Chapitre 7 : Statistiques descriptives

I- Population, individus, caractère, effectif , fréquence

Faire une étude statistique, c'est étudier un certain **caractère** dans une **population** donnée. On relève pour cela les valeurs du caractère par individu. L'ensemble des données ainsi récoltées constitue les données brutes qu'il est nécessaire de trier et de regrouper avant d'en faire l'étude. A noter que le caractère étudié peut être **qualitatif** (couleur d'une voiture) ou **quantitatif** (taille d'un individu)

Exemple : La population étudiée est la classe de seconde 1. Les individus sont les élèves.

- l'effectif total est 29
- on peut étudier les caractères suivant :
  - la couleur des yeux : caractère qualitatif
  - la taille ou le nombre de frères et sœurs : caractère quantitatif

Effectif et fréquence

- L'**effectif** d'une valeur du caractère est le nombre d'individus de la population correspondant à cette valeur
- La **fréquence** d'une valeur du caractère est le quotient de l'effectif de cette valeur par l'effectif total :

$$f = \frac{\text{effectif de la valeur}}{\text{effectif total}}$$

Dans l'exemple d'introduction sur les paquets de café, le tableau des effectifs est :

Masse (en grammes)	249,6	249,7	249,8	249,9	250	250,1	250,2	250,3
Effectif	2	2	2	5	4	5	3	1

II- Moyenne pondérée

Définition : On considère la série statistique suivante :

Valeur du caractère	x_1	x_2	x_3	...	x_p
Effectif	n_1	n_2	n_3	...	n_p

L'effectif total est N = $n_1+n_2+\ldots+n_p$

La moyenne de la série est : $\bar{x} = \dfrac{n_1\times x_1+n_2\times x_2+\ldots+n_p\times x_p}{N}$

Remarques :

1) On peut aussi calculer la moyenne à l'aide des fréquences : $\bar{x} = f_1 x_1+x_2 f_2+\ldots+f_p x_p$

2) La moyenne ainsi calculée est dite **pondérée** car on affecte à chaque valeur $x_1, x_2, \ldots, x_p$ son poids $n_1, n_2, \ldots, n_p$ dans l'étude

Exemple

Masse (en grammes)	249,6	249,7	249,8	249,9	250	250,1	250,2	250,3
Effectif	2	2	2	5	4	5	3	1

Pour le café, la masse moyenne d'un paquet de l'échantillon est donnée par :

$$\bar{x} = \frac{2\times 249{,}6+3\times 249{,}7+\ldots+1\times 250{,}5}{2+3+\ldots+1} = \frac{5999{,}1}{24} = 249{,}9625$$

III- Indicateur de dispersion

a) L'écart type

Définition : L'écart type d'une série statistique , notée σ , permet de mesurer la dispersion d'une série statistique autour de la moyenne :

- **plus l'écart type est petit** , plus les valeurs de la série sont concentrées autour de la moyenne donc la série peut être qualifiée d'homogène
- **plus l'écart type est grand** , plus les valeurs de la série sont éloignées de la moyenne donc la série est moins homogène

Remarque : La plupart du temps, l'écart type sera calculé à la calculatrice mais il existe des formules pour le calculer :

- pour une série $x_1, x_2, \ldots, x_p$ de moyenne m , l'écart type est :

$$\sigma = \sqrt{\frac{(x_1-m)^2+(x_2-m)^2+\ldots+(x_p-m)^2}{N}}$$

- Si chaque valeur $x_1, x_2, \ldots, x_p$ a pour coefficient $n_1, n_2, \ldots, n_p$, l'écart type est :

$$\sigma = \sqrt{\frac{n_1(x_1-m)^2+n_2(x_2-m)^2+\ldots+n_p(x_p-m)^2}{N}}$$

b) Etendue , Médiane et quartiles

Définitions On considère une série statistique ORDONNEE

1) l'**étendue** de la série correspond à la différence entre la plus grande et la plus petite des valeurs prises par le caractère
2) La **médiane Me** partage la série en deux groupes de même effectif
3) Le **quartile 1** notée **Q1** est la plus petite valeur de la série telle **qu'au moins le quart** de l'effectif est inférieur ou égal à Q1
4) Le **quartile 3** notée **Q3** est la plus petite valeur de la série telle **qu'au moins les trois quarts** de l'effectif est inférieur ou égal à Q3

On veut déterminer la médiane et les quartiles des résultats de la première composition des seconde III
Exemple Voici les résultats de cette première composition

11	6	5	7	10	7	5	9	8	14	3	18	12	13	11
8	5	7	9	6	11	12	11	6	9					

La 1ère étape consiste à classer ces notes dans l'**ordre croissant**

Notes												
effectif												

1) L'étendue de cette série est E = Valeur max – Valeur min =

2) **Détermination de la médiane**

On tient compte de l'effectif total N = 25 • • • • •

- **L'effectif est impair.** $\frac{25}{2}$ = 12,5 La médiane est la note donc Me =

Si on retire la note 18 à cette série, l'effectif total est alors 24 • • • •

- **L'effectif est alors pair.** $\frac{24}{2}$ = 12 . La médiane est entre la 12ème et la 13ème note . Me =

2) Détermination des quartiles

N = 25 et $\frac{N}{4} = \frac{25}{4} = 6{,}25$. Le premier quartile est la note c'est à dire Q1 =

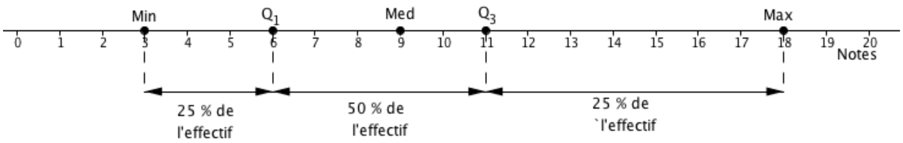
$\frac{3N}{4} = \frac{75}{4} = 18{,}75$. Le troisième quartile est la note c'est à dire Q3 =

3) Une méthode plus rapide

Si l'on dispose du tableau des effectifs cumulés croissants, la lecture de ces différentes valeurs est immédiate

Note	3	5	6	7	8	**9**	10	11	12	13	14	18
Effectif												
Effectif cumulés croissants												

Remarque : On peut visualiser ces paramètres à l'aide du schéma ci-dessous :



> **Définition :**
> On appelle **intervalle interquartile** l'intervalle $[Q_1 ; Q_3]$
> On appelle **écart interquartile** la quantité $Q_3 - Q_1$

Remarques :

L'intervalle interquartile contient donc 50 % de l'effectif.

L'écart interquartile est une mesure de dispersion des données autour de la médiane :

- plus il est grand, plus les données sont dispersées autour de la médiane
- plus il est petit, plus les données sont proches de la médiane

V- Une propriété de la moyenne : la linéarité de la moyenne

> **Définition :** Soit a et b deux réels. On considère une série statistique $x_1, x_2, ..., x_p$ de moyenne m.
> - Si on multiplie par a toutes les valeurs de la série, alors la moyenne M de la nouvelle série est obtenue en multipliant par a la moyenne de la série de départ : M = $a \times m$
> - Si on ajoute b à toutes les valeurs de la série, alors la moyenne M de la nouvelle série est obtenue en ajoutant b à la moyenne de la série de départ M = $m+b$
> - Si on effectue les deux transformations précédentes , la moyenne de la série $ax_1+b, ax_2+b, ..., ax_p+b$ est am+b