I. SERIES STATISTIQUES

Réaliser une étude statistique consiste à classer les **individus** d'une **population** en fonction d'un **caractère** (ou **variable**).

Exemples :
- Classer les **élèves** d'une **classe** en fonction de leur **âge**.
- Classer les **voitures** garées sur un **parking** en fonction de leur **couleur**.
- Classer les **joueurs** d'une **équipe de foot** en fonction de leur **poste**.
- Classer des **forfaits** d'un **opérateur téléphonique** en fonction de leur **prix**.

Le caractère étudié peut être **qualitatif** (couleur, poste des joueurs de foot…) ou **quantitatif** (âge, prix…).
En regroupant les valeurs du caractère sur toute une population, on obtient une **série statistique**.

Que l'on peut représenter sous forme d'une **liste** ou d'un **tableau**.

Exemple : Dans une équipe de football, on demande aux joueurs leur poste sur le terrain et on obtient les résultats suivants (Défenseur, Gardien, Milieu, Avant) :
→ sous forme de liste : D ; A ; G ; D ; A ; M ; M ; D ; M ; A ; D
→ sous forme de tableau :

Poste (caractère)	G	D	M	A	TOTAL
Effectif	**1**	**4**	**3**	**3**	**11**
Fréquence (%)	**0,091 (9,1 %)**	**0,364 (36,4 %)**	**0,273 (27,3 %)**	**0,273 (27,3 %)**	**1 (100 %)**

Pour chaque valeur du caractère, on peut indiquer l'**effectif** (le nombre d'individus) ou la **fréquence** (la proportion d'individu par rapport à la totalité de la population).

Cette fréquence peut s'exprimer sous la forme :
- d'un nombre décimal entre 0 et 1 (Exemple : $0{,}057$)
- d'un pourcentage (Exemple : $5{,}7\ \%$)
- d'une fraction (Exemple : $\frac{2}{35}$)

En règle générale :

Valeur du caractère	x_1	x_2	…	x_p	TOTAL
Effectif	n_1	n_2	…	n_p	$\sum_{i=1}^{p} n_i = N$
Effectif cumulé croissant	n_1	n_1+n_2	…	$n_1+n_2+\ldots+n_p$	
Effectif cumulé décroissant	$n_1+n_2+\ldots+n_p$	$n_2+\ldots+n_p$	…	n_p	
Fréquence	$f_1=\frac{n_1}{N}$	$f_2=\frac{n_2}{N}$	…	$f_p=\frac{n_p}{N}$	1

II. CARACTERISTIQUES DE DISPOSITION

a. Moyenne simple :

Soit x un caractère qui prend les valeurs $x_1, x_2, \ldots, x_p$, alors la moyenne de cette série statistique est :

$$\overline{x} = \frac{\sum_{i=1}^{p} x_i}{p}$$

***Exemple* :** Voici les 8 notes (sur 20) d'un élève ce trimestre : 12 ; 15 ; 12 ; 13 ; 10 ; 19 ; 11 ; 7

Alors : $\overline{x} = \frac{13 + 11 + 12 + 13 + 10 + 19 + 11 + 7}{8} = \frac{96}{8} = 12$

b. Moyenne pondérée (coefficientée) :

Soit x un caractère qui prend les valeurs $x_1, x_2, \ldots, x_p$.

Les effectifs respectifs de chaque valeur sont $n_1, n_2, \ldots, n_p$.

L'effectif total est donc $\sum_{i=1}^{p} n_i = N$

Alors la moyenne de cette série statistique est : $\bar{x} = \dfrac{\sum_{i=1}^{p} n_i x_i}{p}$

Exemple **:** Dans une classe, 7 élèves on eu 12, 3 élèves ont eu 15, 5 élèves ont eu 9 et 1 seul élève a eu 18.

Alors la moyenne de la classe est : $\bar{x} = \dfrac{7 \times 12 + 3 \times 15 + 5 \times 9 + 1 \times 18}{7 + 3 + 5 + 1} = \dfrac{192}{16} = 12$

c. Médiane :

C'est la valeur (de l'âge) qui se trouve au MILIEU de la série, qui la partage en deux séries d'effectif égal.

Exemple **:** voici les âges (par ordre croissant) de 22 individus :

$$\underbrace{21 \quad 22 \quad 22 \quad 24 \quad 26 \quad 26 \quad 26 \quad 26 \quad 28 \quad 28 \quad \mathbf{28}}_{\text{11 individus}} \;\Big|\; \underbrace{\mathbf{28} \quad 28 \quad 29 \quad 29 \quad 30 \quad 31 \quad 31 \quad 32 \quad 32 \quad 34 \quad 35}_{\text{11 individus}}$$

Médiane = **28**

La médiane de cette série statistique est de 28 ans.

Remarques :

- Dans le cas où l'effectif de la série est impair, la « ligne de partage » est située juste sur une valeur : C'est la valeur médiane.
  - Dans le cas où l'effectif de la série est pair (dans notre exemple), la « ligne de partage » est située juste entre deux valeurs de la série. Si ces deux valeurs sont différentes, on prend leur demi-somme pour valeur médiane.

III. CARACTERISTIQUES DE DISPERSION

Deux séries de données peuvent avoir des moyennes et médianes très proches, tout en étant constituées des données très différentes. Pour les comparer, on calcule deux **caractéristiques de dispersion** :

a. Etendue :

C'est la différence entre le Maximum et le Minimum de la série.

Dans l'exemple du **II.c.** l'étendue est : 35 – 21 = 14 ans

b. Les quartiles

Ce sont les valeurs qui partagent la série en quatre séries d'effectif égal. En fait, on ne détermine que le 1er et 3ème quartile, puisque le 2ème quartile est la médiane.

Le 1er quartile est **le plus petit nombre Q_1** tel que **25%** des données sont inférieures ou égales à Q_1.

Le 3ème quartile est **le plus petit nombre Q_3** tel que **75%** des données sont inférieures ou égales à Q_3.

Dans l'exemple précédent :

1er quartile : 25% de 22 = $22 \times \dfrac{25}{100} = 22 \times \dfrac{1}{4} = 5{,}5$ → 6ème valeur donc Q_1 = 26

3ème quartile : 75% de 22 = $22 \times \dfrac{75}{100} = 22 \times \dfrac{3}{4} = 16{,}5$ → 17ème valeur donc Q_3 = 31

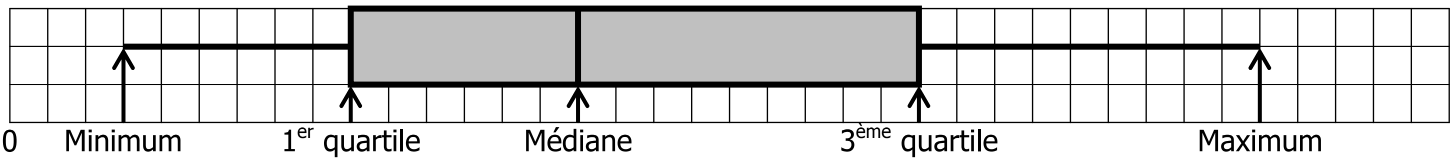
c. L'écart interquartile :

C'est tout simplement la différence $Q_3 - Q_1$.

Dans l'exemple précédent, $Q_3 - Q_1$ = 31 – 26 = 5 ans

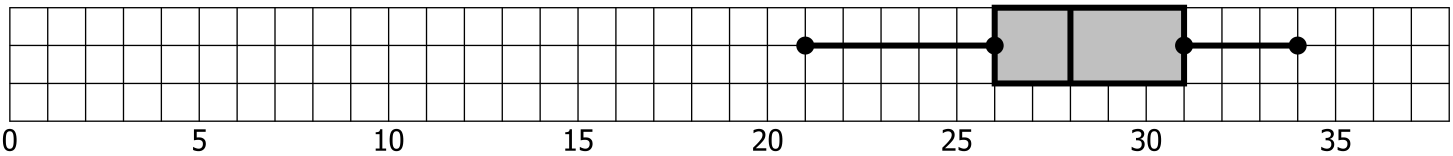
d. Le diagramme « boite à moustache » :

C'est un diagramme qui résume les caractéristiques de position (médiane, quartiles, extremums), sous la forme suivante :



Ce diagramme est principalement utilisé pour comparer un même caractère dans deux populations de tailles différentes.

Le diagramme correspondant à l'exemple précédent est :



e. variance et écart-type :

Formules : $V(x) = \frac{1}{N}\sum_{i=1}^{n} n_i \left(x_i - \overline{x}\right)^2 = \frac{1}{N}\sum_{i=1}^{n} n_i x_i^2 - \left(\overline{x}\right)^2$ et $\sigma_x = \sqrt{V(x)}$

Plus facile à retenir, la variance est « *la différence entre la moyenne des carrés et le carré de la moyenne* » obtenue

Exemple

Voici les 8 notes (sur 20) d'un élève ce trimestre :

									Total
Notes	12	15	12	13	10	19	11	7	99
(Notes)²	144	225	144	169	100	361	121	49	1313

Moyenne des notes : $\frac{12+15+12+13+10+19+11+7}{8} = \frac{99}{8} = 12{,}375$

donc (Moyenne des notes)^2 : $12{,}375^2$ = 153,140625

Moyenne des (notes)^2 : $\frac{144+225+...+49}{8} = \frac{1313}{8} = 164{,}125$

La variance est donc : 164,125 – 153,140625 $\approx$ **10,984**

D'après l'exemple précédent : $\sigma \approx \sqrt{10{,}984} \approx$ **3,314**

L'écart-type peut être interprété comme « l'écart moyen par rapport à la moyenne d'une série de valeurs ».

Plus il est grand, plus les valeurs sont dispersées.

Plus il est petit, plus les valeurs sont centrées autour de la moyenne.