

# Chapitre 01 : Statistique descriptive

## 1.1 Introduction :

La statistique descriptive est un ensemble de méthodes permettant de d'écrire et d'analyser des phénomènes susceptibles d'être dénombrés et classés. Elle a pour but de d'écrire et non d'expliquer.

### 1.1.1 Concepts de base :

- **Population** : On appelle population l'ensemble étudié c'est-à-dire l'ensemble sur lequel porte notre étude statistique.
- **Individu ou unité statistique** : On appelle individu tout élément de la population.
- **Echantillon** : est un sous ensemble de la population étudiée. Nombre d'individus dans un échantillon noté  $n$  est appelé taille de l'échantillon.
- **Caractère (Variable statistique)** : propriété commune aux individus de la population, que l'on veut étudier.

**Remarque 1.1** : Soit  $\Omega$  un ensemble. On appelle et on note  $Card(\Omega)$ , le nombre d'éléments de  $\Omega$ .

$$Card(\Omega) := \text{nombre d'éléments de } \Omega := N.$$

- **Modalités** : Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci.

**Exemple 1.1.** Nous résumons les différents concepts dans cet exemple :

**Population** : l'ensemble des tous les employés d'une usine.

**Individu** : chaque employé de l'usine.

**Caractère** : le salaire, l'état matrimonial, le nombre d'enfants, ... etc.

**Les modalités du caractère** : marié, célibataire, divorcé et veuf sont les modalités de l'état matrimonial, par exemple.

### 1.1.2 Types de caractères :

On distingue deux types de caractères : **qualitatif et quantitatif**.

#### Caractère qualitatif

**Définition 1.1.** Un caractère est dit qualitatif lorsque ses modalités ne sont pas mesurables.

**Exemple 1.2.** Les couleurs du pelage : l'ensemble des modalités est  
 $\{\text{noir, marron, blanc, ...}\}$

#### Caractère quantitatif

**Définition 1.2.** Un caractère est dit quantitatif lorsque ses modalités sont des nombres. On lui donne souvent le nom de **variable statistique**.

Une variable statistique peut être :

- ▶ **Discrète** : Si elle prend des valeurs isolées.

**Exemple 1.3.** Le nombre d'enfants d'une famille.

► **Continue :** Lorsqu'elle peut prendre n'importe quelle valeur dans son domaine de variation.

**Remarque 1.2.** Dans le cas continu, le nombre de ces valeurs est toujours très grand. Dans ce cas, on regroupe toutes ces valeurs en classes.

En général, toutes les grandeurs liées à l'espace (longueur, surface, volume, ...), au temps (âge), à la masse (poids, teneur, ...) ou à des combinaisons (vitesse, débit, ...) sont des variables statistiques continues.

## 1.2 Tableaux statistiques et représentations graphiques :

Soit une population composée de  $n$  individus, sur laquelle on a étudié un caractère possédant  $k$  valeurs possibles. Ces valeurs  $x_1, x_2, \dots, x_k$  sont des modalités (cas qualitatif) ou des nombres (cas quantitatif).

Soient :

$n_1$  le nombre d'individus ayant pris la valeur  $x_1$

$n_2$  le nombre d'individus ayant pris la valeur  $x_2$

⋮

$n_k$  le nombre d'individus ayant pris la valeur  $x_k$

$n_i$  est appelé **fréquence** ou **effectif** de la valeur  $x_i$  et  $n$  est l'effectif total.

On appelle **fréquence relative** ou **effectif relatif** de la valeur  $x_i$  la quantité :

$$f_i = \frac{n_i}{n} \quad (\text{ou en \% } f_i = \frac{n_i}{n} \times 100\%)$$

C'est la proportion d'individus ayant pris la valeur  $x_i$ .

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = n$$

$$\sum_{i=1}^k f_i = f_1 + f_2 + \dots + f_k = \frac{n_1}{n} + \frac{n_2}{n} + \dots + \frac{n_k}{n} = 1$$

Modalités ( $x_i$ )	Effectifs ( $n_i$ )	Fréquences relatives ( $f_i = \frac{n_i}{n}$ )
$x_1$	$n_1$	$f_1 = \frac{n_1}{n}$
$x_2$	$n_2$	$f_2 = \frac{n_2}{n}$
⋮	⋮	⋮
$x_k$	$n_k$	$f_k = \frac{n_k}{n}$
Total	$\sum_{i=1}^k n_i = n$	$\sum_{i=1}^k f_i = 1$

Tableau des effectifs et des fréquences relatives

### 1.2.1 Tableau statistique relatif à un caractère qualitatif et sa représentation graphique :

Nous allons utiliser souvent l'exemple ci-dessous pour illustrer les énoncés de cette section.

**Exemple 1.4.** On veut étudier les lois de Mendel sur le caractère couleur de la fleur de Balsamine. Pour cela on étudiera le croisement des plantes hétérozygotes. On obtient quatre couleurs : pourpre, rose, blanc-lavande et blanche.

**Population :** les plantes de Balsamine.

**Individu :** une plante.

**Caractère étudié :** couleur de la fleur.

#### I) Tableau statistique

On définit les effectifs cumulés  $N_i$  par la somme des effectifs correspondants aux valeurs de la variable statistique inférieure à  $x_{i+1}$ .

C'est-à-dire :

$$N_1 = n_1$$

$$N_2 = n_1 + n_2$$

$$N_3 = n_1 + n_2 + n_3$$

⋮

$$N_i = n_1 + n_2 + n_3 + \dots + n_i$$

De la même façon on définit la fréquence relative cumulée  $F_i$  par :  $F_i = \sum_{j=1}^i f_j$ ,

d'où :  $F_i = \frac{N_i}{n}$ .

Modalités $x_i$	Effectif $n_i$	Fréquences relatives $f_i$	$f_i$ en %	Effectifs cumulés $N_i$	Fréquences relatives cumulées $F_i$
Pourpre	1790	0.5778	57.78%	1790	0.5778
Rose	547	0.1766	17.66%	2337	0.7544
Blanc-Lavande	548	0.1769	17.69%	2885	0.9313
Blanche	213	0.0688	6.88%	3098	1
Total	3098	$\sum_{i=1}^4 f_i = 1$	100%		

#### II) Représentation graphique

L'information résumée dans un tableau statistique se traduit par un graphique pour en réaliser une synthèse visuelle.

### a) Représentation par tuyaux d'orgue (diagramme en colonnes)

Dans ce cas, le graphe s'obtient en construisant autant de colonnes que des modalités du caractère qualitatif. Ces colonnes sont des rectangles de bases constantes et de hauteurs proportionnelles aux fréquences relatives.

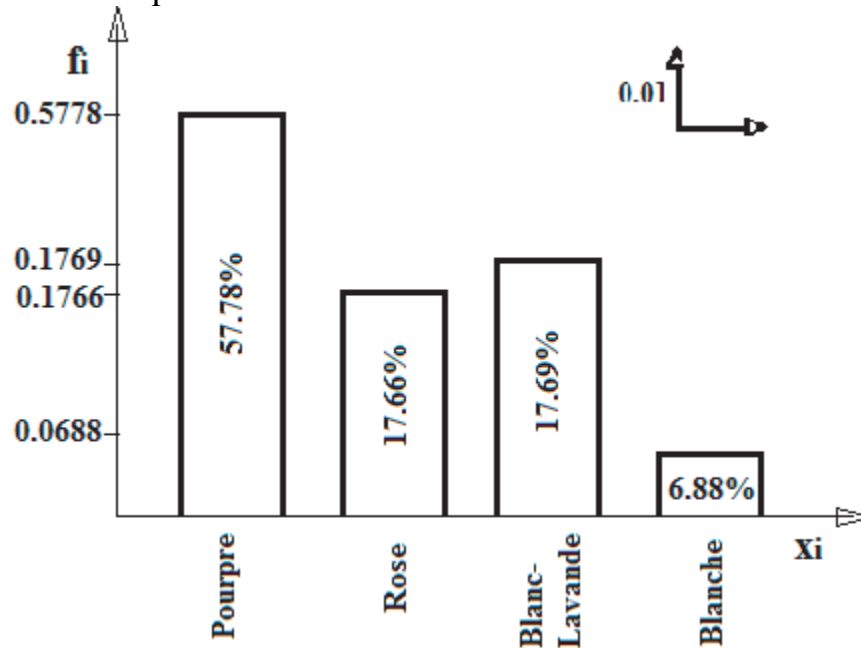


Figure 1.1 : Représentation en tuyaux d'orgue des fréquences relatives.

### b) Représentation par le diagramme circulaire (camembert)

$$\alpha_i = f_i \times 360^\circ = \frac{n_i}{n} \times 360^\circ$$

Les angles correspondant de l'exemple sont :

$$\alpha_1 = 0.5778 \times 360^\circ = 208.01^\circ \rightarrow \textit{Pourpre}$$

$$\alpha_2 = 0.1766 \times 360^\circ = 63.58^\circ \rightarrow \textit{Rose}$$

$$\alpha_3 = 0.1769 \times 360^\circ = 63.68^\circ \rightarrow \textit{Blanc - Lavande}$$

$$\alpha_4 = 0.0688 \times 360^\circ = 24.77^\circ \rightarrow \textit{Blanche}$$

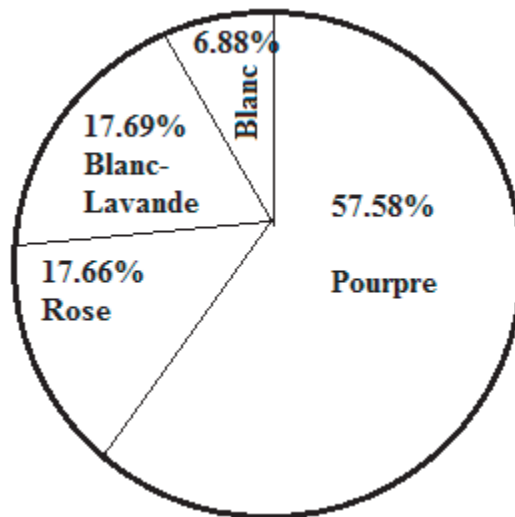


Figure 1.2 – Diagramme en camembert des fréquences relatives.

## 1.2.2 Tableaux statistiques relatifs à un caractère quantitatif et représentations graphiques :

### 1) Cas d'une variable statistique discrète :

**Exemple 1.5.** Lors d'un contrôle d'une chaîne de médicaments, on s'intéresse au nombre de comprimés défectueux dans un lot. L'étude de 200 lots a donné les résultats suivants :

- ▶ 75 lots ont 0 comprimés défectueux ;
- ▶ 53 lots ont 1 comprimé défectueux ;
- ▶ 39 lots ont 2 comprimés défectueux ;
- ▶ 23 lots ont 3 comprimés défectueux ;
- ▶ 9 lots ont 4 comprimés défectueux ;
- ▶ 1 lot a 5 comprimés défectueux.

**Population :** l'ensemble des lots des médicaments.

**Individu :** un lot.

**Caractère étudié :** nombre de comprimés défectueux.

**Modalités :** 0, 1, 2, 3, 4 et 5.

#### I) Tableau statistique

Modalités $x_i$	Effectif $n_i$	Fréquence relative $f_i$	$f_i$ en %	$N_i \nearrow$	$F_i \nearrow$
0	75	0.375	37.5%	75	0.375
1	53	0.265	26.5%	128	0.64
2	39	0.195	19.5%	167	0.835
3	23	0.115	11.5%	190	0.95
4	9	0.045	4.5%	199	0.995
5	1	0.005	0.5%	200	1
<b>Total</b>	<b>200</b>	<b>1</b>	<b>100%</b>		

#### II) Représentation graphique

a) Diagramme en bâtons (Polygone des fréquences relatives)

On utilise le diagramme en bâtons pour représenter les effectifs  $n_i$  et les fréquences relatives  $f_i$ . À chaque marque correspond un bâton. Les hauteurs des bâtons sont proportionnelles aux effectifs (aux fréquences relatives) représentés.

Dans le cas du graphe des fréquences relatives, en joignant les sommets des bâtons on obtient le polygone des fréquences relatives.

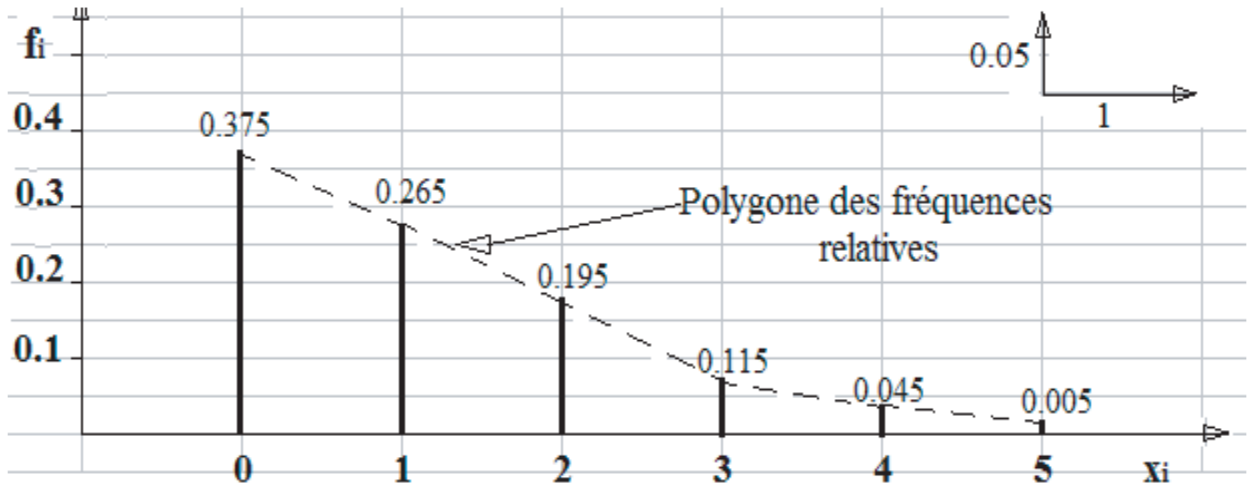


Figure 1.3 – Diagramme en escalier des fréquences relatives.

b) Courbe des fréquences relatives cumulées (courbe cumulative).

**Remarque 1.3.**

$$\sum_{i=1}^k f_i = 1$$

$$F_i = 0 \text{ si } x_i < x_1$$

$$F_i = 1 \text{ si } x_i < x_k$$

où  $x_1$  est la plus petite valeur observées et  $x_k$  est la plus grande valeur observée.

Soit  $X$  une variable statistique discrète et  $x_1, x_2, \dots, x_k$  les valeurs rangées dans l'ordre croissant. La fonction de répartition d'une v.s. discrète est définie de  $R$  dans  $[0; 1]$  et est donnée par :

$$F(x) = \begin{cases} 0, & \text{si } x < x_1; \\ f_1, & \text{si } x_1 \leq x < x_2; \\ f_1 + f_2, & \text{si } x_2 \leq x < x_3; \\ \vdots & \\ f_1 + f_2 + \dots + f_i, & \text{si } x_i \leq x < x_{i+1}; \\ \vdots & \\ 1, & \text{si } x \geq x_k. \end{cases}$$

Ecrivons la fonction de répartition de la variable statistique  $X$  de l'exemple des comprimés défectueux (l'exemple 1.5) :

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 0.375, & \text{si } 0 \leq x < 1; \\ 0.64, & \text{si } 1 \leq x < 2; \\ 0.835, & \text{si } 2 \leq x < 3; \\ 0.95, & \text{si } 3 \leq x < 4; \\ 0.995, & \text{si } 4 \leq x < 5; \\ 1, & \text{si } x \geq 5. \end{cases}$$

La courbe cumulative est la représentation graphique des fréquences relatives cumulées. Dans le cas discret, la courbe cumulative est une courbe en escalier, dont les paliers horizontaux ont pour coordonnées  $(x_i, F_i)$ .

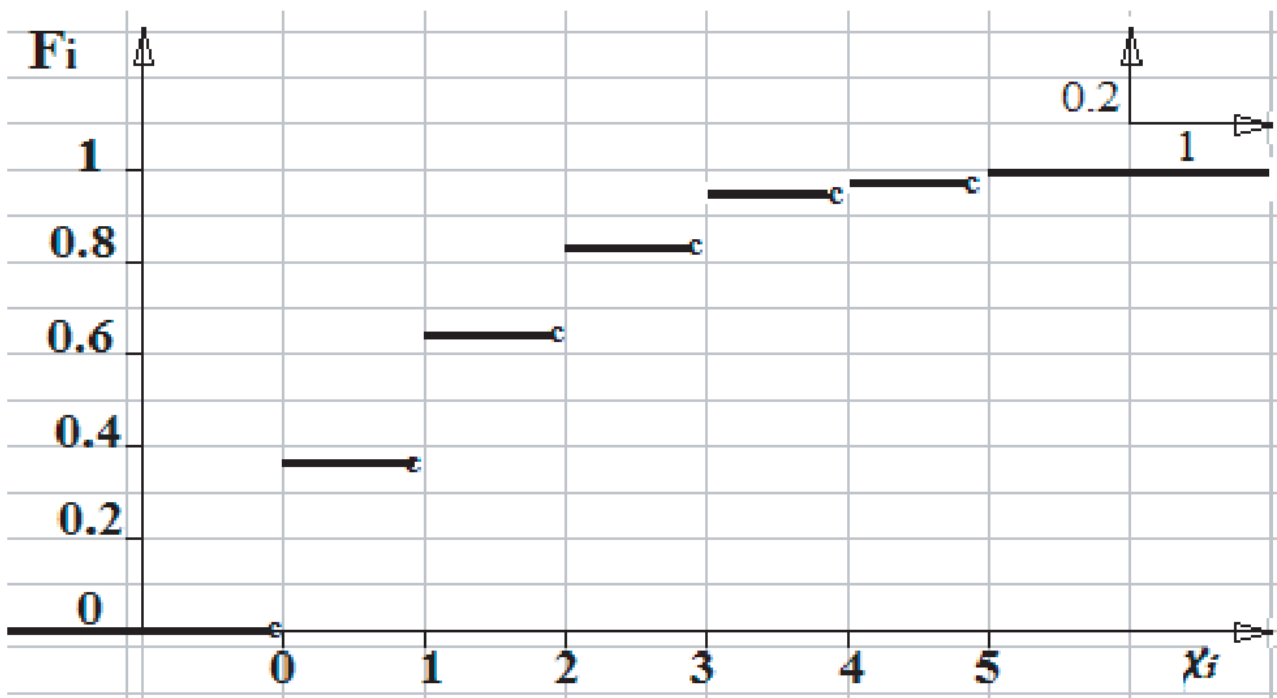


Figure 1.7 – Courbe des fréquences relatives cumulées (courbe cumulative).

### III) Paramètre de Position (caractéristique de tendance centrale) :

Les indicateurs statistiques de tendance centrale (dits aussi de position) considérés fréquemment sont **la moyenne, la médiane et le mode**.

#### • La moyenne arithmétique :

Soient  $X$  une variable statistique discrète et  $x_1, x_2, \dots, x_k$  ses valeurs, pour lesquelles correspondent les effectifs  $n_1, n_2, \dots, n_k$  ; avec  $n = \sum_{i=1}^k n_i$  l'effectif total.

La moyenne arithmétique notée  $\bar{X}$  de cette série statistique, est définie par :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i x_i$$



**Remarque 1.4.**

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} \\ &= \frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \dots + \frac{n_k}{n} x_k \\ &= f_1 x_1 + f_2 x_2 + \dots + f_k x_k \\ &= \sum_{i=1}^k f_i x_i,\end{aligned}$$

où  $f_i$  est la fréquence relative.

Dans notre exemple, la moyenne arithmétique de 200 lots de comprimés est :

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^6 n_i x_i \\ &= \frac{75 \times 0 + 53 \times 1 + 39 \times 2 + 23 \times 3 + 9 \times 4 + 1 \times 5}{200} \\ &= \frac{241}{200} = 1.205\end{aligned}$$

● **La médiane :**

La médiane est la valeur pour laquelle, on a le même nombre d'individus à gauche et à droite dans un échantillon. Elle correspond au milieu de la distribution.

Pour déterminer la médiane d'un échantillon ou d'une population :

1. on classe les individus par ordre croissant ;
2. on prend celui du milieu.

**Remarque 1.5.** La médiane  $Me$  est la valeur qui se trouve au centre de la série statistique :

► Si la valeur de l'effectif total  $n$  est impaire, alors la médiane  $Me$  est la valeur qui se trouve à l'ordre  $\frac{n+1}{2}$  :

$$Me = x_{\frac{n+1}{2}}$$

► Si la valeur de l'effectif total  $n$  est paire, alors la médiane  $Me$  est la moyenne des valeurs qui se trouve à l'ordre  $\frac{n}{2}$  et  $\frac{n}{2} + 1$  :

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

**Exemple 1.13 :** Soit un échantillon de 11 personnes dont le poids en kg est :

45,48,89,74,55,62,56,68,49,52,63

Les poids classés par ordre croissant sont :

45,48,49,52,55 , **56** , 62,63,68,74,89

La médiane est  $Me = 56$ .

Si le nombre d'individus est pair,  $n = 12$ , on prend la moyenne entre les deux valeurs centrales.

45,48,49,52,55 , 56, 62 , 63,68,74,89,91

La médiane est

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{\frac{12}{2}} + x_{\frac{12}{2}+1}}{2} = \frac{x_6 + x_7}{2} = \frac{56 + 62}{2} = 59$$

**Remarque 1.6.** Pour les ensembles de données plus grand, le calcul de la médiane se fait à partir des effectifs ou des fréquences cumulées. La médiane est la valeur de la variable à laquelle est associé un effectif cumulé égal à  $\frac{n}{2}$ , ou une fréquence cumulée égale à 0.5 ( $n$  étant effectif total de la population).

Si ces valeurs d'un effectif cumulé égal à  $\frac{n}{2}$  ou d'une fréquence cumulée égale à 0.5 ne correspondent pas exactement à une valeur de la variable et tombent entre de lignes de la distribution, aucune valeur possible de la variable ne partage alors exactement la population en deux sous-ensembles égaux. Par convention, on retient dans ce cas comme médiane **la valeur de la variable immédiatement supérieure**.

• **Le mode :**

Le mode d'une variable statistique est la valeur qui a le plus grand effectif partiel (ou la plus grande fréquence partielle) et il est dénoté par **Mo**.

**Remarque 1.7.** On peut avoir plus d'un mode ou rien.

**Exemple 1.15.**

1. Dans l'exemple des comprimés défectueux, le mode est 0.

2. Dans l'exemple des observations suivantes : 2, 2, 3, 5, 6, 6, 7, 8, 9, 7, 10 ; il y a trois modes : 2, 6 et 7.

**IV) Paramètre de Dispersion :**

• **Etendue :**

On appelle étendue, notée  $e$ , la différence entre la plus grande valeur et la plus petite valeur observée.

**Exemple 1.16.**

Soit  $X$  l'ensemble d'observation suivant :

$$X = \{5, 8, 12, 12, 14, 16, 20, 20\}$$

L'étendue de  $X$  est  $e = 20 - 5 = 15$ .

• **Variance :**

La variance notée  $V(X)$  est la moyenne des carrés des écarts des observations à la moyenne. Elle est définie par :

$$V(x) = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2$$

**Propriétés :**

1)  $V(x) \geq 0$

2)  $V(x) = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{X}^2 = \sum_{i=1}^k f_i x_i^2 - \bar{X}^2$

• **Ecart type :**

L'écart type noté  $\sigma_x$  est la racine carrée de  $V(x)$  :

$$\sigma_x = \sqrt{V(x)}$$

**Remarque 1.8.** Lorsqu'on compare les observations de deux variables statistiques  $X$  et  $Y$ , celle qui possède l'écart-type le plus élevé est la plus **dispersée**.

• **Les quartiles :**

On utilise couramment les quartiles  $Q_1, Q_2$  et  $Q_3$ .

$Q_1$  est le quartile d'ordre  $\frac{1}{4}$ , représente 25% de l'échantillon ;

$Q_2$  est le quartile d'ordre  $\frac{2}{4}$ , représente 50% de l'échantillon ;

$Q_3$  est le quartile d'ordre  $\frac{3}{4}$ , représente 75% de l'échantillon.

► **Le premier quartile  $Q_1$**

$Q_1$  est la valeur  $x_i$  dont le rang (la position) est le plus petit entier qui suit  $\frac{n}{4}$ .

**Exemple 1.17.** Dans l'exemple des observations suivantes : 2, 3, 4, 5, 6, 6, 7, 7, 8, 9, 10 ; on a :

$n = 11$  et  $\frac{n}{4} = \frac{11}{4} = 2.25$ . Le plus petit entier qui suit  $\frac{n}{4} = 2.25$  est 3, alors  $Q_1$  est la troisième valeur. D'où  $Q_1 = x_3 = 4$ .

**Exemple 1.18.** Reprenons l'exemple des comprimés défectueux, où on a  $\frac{n}{4} = \frac{200}{4} = 50$  et la valeur pour laquelle les effectifs cumulés atteignent ou dépassent pour la première fois 50 est "moins 1", c'est-à-dire la valeur de 0, alors  $Q_1 = 0$ .

► **Le troisième quartile  $Q_3$**

$Q_3$  est la valeur  $x_i$  dont le rang (la position) est le plus petit entier qui suit  $\frac{3n}{4}$ .

**Exemple 1.19.** Dans le même exemple des observations : 2, 3, 4, 5, 6, 6, 7, 7, 8, 9, 10 ; on a :

$n = 11$  et  $\frac{3n}{4} = \frac{3 \times 11}{4} = 8.25$ . Le plus petit entier qui suit  $\frac{3n}{4} = 8.25$  est 9, alors  $Q_3$  est la 9<sup>ème</sup> valeur. D'où  $Q_3 = x_9 = 8$ .

**Exemple 1.20.** Dans l'exemple des comprimés défectueux, on a  $\frac{3n}{4} = \frac{3 \times 200}{4} = 150$  et la valeur où les effectifs cumulés atteignent ou dépassent pour la première fois 150 est "moins 3", c'est-à-dire la valeur de 2, alors  $Q_3 = 2$ .

**2) Cas d'une variable statistique continue :**

Lorsque la variable statistique est continue les données sont regroupées en classes

$$[e_0, e_1[, [e_1, e_2[, [e_2, e_3[, \dots, [e_{k-1}, e_k[$$

Les modalités  $x_i$  représentent les centres  $c_i$  des classes  $[e_{i-1}, e_i[$ , avec :

►  $c_i = \frac{e_{i-1} + e_i}{2}, i \in \{1, 2, \dots, k\}$  ;

►  $e_{i-1}$  est appelé l'extrémité inférieure de la classe  $[e_{i-1}, e_i[$  ;

►  $e_i$  est appelé l'extrémité supérieure de la classe  $[e_{i-1}, e_i[$  ;

►  $a_i = e_i - e_{i-1}$  est l'amplitude de la classe  $[e_{i-1}, e_i[$  ;

►  $e_k - e_0$  est appelé l'étendue de la variable statistique.

**Exemple 1.21.** Une étude faite sur la taille d'un groupe d'étudiants (en mètre) a donné les résultats suivants :

## I) Tableau statistique

Classes	Centre $C_i$	Effectif $n_i$	Fréq. Rel $f_i$	$f_i$ en %	$a_i$	Effectif. Cum $N_i \nearrow$	Fréq. Cum $F_i \nearrow$
[1,5; 1,6[	1.55	8	0.08	8%	0.1	8	0.08
[1,6; 1,7[	1.65	33	0.33	33%	0.1	41	0.41
[1,7; 1,8[	1.75	31	0.31	31%	0.1	72	0.72
[1,8; 1,9[	1.85	22	0.22	22%	0.1	94	0.94
[1,9; 2[	1.95	6	0.06	6%	0.1	100	1
Total		n=100	1	100%			

**Population** : les étudiants du groupe.

**Individu** : un étudiant.

**Caractère étudié** : la taille d'un étudiant.

**Type** : quantitatif continu.

**Modalités** : [1,5; 1,6[ ; [1,6; 1,7[ ; ... ; [1,9; 2[.

## II) Représentation graphique

### a) Histogramme

Dans le cas d'une variable statistique continue on utilise l'histogramme.

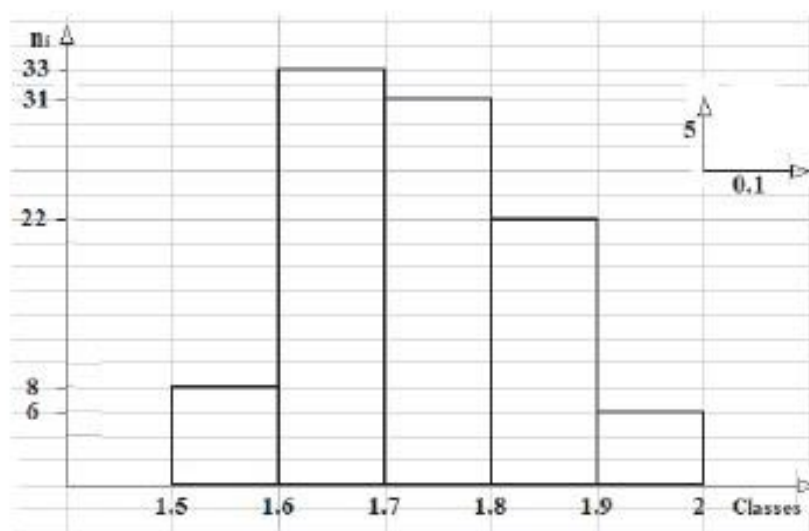


Figure 1.8- L'histogramme des effectifs de l'exemple 1.21.

### b) Courbe cumulative

La courbe cumulative est donnée dans la figure suivante :

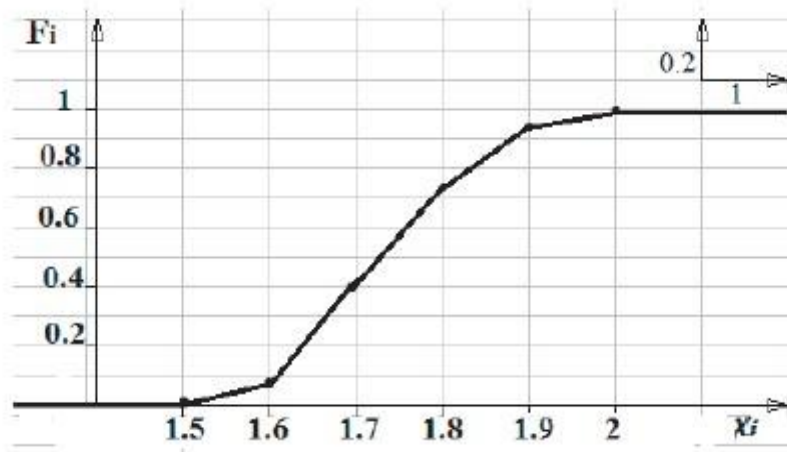


Figure 1.9- Courbe des fréquences relatives cumulées du cas continu (exemple 1.21)

### III) Paramètre de Position (caractéristique de tendance centrale) :

#### ● Moyenne arithmétique :

Dans le cas d'une variable statistique continue, les observations sont groupées dans des classes, et nous avons la même formule que le cas discret, sauf qu'on remplace les  $x_i$  par les centres des classes  $c_i$  :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^k n_i c_i = \sum_{i=1}^k f_i c_i$$

**Exemple 1.22.** La moyenne arithmétique de la taille des étudiants de l'exemple 1.21 est :

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^k n_i c_i \\ &= \frac{(8 \times 1.55) + (33 \times 1.65) + (31 \times 1.75) + (22 \times 1.85) + (6 \times 1.95)}{100} \\ &= \frac{173,5}{100} = 1.735 \end{aligned}$$

#### ● La médiane :

Dans ce cas la médiane est donnée par la formule suivante :

$$Me = L_i + \left( \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{Me}} \right) \times a$$

- ▶  $L_i$  : borne inférieure de la classe médiane (classe qui divise l'effectif en deux) ;
- ▶  $n$  : effectif total ;
- ▶  $\sum_{i=1}^{<Me} n_i$  : somme des effectifs correspondant à toutes les classes inférieures à la classe médiane ;
- ▶  $n_{Me}$  : effectif de la classe médiane ;
- ▶  $a$  : amplitude de la classe médiane.

**Exemple 1.23.**

Prenons l'exemple 1.21. La classe médiane correspond à la classe où les effectifs cumulés atteignent ou dépassent pour la première fois le 50%.

Dans notre exemple, la classe médiane est la classe  $[1,7 ; 1,8[$ .

- ▶  $L_i = 1.7$  ;
- ▶  $n = 100$  ;
- ▶  $\sum_{i=1}^{<Me} n_i = 33 + 8$  ;
- ▶  $n_{Me} = 31$  ;
- ▶  $a = 0.1$  ;

$$Me = L_i + \left( \frac{\frac{n}{2} - \sum_{i=1}^{<Me} n_i}{n_{Me}} \right) \times a = 1.7 + \left( \frac{\frac{100}{2} - (33 + 8)}{31} \right) \times 0.1 = 1.7290.$$

• **Le mode :**

Dans le cas des données groupées en classes, le mode se calcule par la formule :

$$Mo = L_i + \left( \frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot a$$

- ▶  $L_i$  : borne inférieure de la classe modale (classe correspondant au plus grand effectif) ;
- ▶  $\Delta_1$  : excédent de l'effectif de la classe modale par rapport à l'effectif de la classe précédente ;
- ▶  $\Delta_2$  : excédent de l'effectif de la classe modale par rapport à l'effectif de la classe suivante ;
- ▶  $a$  : amplitude de la classe modale.

**Exemple 1.24.** Prenons les données de l'exemple du cas quantitatif continu (exemple 1.21) et calculons le mode.

- ▶ La classe modale est  $[1,6 ; 1,7[$  ;
- ▶  $L_i = 1.6$  ;
- ▶  $\Delta_1 = 33 - 8$  ;
- ▶  $\Delta_2 = 33 - 31$  ;
- ▶  $a = 0.1$ .

Le mode est

$$Mo = L_i + \left( \frac{\Delta_1}{\Delta_1 + \Delta_2} \right) \cdot a = 1.6 + \left( \frac{33 - 8}{(33 - 8) + (33 - 31)} \right) \times 0.1 = 1.6926 \text{ mètre.}$$

#### IV) Paramètre de Dispersion :

- **Etendue :**

On appelle étendue, notée  $e$ , la différence entre la plus grande valeur et la plus petite valeur observée.

Dans l'exemple 1.21 l'étendue est égale à :

$$e = 2 - 1.5 = 0.5$$

- **Variance :**

La variance est la quantité :

$$V(X) = \frac{1}{n} \sum_{i=1}^k n_i (c_i - \bar{x})^2 = \sum_{i=1}^k f_i (c_i - \bar{x})^2$$

Pour le calcul, on utilise

$$V(X) = \frac{1}{n} \sum_{i=1}^k n_i c_i^2 - \bar{x}^2 = \sum_{i=1}^k f_i c_i^2 - \bar{x}^2.$$

- **Ecart type :**

L'écart-type noté  $\sigma_x$  est la racine carrée de  $V(X)$  :

$$\sigma_x = \sqrt{V(X)}.$$

#### **Exemple 1.25.**

La variance de l'exemple 1.21 est :

$$\begin{aligned} V(X) &= \frac{1}{n} \sum_{i=1}^k n_i c_i^2 - \bar{x}^2 \\ &= \frac{(8 \times 1.55^2) + (33 \times 1.65^2) + (31 \times 1.75^2) + (22 \times 1.85^2) + (6 \times 1.95^2)}{100} - (1.735)^2 \\ &= 0.010875 \end{aligned}$$

L'écart-type  $\sigma_x = \sqrt{V(X)} = \sqrt{0.010875} = 0.1043$ .

- **Les quartiles :**

- ▶ **Le premier quartile  $Q_1$**

Dans ce cas le premier quartile est donné par la formule suivante :

$$Q_1 = L_i + \left( \frac{\frac{n}{4} - \sum_{i=1}^{<Q_1} n_i}{n_{Q_1}} \right) \times a$$

- ▶  $L_i$  : borne inférieure de la classe de  $Q_1$  ;

- ▶  $n$  : effectif total ;

- ▶  $\sum_{i=1}^{<Q_1} n_i$  : somme des effectifs correspondant à toutes les classes inférieures à la classe de  $Q_1$  ;

- ▶  $n_{Q_1}$  : effectif de la classe de  $Q_1$  ;

- ▶  $a$  : amplitude de la classe de  $Q_1$ .

**Exemple 1.26.** Prenons le tableau de l'exemple 1.21. La classe de  $Q_1$  correspond à la classe où les effectifs cumulés atteignent ou dépassent pour la première fois 25% de l'effectif total.

Dans notre exemple, la classe de  $Q_1$  est la classe [1; 6; 1: 7[.

- ▶  $L_i = 1.6$  ;
- ▶  $n = 100$  ;
- ▶  $\sum_{i=1}^{<Q_1} n_i = 8$  ;
- ▶  $n_{Q_1} = 33$  ;
- ▶  $a = 0.1$

$$Q_1 = L_i + \left( \frac{\frac{n}{4} - \sum_{i=1}^{<Q_1} n_i}{n_{Q_1}} \right) \times a = 1.6 + \left( \frac{\frac{100}{4} - 8}{33} \right) \cdot 0.1 = 1.65515.$$

### ▶ Le troisième quartile $Q_3$

Dans ce cas le premier quartile est donné par la formule suivante :

$$Q_3 = L_i + \left( \frac{\frac{3n}{4} - \sum_{i=1}^{<Q_3} n_i}{n_{Q_3}} \right) \times a$$

- ▶  $L_i$  : borne inférieure de la classe de  $Q_3$  ;
- ▶  $n$  : effectif total ;
- ▶  $\sum_{i=1}^{<Q_3} n_i$  : somme des effectifs correspondant à toutes les classes inférieures à la classe de  $Q_3$  ;
- ▶  $n_{Q_3}$  : effectif de la classe de  $Q_3$  ;
- ▶  $a$  : amplitude de la classe de  $Q_3$ .

**Exemple 1.25.** La classe de  $Q_3$  de l'exemple 1.21, correspond à la classe où les effectifs cumulés atteignent ou dépassent pour la première fois 75% de l'effectif total.

Dans notre exemple, la classe de  $Q_3$  est la classe [1.8; 1.9[.

- ▶  $L_i = 1.8$  ;
- ▶  $n = 100$  ;
- ▶  $\sum_{i=1}^{<Q_3} n_i = 8 + 33 + 31$  ;
- ▶  $n_{Q_3} = 22$  ;
- ▶  $a = 0.1$ .

$$Q_3 = L_i + \left( \frac{\frac{3n}{4} - \sum_{i=1}^{<Q_3} n_i}{n_{Q_3}} \right) \times a = 1.8 + \left( \frac{\frac{3 \times 100}{4} - (8 + 33 + 31)}{22} \right) \cdot 0.1 = 1.8136$$