport pratiqué ...) : les modalités ne sont pas des nombres.

Exemple 1.1 Le tableau suivant recense les salaires annuels bruts en kilo euros de vingt salariés d'une entreprise :

Salaire x_i	21	22	22	81	37	24	24	23	23	24	23	22	22	23	37	23	21	23	24	24
---------------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Population=	Modalités																				
Caractère=	Effectif																				
Effectif total=	Fréquence																				
Mode=	Cumul des Freq.																				
Étendue=																					

2 Caractéristiques d'une série

2.1 Moyenne

Définition 2.1 *Moyenne* On considère une série statistique dont les p valeurs sont données par x_1, x_2, \dots, x_p , d'effectifs associés n_1, n_2, \dots, n_p :

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

La **moyenne** de cette série est le nombre noté \bar{x} qui vaut :

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

Exercice 1 Calculer la moyenne de la série A des notes :

Notes	2	3	4	5	6	7	8	9	10	11	13	15	16	Tot
Effectifs	1	2	1	1	2	3	5	6	2	3	2	1	1	30
Fréquences	0,03	0,07	0,03	0,03	0,07	0,10	0,17	0,20	0,07	0,10	0,07	0,03	0,03	1

Exercice 2 Lorsque la série est regroupée en **classes**, on calcule la moyenne en prenant pour valeurs x_i le centre de chaque classe; ce centre est obtenu en faisant la moyenne des deux extrémités de la classe.

Notes	[0;4[[4;8[[8;12[[12;16[[16;20]
Centre des classes					
Effectifs	3	7	16	3	1

Si on regroupe par classe d'amplitude 4, une estimation de la moyenne est : $\bar{x} =$

2.2 Variance, écart-type

Définition 2.2 La variance est la moyenne des carrés des écarts à la moyenne :

$$\text{Var}(x) = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{N} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{N} \left(\sum_{i=1}^N x_i^2 \right) - \bar{x}^2$$

L'écart-type est la racine carrée de la variance :

$$\sigma(x) = \sqrt{\text{Var}(x)}$$

Exercice 3 Calculer $\text{Var}(x)$ et $\sigma(x)$ pour la série A des notes.

2.3 Médiane, quartiles

Définition 2.3 Médiane On appelle **médiane** d'une série statistique considère une série statistique une valeur, notée **Me**, telle que :

1. au moins 50% des valeurs sont inférieures ou égales à Me.
2. au moins 50% des valeurs sont supérieures ou égales à Me.

Exercice 4 1. Déterminer la médiane de la série ordonnée : 2 – 3 – 5 – 5 – 11 – 12 – 12 – 17 – 18.

2. Calculer le pourcentage de valeurs inférieures ou égales à Me.
3. Calculer le pourcentages de valeurs supérieures ou égales à Me.

↪ *Solution* : $\underbrace{2 - 3 - 5 - 5}_{4 \text{ valeurs}} - \underbrace{11}_1 - \underbrace{12 - 12 - 17 - 18}_{4 \text{ valeurs}}$ donc Me=11.

Exercice 5 Déterminer la médiane de la série A des notes.

Remarque : mieux vaut ne pas retenir la formule $Me = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$ mais retenir que :

1. si N est impair alors $N = 2k + 1$ donc la série ordonnée s'écrit : $k - 1 - k$ donc Me est la valeur de rang $k + 1$.
2. si N est pair alors $N = 2k$ donc la série s'écrit : $k - k$ donc Me est la moyenne des termes de rangs k et $k + 1$.

Définition 2.4 Quartiles On considère une série statistique dont les valeurs sont rangées dans **l'ordre croissant**.

1. le **premier quartile** noté Q_1 est la plus petite valeur de la série telle que au moins **25%** des valeurs sont inférieures ou égales à Q_1 .
2. le **troisième quartile** noté Q_3 est la plus petite valeur de la série telle que au moins **75%** des valeurs sont inférieures ou égales à Q_3 .

Remarque Il existe plusieurs définitions de la médiane, des quartiles. Dans tous les cas, on peut dire que 25% des valeurs de la série sont inférieures ou égales à Q_1 , 50% des valeurs de la série sont inférieures ou égales à la médiane et 75% des valeurs de la série sont inférieures ou égales à Q_3 . (Les quartiles divisent la série en quarts)

Exercice 6 Déterminer les quartiles de la série : $5 - 8 - 3 - 7 - 15 - 13 - 3 - 8 - 2 - 11 - 12 - 19 - 6$.

↪ *Solution* :

1. On ordonne : $2 - 3 - 3 - 5 - 6 - 7 - 8 - 8 - 11 - 12 - 13 - 15 - 19$: il y a $N = 13$ valeurs.
2. On calcule 25% de 13 : $0,25 \times 13 = 3,25$: Q_1 est la 4^e valeur de la série dont $Q_1 = 5$.
3. On calcule 75% de 13 : $0,75 \times 13 = 9,75$: Q_3 est la 10^e valeur de la série dont $Q_3 = 11$.

Exercice 7 Déterminer les quartiles de la série A des notes.

3 Mesures de dispersion d'une série statistique

Définition 3.1 1. *L'étendue d'une série statistique est la différence entre la plus grande valeur de la série et la plus petite.*

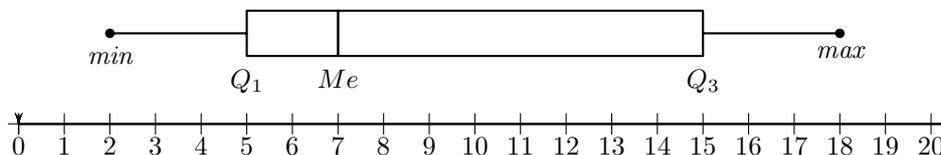
2. *L'écart interquartile d'une série statistique est la différence $Q_3 - Q_1$.*

Remarque La **variance** est aussi une mesure de dispersion. Plus l'écart-type est petit, plus les valeurs de la série se concentrent autour de la moyenne.

Définition 3.2 *Diagramme en boîte*

Un **diagramme en boîte** se construit de la manière suivante :

1. les valeurs du caractère sont représentées sur un axe (vertical ou horizontal) ;
2. on place sur cet axe, le minimum, le maximum, les quartiles et la médiane de la série ;
3. on construit alors un rectangle parallèlement à l'axe, dont la longueur est l'interquartile et la largeur arbitraire.



Remarque Un tel diagramme permet non seulement de visualiser la dispersion d'une série mais aussi de comparer plusieurs séries entre elles (Voir activité).

On peut y incorporer aussi le premier décile et neuvième décile.

4 Effectifs et fréquences cumulées

Définition 4.1 *ECC, ECD, FCC, FCD*

1. *L'effectif cumulé croissant (respectivement décroissant) d'une valeur est la somme des effectifs des valeurs inférieures (respectivement supérieures) ou égales à cette valeur.*
2. *la fréquence cumulée croissante (respectivement décroissante) d'une valeur est la somme des fréquences des valeurs inférieures (respectivement supérieures) ou égales à cette valeur.*

Exercice 8 Dans la série A, donner les effectifs cumulés **croissants** :

Notes	2	3	4	5	6	7	8	9	10	11	13	15	16	Tot
Effectifs	1	2	1	1	2	3	5	6	2	3	2	1	1	30
ECC	1													

Remarque *Intérêts* :

1. Recherche de la médiane, des quartiles, déciles.
2. Construction du polygone des ECC, ECD, FCC, FCD.