Matière : Mathématiques
Professeur : Yahya MATIOUI

# Statistique Descriptive

## Population statistique / Caractère

### Population statistique

**Definition 1** *La population statistique est l'ensemble qui fait l'objet de l'étude et chaque élément de cet ensemble est appelé " individu" ou " unité statistique".*

***Example 2*** *L'étude suivante donne une répartition de* 10 *élèves suivant le nombre de villes visitées par chacun d'eux :*

$$5 - 4 - 3 - 1 - 1 - 2 - 4 - 5 - 4 - 4$$

*Dans cet exemple la population statistique est l'ensemble des éléves.*

### Caractère

**Definition 3** *La caractère qu'on veut étudier chez une population statistique s'appelle "le caractère" ou "la variable statistique". Le caratère peut être **quantitatif** ou **qualitatif.***

### Somme généralisée

Pour exprimer d'une façon simple la somme de $n$ réels, $x_1, x_2, x_3, ..., x_n$ on utilise le symbole $\sum$ de la manière suivante : $x_1 + x_2 + ... + x_n = \sum\limits_{k=1}^{k=n} x_k$, il se lit somme de $x_k$ lorsque $k$ varie 1 jusqu'à $n$.

Soient $x_1, x_2, x_3, ..., x_n$ des réels et $i$ un réel, alors : $\sum\limits_{k=1}^{k=n} i x_k = i \sum\limits_{k=1}^{k=n} x_k$.

Soient $x_1, x_2, x_3, ..., x_n$ et $y_1, y_2, y_3, ..., y_n$ des réels alors : $\sum\limits_{k=1}^{k=n} (x_k + y_k) = \sum\limits_{k=1}^{k=n} x_k + \sum\limits_{k=1}^{k=n} y_k$.

Soient $x_1, x_2, x_3, ..., x_n$ des réels alors $\dfrac{x_1 n_1 + x_2 n_2 + ... + x_k n_k}{N} = \dfrac{1}{N} \sum\limits_{k=1}^{k=n} x_k n_k = \sum\limits_{k=1}^{k=n} \left( \dfrac{n_k}{N} \right) x_k$

## Types de caractères

### Caractère quantitatif

**Definition 4** *Le caractère quantitatif est un caractère qui peut s'exprimer par des nombres, on distingue le caractère quantitatif **discret** et le caractère quantitatif **continu.***

### Caractère discret

**Definition 5** *Le caractère quantitatif* ***discret*** *est celui qui prend des valeurs isolées, comme le numéro du mois de naissance d'un élève par exemple.*

**Example 6** *Si on étudie le nombre de frères et soeurs des élèves d'une classe, les modalités possibles seront* : $0, 1, 2, 3, ...$

*cette série statistique à caractère discret.*

### Caractère continu

**Definition 7** *Le caractère quantitatif continu est celui qui prend des valeurs très proches, dans ce cas les valeurs du caractère sont rassemblées dans des intervalles qu'on appelle aussi "classes".*

**Example 8** *Les vitesses de* 150 *voitures ont été détectées sur l'autoroute entre Rabat et Casa, on a obtenu le tableau suivant* :

| *vitesse* | $[50, 70[$ | $[70, 90[$ | $[90, 110[$ | $[110, 130[$ | $[130, 150[$ |
|---|---|---|---|---|---|
| *Nombre des voitures* | 10 | 40 | 60 | 25 | 15 |

*Dans cet exemple la population statistique est l'ensemble des voitures, et le caractère* : *les vitesses de* 150 *voitures.* (*C'est un caractère quantitatif* ).

## Caractère qualitatif

**Definition 9** *Le caractère qualitatif est un caractère qu'on ne peut pas associer à un ensemble numérique discret ou continu. On parle de caractère qualitatif quand ce caractère n'est pas chiffré* (*langue, préférence, secteur, couleur,...*)

# Effectif- Effectifs cumulés-Fréquence-Pourcentage

## Effectif

**Definition 10** *L'effectif d'une donnée dans un relevé statistique correspond au nombre de fois ou la donnée apparait. L'effectif total correspond à la somme de tous les effectifs.*

**Example 11** *Voici la liste des notes des élèves d'une classe du tronc commun science lors d'un devoir de mathématiques.*

$$2 - 2 - 2 - 5 - 5 - 6 - 11 - 12 - 13 - 15 - 7 - 9 - 3 - 14$$

| *Les notes* $x_i$ | 2 | 3 | 5 | 6 | 7 | 9 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Effectif* $n_i$ | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*L'effectif total est* : $N = 3 + 1 + 2 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 = 14$

### Effectifs cumulés

**Definition 12** *Si dans une série statistique, les valeurs d'un caractère peuvent être ordonnées, l'effectif cumulé de la valeur $x$ est la somme des effectifs de toutes les valeurs inférieures ou égales à $x$. Il s'agit ici d'effectif cumulé croissant, on pourrait de même définir un effectif cumulé décroissant en prenant la somme des effectifs de toutes les valeurs supérieures ou égale à $x$.*

**Example 13** *D'après l'exemple précédent on a :*

| *Les notes* $x_i$ | 2 | 3 | 5 | 6 | 7 | 9 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Effectif* $n_i$ | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| *Effectif cumulé croissant* $N_i$ | 3 | 4 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |

### Fréquence

**Definition 14** *La fréquence d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total.*

$$f_i = \frac{n_i}{N}$$

**Example 15** *Dans l'exemple précédent on obtient le tableau suivant :*

| *Les notes* $x_i$ | 2 | 3 | 5 | 6 | 7 | 9 | 11 | 12 | 13 | 14 | 15 | *total* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Effectif* $n_i$ | 3 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 14 |
| *Fréquence* $f_i$ | $\frac{3}{14}$ | $\frac{1}{14}$ | $\frac{2}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | $\frac{1}{14}$ | 1 |

### Pourcentage

**Definition 16** *Le pourcentage relatif à la valeur $x_i$ est :* $p_i = f_i \times 100 = \dfrac{n_i}{N} \times 100$.

## Paramètres de position

### Le mode

**Definition 17** *Le mode d'une série statistique $(x_i, n_i)_{1 \leq i \leq k}$ c'est la valeur du caractère ou la classe correspondant au plus fort effectif.*

**Example 18** *Dans l'exemple précédent le mode est la note* 2.

### La moyenne arithmétique

**Definition 19** *La moyenne d'une série statistique $(x_i, n_i)_{1 \leq i \leq k}$ est le nombre réel noté $\overline{x}$ ou $m$ tel que :*

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + ... + n_k x_k}{N} = \frac{\displaystyle\sum_{i=1}^{k} n_i x_i}{N} \quad avec \ : \ N = n_1 + n_2 + ... + n_k$$

*en fonction des fréquences, on obtient :*

$$\overline{x} = f_1 x_1 + f_2 x_2 + ... + f_k x_k$$

**Example 20** *Dans l'exemple précédent la moyenne de la classe dans ce devoir est :*

$$\begin{aligned}\overline{x} &= \frac{2\times 3+3\times 1+5\times 2+6\times 1+7\times 1+9\times 1+11\times 1+12\times 1+13\times 1+14\times 1+15\times 1}{14} \\ &= \frac{53}{7} \backsimeq 7,57\end{aligned}$$

## La médiane

**Definition 21** *La médiane d'une série statistique notée $Me$ est le nombre tel que 50% au moins des individus ont une valeur du caractère inférieure ou égale à ce nombre et 50% au moins des individus ont une valeur du caractère supérieure ou égale à ce nombre.*

**Example 22** *Déterminer la médiane de la série statistique suivante :* $2-2-2-5-6-7-7-8-8$.

*Cette série statistique contient* 9 *valeurs. Donc :*

$$\underbrace{2-2-2-5}-\underbrace{6}_{\searrow Me}-\underbrace{7-7-8-8}$$

*L'effectif total est impair la médiane $Me$ est la valeur située au milieu. Donc :* $Me = 6$.

**Example 23** *Déterminer la médiane de la série statistique suivante :* $3-5-5-6-7-8-8-9-9-9$.

$$\underbrace{3-5-5-6}-\underbrace{7-8}-\underbrace{8-9-9-9}$$

*Cette série statistique contient* 10 *valeurs. Donc :*

*L'effectif total est pair la médiane $Me$ est la demi-somme des* 2 *valeurs situées au milieu. Donc :* $Me = \dfrac{7+8}{2} = 7,5$.

**Remark 24** *On considère une liste de $N$ données rangées par ordre croissant.*

- *Si la série est de taille impair ($N = 2n+1$), la médiane est la donnée de rang $n+1$.*
- *Si la série est de taille pair ($N = 2n$), la médiane est la demi-somme des données de rang $n$ et de rang $n+1$.*

## Série statistique en classes

Lorsque les données d'une série sont trop nombreuses, on range ces données par intervalles appelés classes..

Si $x_i$ et $x_{i+1}$ sont les valeurs de deux caractères de la série statistique, la classe correspondante est dite classe $x_i - x_{i+1}$ . Elle est notée $[x_i, x_{i+1}[$. La valeur $x_i$ est incluse et $x_{i+1}$ est une valeur exclue.

| classe $I_i$ | $[x_1, x_2[$ | $[x_2, x_3[$ | ... | $[x_{k-1}, x_k[$ |
|---|---|---|---|---|
| Effectif $n_i$ | $n_1$ | $n_2$ | ... | $n_k$ |

## La moyenne arithmétique

*Soit* $(I_i, n_i)_{1\leq i\leq k}$ *une série statistique en classes et* $N$ *est l'effectif total et* $c_i$ *est le centre de l'intervalle* $I_i = [x_{i-1}, x_i[$. *La moyenne arithmétique de cette série statistique est le nombre* :

$$\overline{x} = \frac{n_1c_1 + n_2c_2 + ... + n_kc_k}{N} = \frac{\sum_{i=1}^{k} n_ic_i}{N} \quad avec: \ c_i = \frac{x_{i-1} + x_i}{2}$$

**Example 25** *Soit la série statistique suivante* :

| *Caractère* $I_i$ | $[0,10[$ | $[10,20[$ | $[20,30[$ | $[30,40[$ | $[40,50[$ |
|---|---|---|---|---|---|
| *Effectifs* $n_i$ | 13 | 4 | 40 | 38 | 5 |
| *Centre* $c_i$ | 5 | 15 | 25 | 35 | 45 |

*l'effectif total est* : $N = 13 + 4 + 40 + 38 + 5 = 100$. *Donc* :

$$\begin{aligned} \overline{x} &= \frac{n_1c_1 + n_2c_2 + n_3c_3 + n_4c_4 + n_5c_5}{N} \\ &= 26,8 \end{aligned}$$

## La médiane d'une série statistique en classes

*Soit* $(I_i, n_i)_{1\leq i\leq k}$ *une série statistique en classes. On pose* $I_i = [a_{i-1}, a_i]$ *et* $N_i$ *l'effectif cumulé correspondant à la classe* $I_i$ *et soit* $p$ *un entier naturel tel que* : $N_{p-1} \leq \frac{N}{2} < N_p$. *La médiane de cette série statistique est le nombre* $Me$ *qui appartient à la classe* $[a_{p-1}, a_p[$ *et on a* :

$$(Me - a_{p-1})(N_p - N_{p-1}) = \left(\frac{N}{2} - N_{p-1}\right)(a_p - a_{p-1})$$

**Example 26** *Soit la série statistique suivante* :

| *Classe* $I_i$ | $[0,10[$ | $[10,20[$ | $[20,30[$ | $[30,40[$ | $[40,50[$ |
|---|---|---|---|---|---|
| *Effectif* $n_i$ | 13 | 4 | 40 | 38 | 5 |

*Déterminer la médiane de cette série statistique.*
*On commence par ajouter une ligne pour les effectifs cumulés.*

| *Classe* $I_i$ | $[0,10[$ | $[10,20[$ | $[20,30[$ | $[30,40[$ | $[40,50[$ |
|---|---|---|---|---|---|
| *Effectif* $n_i$ | 13 | 4 | 40 | 38 | 5 |
| *Effectif cumulé* $N_i$ | 13 | 17 | 57 | 95 | 100 |

*On a* : $N = 100$. *Donc* $\frac{N}{2} = 50$ *et* $17 \preceq 50 \prec 57$ *de plus* $[a_{p-1}, a_p[ = [20,30[$ *et comme* $N_p = 57$ *et* $N_{p-1} = 17$, *alors* : $Me \in [20,30[$.

*D'autre part, on sait que* : $(Me - a_{p-1})(N_p - N_{p-1}) = \left(\frac{N}{2} - N_{p-1}\right)(a_p - a_{p-1})$. *Donc* :

$$(Me - 20)(57 - 17) = (50 - 17)(30 - 20) \ \text{éq} : Me = \frac{330}{40} + 20 = \frac{113}{4} = 28,25$$

## Paramètre de dispersion : $Etendue, Ecart-moyen, Variance$ et $Ecart$-$type$

### Etendue

**Definition 27** *C'est la différence entre les valeurs extrêmes.*

**Example 28** *Dans l'exemple précédent les notes des élèves (paragraphe* $3)$ *,on a* $2$ *est la valeur minimale et* $15$ *est la valeur maximale. Donc l'étendue est égale à :* $15 - 2 = 13$.

### Ecart-moyen

**Definition 29** *Soit* $(x_i, n_i)_{1 \leq i \leq k}$ *une série statistique et* $\overline{x}$ *sa moyenne arithmétique. L'écart-moyen est défini par :*

$$e = \frac{n_1 |x_1 - \overline{x}| + n_2 |x_2 - \overline{x}| + ... + n_k |x_k - \overline{x}|}{N} = \frac{\sum\limits_{i=1}^{k} n_i |x_i - \overline{x}|}{N}$$

**Remark 30** *pour faire le calcul de l'écart-type, on ajoute les lignes* $|x_i - \overline{x}|$ *et* $n_i |x_i - \overline{x}|$ *au tableau statistique.*

### Variance et ecart-type

**Definition 31** *La variance* $V$ *d'une série statistique de moyenne arithmétique* $\overline{x}$ *dont les valeurs du caractère sont* $x_1, x_2, x_{3,}..., x_k$ *et les effectifs correspondants sont :* $n_1, n_2, n_{3,}..., n_k$ *est égale à :*

$$V = \frac{n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + ... + n_k (x_k - \overline{x})^2}{N} = \frac{\sum\limits_{i=1}^{k} n_i (x_i - \overline{x})^2}{N}$$

*L'écart-type* $\sigma$ *d'une série statistique de variance* $V$ *est égal à :* $\sigma = \sqrt{V}$.

**Remark 32** *L'écart-type exprime la dispersion des valeurs d'une série statistique autour de sa moyenne. Les valeurs extrêmes influencent l'écart-type.*

**Proposition 33** *La variance d'une série statistique* $(x_i, n_i)_{1 \leq i \leq k}$ *est :* $V = \dfrac{n_1 x_1^2 + n_2 x_2^2 + ... + n_k x_k^2}{N} - (\overline{x})^2 = \dfrac{1}{N} \sum\limits_{i=1}^{k} n_i x_i^2 - \overline{x}^2$ *tel que* $\overline{x}$ *est la moyenne arithmétique.*

**Example 34** *Soit la série statistique suivante :*

| | | | |
|---|---|---|---|
| *Caractère* $x_i$ | 1 | 2 | 7 |
| *Effectif* $n_i$ | 5 | 4 | 1 |

- *La moyenne arithmétique de cette série statistique* :

$$\overline{x} = \frac{5 \times 1 + 2 \times 4 + 7 \times 1}{5 + 4 + 1} = 2$$

- *L'écart-moyen*

  *On ajoute la ligne de* $|x_i - \overline{x}|$ *au tableau statistique.*

| *caractère* $x_i$ | 1 | 2 | 7 |
|---|---|---|---|
| *Effectif* $n_i$ | 5 | 4 | 1 |
| $\|x_i - \overline{x}\|$ | $\|1 - 2\| = 1$ | $\|2 - 2\| = 0$ | $\|7 - 2\| = 5$ |

*Donc*

$$e = \frac{n_1 |x_1 - \overline{x}| + n_2 |x_2 - \overline{x}| + n_3 |x_3 - \overline{x}|}{N} = \frac{5 \times 1 + 4 \times 0 + 1 \times 5}{10} = 1$$

- *La variance*

$$V = \frac{n_1 (x_1 - \overline{x})^2 + n_2 (x_2 - \overline{x})^2 + n_3 (x_3 - \overline{x})^2}{N} = \frac{5 \times 1 + 2 \times 0 + 1 \times 25}{10} = 3$$

- *L'écart-type*

  *on a* : $\sigma = \sqrt{V} = \sqrt{3}$.

# Représentation graphique des données statistiques

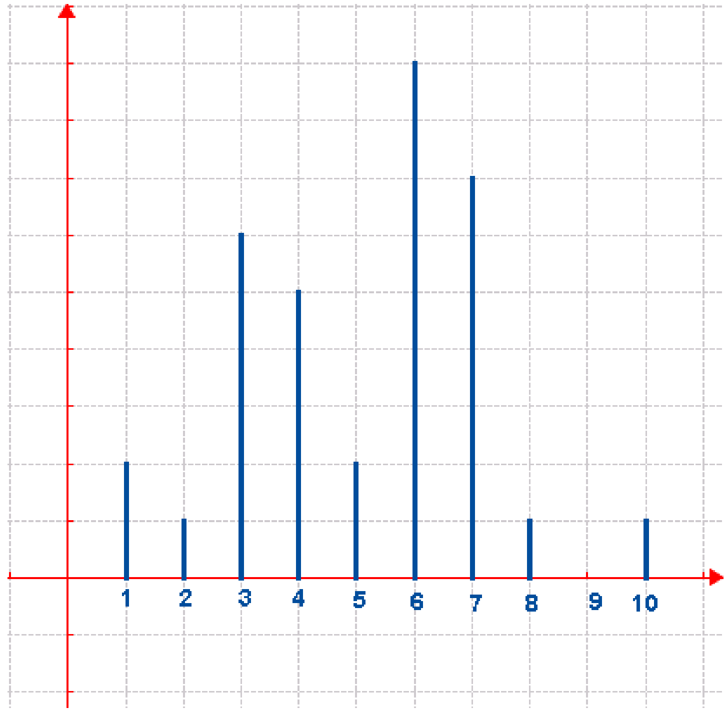
## Variable discrète

Si l'on souhaite représenter les effectifs correspondant à une variable discrète, on obtient un diagramme en bâtons en portant, dans un repère orthogonal, les valeurs $x_i$ en abscisses, les effectifs $n_i$ en ordonnées .Le polygone des effectifs est la ligne polygonale obtenue en joignant les extrémités des bâtons ( à la régle). Le polygone permet de visualiser plus facilement la variable discrète. On peut de la même façon représenter le diagramme en bâtons des effectifs cumulés croissants ou décroissant (ou des fréquences), ainsi que les polygones correspondants.

**Example 35** *Dans une classe, les notes obtenues à un devoir de mathématiques sont* :

| *Note* $x_i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| *Effectif* $n_i$ | 2 | 1 | 6 | 5 | 2 | 9 | 7 | 1 | 0 | 1 |

*Le diagramme à bâtons correspondant est :*




*On remplace parfois l'effectif par la fréquence, ce qui donne bien sûr le même aspect au diagramme.*

## Variable continue

On trace principalement des histogrammes, c'est à dire une suite de rectangles dont l'aire est proportionnelle à l'effectif. Dans la pratique, on porte en abscisses les classes, et en ordonnées les effectifs, si les classes ont même amplitudes. On trace alors des rectangles dont les bases représentent les amplitudes de la classe, et en hauteur l'effectif correspondant. La méthode est la même lorsqu'on représente les fréquences les effectifs cumulés croissant ou décroissant ( et les fréquences du même genre).

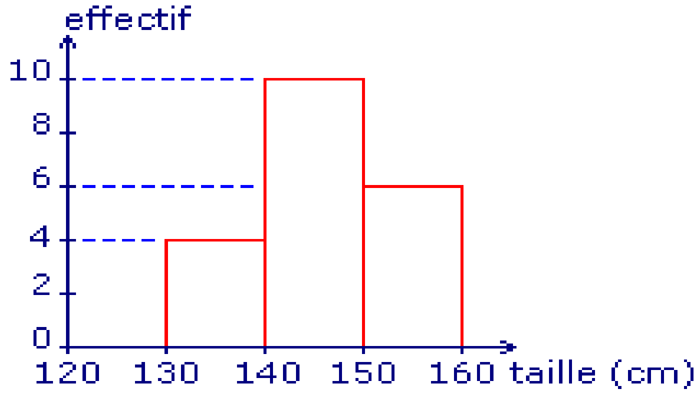
**Example 36** *Voici la répartition des tailles des enfants d'un club de sport.*

| *Taille* | $[130, 140[$ | $[140, 150[$ | $[150, 160[$ |
|---|---|---|---|
| *Effectif* | 4 | 10 | 6 |

*Les valeurs du caractère étudié (la taille) se présentent sous forme d'intervalles. On construit un histogramme avec :*

- *Sur l'axe horizontal, les tailles.*
- *Sur l'axe vertical, les effectifs.*

- *Les hauteurs des barres sont proportionnelles aux effectifs représentés.*




# FIN

Pr : **Yahya MATIOUI**

**www.etude – generale.com**