Université Mohamed Boudiaf - M'sila
Faculté des Sciences et Technologies
Département de Genie Civil

Cours de Probabilité-Statistiques
chapitre 1: Définitions de base-Séries statistiques à une variable

Merini Abdelaziz*

4 octobre 2022

Table des matières

1 Définitions de base **2**
1.1 Introduction 2
1.2 Notions de base 2
1.3 La populationv-l'individusv- L'échantillon 2
1.4 variables- modalités 3
1.5 Différents types de variables statistiques 3
1.5.1 Variables qualitatives : 3
1.5.2 Variables quantitative : 3

2 Séries statistiques à une variable **4**
2.1 Effectif- Fréquence- Pourcentage. 4
2.1.1 **L'Effectif total :** 4
2.1.2 L'Effectif d'une valeur $\mathbf{x_i}$ 4
2.1.3 **La Fréquence :** 4
2.2 Effectif cumulé, Fréquence cumulée. 5
2.2.1 Effectif cumulé : 5
2.2.2 La fréquence cumulée croissante (FCD) d'une valeur (ou d'une classe) : 5
2.2.3 La Fréquence cumulée décroissante (FCD)d'une valeur (ou d'une classe) : 5

*

3 Représentation des données **6**
3.1 Tableau statistique 6
3.1.1 Cas d'une variable qualitative et quantitative discrèt . . . 6
3.1.2 Cas d'une variable quantitatif continue : 6
3.2 Représentations graphiques 6
3.2.1 Cas d'une variable statistique qualitative 7
3.2.2 Cas d'une variable statistique quantitative 8
3.2.3 5.2.2.2 Cas d'une variable quantitative continue 9

4 Caractéristiques de position **10**
4.1 La moyenne 10
4.1.1 Propriétés de la moyenne 11
4.2 Le mode 11
4.3 La médiane 12

5 Caractéristiques de dispersion **13**
5.1 L'étendue 13
5.2 La variance 13
5.2.1 Propriété 14
5.3 L'écart type 14
5.3.1 Propriété 14
5.4 Coefficient de Variation 14
5.4.1 Propriété 14

1 Définitions de base

1.1 Introduction

La statistique descriptive est un ensemble des méthodes scientifiques qui permettent de rassembler et de présenter et d'analyser les données d'une enquête, d'une étude, d'une expérience . . . etc.

1.2 Notions de base

Nous allons commencer par définir les termes utilisés en statistiques pour désigner les observations chiffrées.

1.3 La populationv-l'individusv- L'échantillon

- **La population** est l'ensemble étudié, cet ensemble est noté Ω.

-Les éléments de cette population sont appelés **individus** ou **unités statistiques**, il est noté ω .
- **L'échantillon** est un sous ensemble de la population statistique.
- **La taille** de la population est le nombre d'individus.

Exemple 1. *Les étudiants de 2^{ieme} année Génie Civil=La population, (unité statistique =étudiant),*

(échantillon= groupe de TD). La taille = nombre des étudiants de 2^{ieme} année Génie Civil.

1.4 variables- modalités

Caractère (variable statistique) : est la propriété d'une unité statistique qui peut prendre différentes valeurs pour différentes unités statistiques, est notée X, Y...etc

Exemple 2. *Une étude sur les étudiants de 2^{ieme} année Génie Civile peut porter sur les différentes variables : leur âge, leur sexe, leur nationalité, leur moyenne de l'année, ...etc.*

Les modalités : sont les différentes valeurs qu'une variable peut prendre.

Exemple 3. *Pour la variable "sexe", les modalités sont masculin et féminin. La variable "la note de TD"; les modalités $\in [0, 20]$. La variable "couleur des yeux" les modalités sont : noir, bleu, marron ou vert.*

Remarque 1 : Pour définir une variable, il faut, après avoir indiqué sur quelle population on travaille, préciser l'ensemble des modalités de la variable.

1.5 Différents types de variables statistiques

Nous distinguons deux catégories de variables : les variables qualitatives et les variables quantitatives.

1.5.1 Variables qualitatives :

Une Variable statistique est dite qualitative lorsque ses modalités ne sont pas mesurables. Elle peut être de type :
a)Nominale : les modalités ne sont pas ordonnées.

Exemple 4. *Couleurs des yeux.*

b) Ordinale : les modalités sont ordonnés.

Exemple 5. *mention au BAC, (passable, assez-bien, bien).*

1.5.2 Variables quantitative :

Une Variables statistique est dite **quantitative** lorsque ses modalités sont mesurables. Elle peut être de type

a) Discrète : les modalités sont dénombrables.

Exemple 6. *nombre d'enfants par famille (les modalités sont 0, 1 , 2, 3, 4...).*

b) Continue : les modalités sont définies sur un intervalle continu.

Exemple 7. *la taille (en centimetre), le poids, note de TD ...*

Résumé types de variables

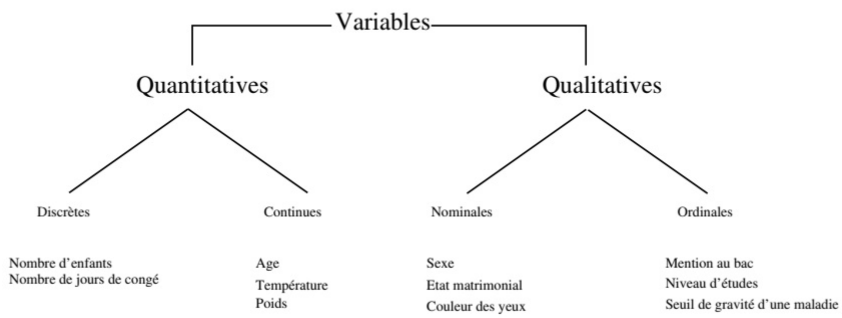

FIGURE 1 – Résumé types de variables

2 Séries statistiques à une variable

Une série statistique est la suite des observations d'une (ou plusieurs) variable(s), relevées sur les individus d'une population.

Les valeurs de la variable X sont notées $x_1, ..., x_i, ..., x_k$.

2.1 Effectif- Fréquence- Pourcentage.

2.1.1 L'Effectif total :

noté N, est le nombre d'individus qui composent la population, on le note $card\,(\Omega) = N$.

2.1.2 L'Effectif d'une valeur $\mathbf{x_i}$

noté n_i est le nombre d'individus associé à cette valeur.

2.1.3 La Fréquence :

noté f_i , est donnée par $f_i = \dfrac{n_i}{N}$.

Le Pourcentage ou fréquence en pourcentage d'un caractère c'est : $p_i = f_i \times 100$.

Propriétés

1) $\sum\limits_{i=1}^{k} n_i = N$.

2) $\sum\limits_{i=1}^{k} f_i = 1$.

3) $0 \leq f_i \leq 1, \forall i \in \{1, 2, ..., k\}$.

Démonstration.

1) par définition.

2) $\sum_{i=1}^{k} f_i = f_1 + f_2 + ... f_k = \frac{n_1}{N} + \frac{n_2}{N} + ... + \frac{n_k}{N} = \frac{n_1 + n_2 + ... + n_k}{N} = \frac{N}{N} = 1.$

3) $\forall i \in \{1, 2, ..., k\} : 0 \leq n_i \leq N$, donc $0 \leq \frac{n_i}{N} \leq \frac{N}{N}$ par suite $0 \leq f_i \leq 1.$

□

Exemple 8. *Une enquête réalisée dans un village porte sur le nombre d'enfants à charge par famille.*

On note x le nombre d'enfants, les résultats sont données par ce tableau :

x_i	0	1	2	3	4	5	6	*total*
n_i	18	32	66	41	32	9	2	200

Nous avons :

Ω : ensemble des familles.

x_i : nombre d'enfants par famille.

$N = 18 + 32 + 66 + 41 + 32 + 9 + 2 = 200.$

x_i	0	1	2	3	4	5	6	*total*
n_i	18	32	66	41	32	9	2	200
f_i	$\frac{18}{200} = 0,09$	$\frac{32}{200} = 0,16$	$0,33$	$0,205$	$0,16$	$0,045$	0.01	1

2.2 Effectif cumulé, Fréquence cumulée.

2.2.1 Effectif cumulé :

Quand les valeurs du caractère sont ordonnées, on peut cumuler les effectifs (les fréquencede) de façon croissante ou décroissante.

L'effectif cumulé croissant (ECC) d'une valeur (ou d'une classe) : est la somme des effectifs de cette valeur (ou de cette classe) et des effectifs (ou des classes) précédentes.

L'effectif cumulé décroissant (ECD) d'une valeur (ou d'une classe) : est la somme des effectifs de cette valeur (ou de cette classe) et des effectifs (ou des classes) suivantes.

2.2.2 La fréquence cumulée croissante (FCD) d'une valeur (ou d'une classe) :

est la somme des fréquences de cette valeur (ou de cette classe) et des fréquences (ou des classes) précédentes

2.2.3 La Fréquence cumulée décroissante (FCD)d'une valeur (ou d'une classe) :

est la somme des fréquences de cette valeur (ou de cette classe) et des fréquences (ou des classes) suivantes.

Exemple 9. *Dans l'exemple précédent : Nous le regardons dans le tableau suivant :*

x_i	0	1	2	3	4	5	6	*total*
n_i	18	32	66	41	32	9	2	200
ECC	18	$18+32=\mathbf{50}$	$18+32+66=116$	157	189	198	200	$\diagup$
ECD	200	182	**150**	84	$2+9+32=43$	$2+9=11$	2	$\diagup$

50 familles qui ont au plus d'un enfant .
150 familles qui ont au moins de 2 enfants

3 Représentation des données

3.1 Tableau statistique

Un tableau statistique est un moyen d'organiser, classifier et ranger par ordre croissant (ou décroisant) les données brutes de la série statistique, pour les bien représenter.

Le choix du type de tableau dépend de la nature des variables.

3.1.1 Cas d'une variable qualitative et quantitative discrèt

Les modalities x_i	x_1	x_2	....	x_k	total
Effectif n_i	n_1	n_2	....	n_k	N
ECC	n_1	n_1+n_2	....	N	$\diagup$
Fréquense f_i	f_1	f_2	....	f_k	1
FCC	f_1	f_1+f_2	...	1	$\diagup$

3.1.2 Cas d'une variable quantitatif continue :

Les classes x_i	$[b_1, b_2[$	$[b_2, b_3[$	....	$[b_{k-1}, b_k[$	total
Centre de classe c_i	c_1	c_2	....	c_k	$\diagup$
Effectif n_i	n_1	n_2	....	n_k	N
ECC	n_1	n_1+n_2	....	N	$\diagup$
Fréquense f_i	f_1	f_2	....	f_k	1
FCC	f_1	f_1+f_2	...	1	$\diagup$

Avec b_i et b_{i+1} sont les bornes de la classe i. et c_i est le centre de cette classe avec $c_i = \dfrac{b_i + b_{i+1}}{2}$,$\forall i \in \{1, 2, ..., k\}$.

3.2 Représentations graphiques

Les représentations graphiques permettent de résumer visuellement des séries statistiques.

Le choix du type de graphe dépend de la nature des variables.

3.2.1 Cas d'une variable statistique qualitative

i) Diagramme circulaire (ou camembert) Le principe du graphe consiste à diviser un cercle ou un disque en secteurs.

Chaque secteur représente une modalité sa surface est proportionnelle à la fréquence de cette modalité.

L'angle d'un secteur est déterminé à l'aide de la règle de trois de la manière suivante :

$$\begin{array}{lcl} N & \rightarrow & 360^0 \\ n_i & \rightarrow & \alpha_i \text{ (degré de la modalité } i\text{)} \end{array}$$

alors

$$\alpha_i = \frac{n_i \times 360}{N}.$$

Remarque 1. $\forall i \in \{1, 2, ..., k\} : \alpha_i = f_i \times 360°$.

Exemple 10. *Soit le tableau d'état civil suivant :*

Les modalities x_i	*Célibataire*	*Marié(e)*	*veuf(ve)*	*Divorcé(e)*	*total*
Effectif n_i	16	12	6	6	40
Fréquense f_i	$\frac{16}{40} = 0.4$	0.30	0.15	0.15	1
Anngle α_i	144	108	54	54	360
pourcentage pi	40	30	15	15	100

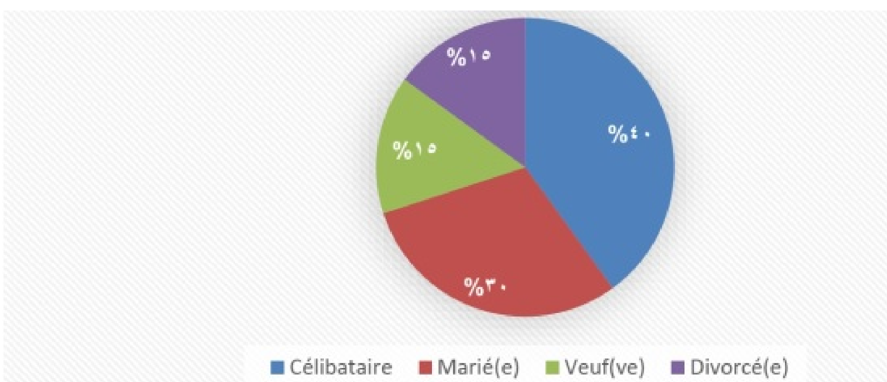

FIGURE 2 – Diagramme circulaire d'´etat civil

ii) Diagrammes en barres (en bandes) (tuyaux d'orgue) Un diagramme en barres est une représentation graphique de données à l'aide de rectangles de même largeur.

Les modalités sont représentées sur l'axe horizontal, les effectifs sur l'axe vertical. chaque modalité correspond une barre.

Les hauteurs des barres sont proportionnelles aux effectifs représentés la modalité.

Exemple 11. *Dans l'exemple précédent, on a*

Les modalities x_i	*Célibataire*	*Marié(e)*	*veuf(ve)*	*Divorcé(e)*	*total*
Effectif n_i	16	12	6	6	40
Fréquense f_i	$\frac{16}{40} = 0.4$	0.30	0.15	0.15	1
hauteur h_i	8	6	3	3	/

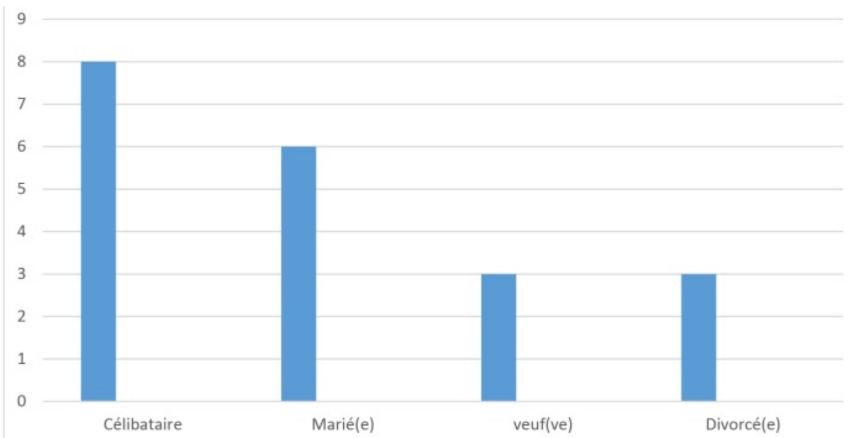

FIGURE 3 – Diagramme en barres d'état civil

3.2.2 Cas d'une variable statistique quantitative

Pour les variables quantitatives, il existe deux types de représentation graphique qui sont :

Cas d'une variable quantitative discrète :

i)Diagramme en bâton Ce diagramme comporte deux axes, un axe horizontal qui représente les valeurs de la variable, et un axe vertical qui représente les Effectif. ou les fréquences, à chaque valeur on associe un segment (bâton) dont sa hauteur est proportionnelle à l'effectif ou à la fréquence de cette modalité.

Exemple 12. *Dans l'exemple 9, on a*

x_i	0	1	2	3	4	5	6	*total*
n_i	18	32	66	41	32	9	2	200

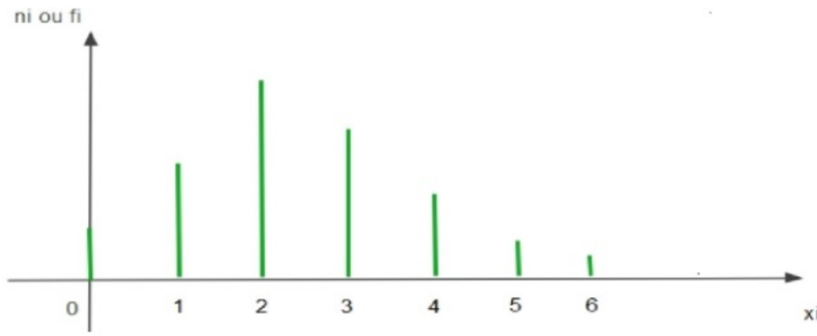

FIGURE 4 – Diagramme en bâton

ii) Polygone des effectifs (ou des fréquences) Le polygone des effectifs (ou des fréquences) est obtenu en reliant les extrémités des bâtons.

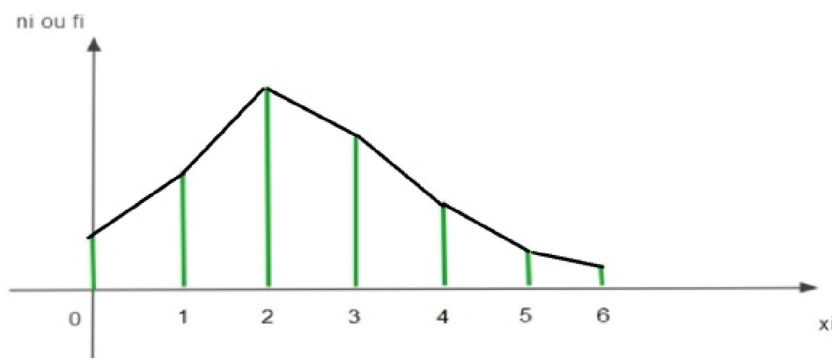

FIGURE 5 – Polygone des effectifs

Polygone des effectifs cumulés

3.2.3 5.2.2.2 Cas d'une variable quantitative continue

i) Histogramme des fréquences (ou effectifs) Lorsque le caractère étudié est quantitatif continu, et lorsque les données sont regroupées en classes, on peut représenter la série par un histogramme :

l'aire de chaque rectangle est alors proportionnelle à l'effectif (ou a la fréquence) associée à chaque classe.

Remarque 2. *Lorsque les classes ont la même amplitude, c'est la hauteur de chaque rectangle qui est proportionnelle à l'effectif.*

Exemple 13. *Soit le tableau suivant :*

Classes	$[0;4[$	$[4;8[$	$[8;12[$	$[12;16[$	$[16;20[$
Effectifs	4	5	11	6	2

Donc

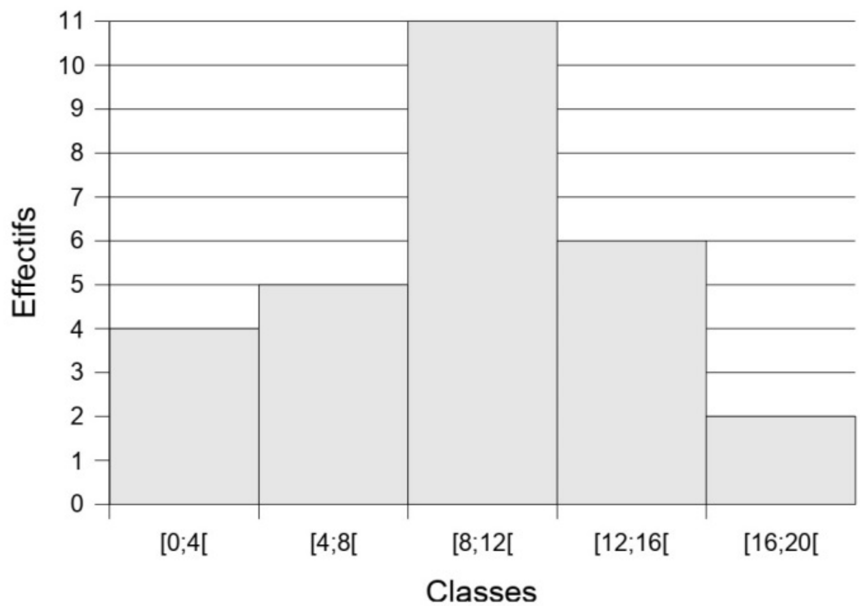

FIGURE 6 – Histogramme

4 Caractéristiques de position

Seuls les caractères quantitatifs se caractérisent par nombres réels appelés caractèristiques de position. les caractèristiques (ou mesures) de tendance centrale sont des outils trés utilisés pour interpréter les données.

Ce sont des valeurs qui résument l'ensemble des données et permettent de se faire une idée de la tendance.

Les caractèristiques de position sont la moyenne, le mode et la médiane .

4.1 La moyenne

la moyenne d'une série statistique (x_i, n_i) est la quantité

$$\bar{x} = \sum_{i=1}^{k} \frac{n_i x_i}{N} = \sum_{i=1}^{k} f_i x_i$$

Remarque 3. *Dans le cas continu, on choisit la valeur x_i égale au centre de la classe correspondante c_i ,c-à-d*

$$\bar{c} = \sum_{i=1}^{k} \frac{n_i c_i}{N} = \sum_{i=1}^{k} f_i c_i$$

4.1.1 Propriétés de la moyenne

i) La somme des écarts à la moyenne est égale à 0

$$\sum_{i=1}^{k} n_i \left(x_i - \bar{x} \right) = 0.$$

ii) Propriété de linéarité : Si $\forall i \in \{1, 2, ..., k\}$, on a : $y_i = a x_i + b$, alors : $\bar{y} = a\bar{x} + b$.

iii) Propriété des moyennes partielles : Soient deux séries statistiques de moyennes respectives $\bar{x}_1$ et $\bar{x}_2$ et d'effectifs respectifs N_1 et N_2.

La moyenne des deux séries est

$$\bar{x} = \frac{N_1 \bar{x}_1 + \bar{x}_2 N_2}{N_1 + N_2}$$

4.2 Le mode

le mode est la valeur la plus fréquente d'une série statistique, il est dénoté par M_o

Remarque 4. *1) Le mode peut etre calculé pour tous les types de variable, quantitative et qualitative.*
2) Le mode n'est pas nécessairement unique.
3) Pour une série répartie en classe on parle de classe modale.

Le mode dans le cas d'une variable quantitative continue On peut calcler le mode pour une variable quantitative continue par la formule suivante :

$$M_o = L_i + \frac{\triangle_1}{\triangle_1 + \triangle_2} a_i$$

Où
- L_i : la borne inférieure de la classe modale.
- a_i : le pas de la classe modale.
- $\triangle_1 = n_0 - n_1$, $\triangle_2 = n_0 - n_2$. ou bien $\triangle_1 = f_0 - f_1, \triangle_2 = f_0 - f_2$.
- n_0 et f_0 sont l'effectif et la fréquence associés à la classe modale.
- n_1 et f_1 sont l'effectif et la fréquence de la classe qui précède la classe modale.
- n_2 et f_2 sont l'effectif et la fréquence de la classe qui suit la classe modale.

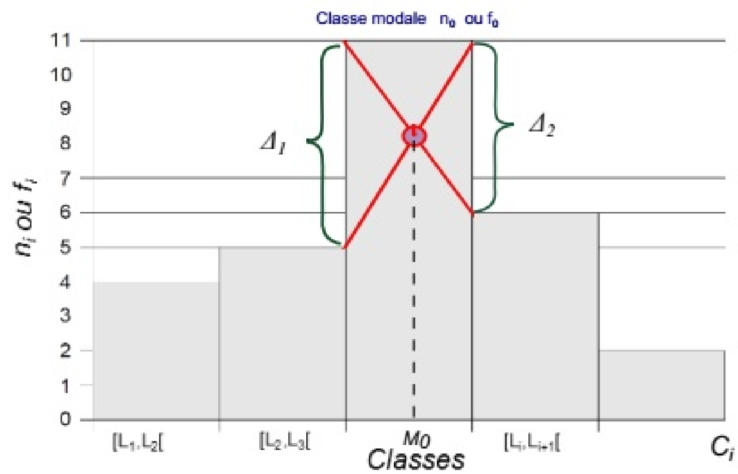

FIGURE 7 – Détermination graphique du mode (cas continu)

Exemple 14. *Dans l'exemple 9 :*

x_i	0	1	2	3	4	5	6	*total*
n_i	18	32	66	41	32	9	2	200

on a $M_o = 2$.

Dans l'exemple 14 :

Classes	$[0;4[$	$[4;8[$	$[8;12[$	$[12;16[$	$[16;20[$
Effectifs	4	5	11	6	2

La classe modale.$=[8;12[$ *. Donc,*

$M_o = 8 + \frac{6}{6+5} \times 4 \simeq 10.18$

4.3 La médiane

La médiane est la valeur de la série (qui n'est pas toujours une valeur de la série) qui partage la série en deux groupes de même effectif (ou de même fréquence).

est dénotée par M_e.

Remarque 5. *Quand les données sont regroupées en classes on parle de classe modale.*

Pour calculer la médiane, il est nécessaire de classer les données de la série de la plus petite à la plus grande.

Comment déterminer la médiane d'une série statistique ?

Pour déterminer la médiane d'une série statistique :

Etape 1 : On range les données dans l'ordre croissant

Etape 2 : On calcule les effectifs cumulés croissants et l'effectif total de la série.

Le calcul est légèrement différent suivant que le nombre de données est pair ou impair.

i) Si $N = 2k + 1$ (i.e. impair), alors $Me = x_{k+1} = x_{\frac{N+1}{2}}$.

ii) Si $N = 2k$ (i.e. pair), alors $M_e = \dfrac{x_k + x_{k+1}}{2} = \dfrac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2}$.

Exemple 15. *Soit la série suivante :* $3; 3; 7; 7; 7; ; 12; 15; 15; 16; 17; 19$. *On a* $N = 11$, *alors* $M_e = x_{\frac{11+1}{2}} = x_6 = 12$.

Soit la série suivante : $3; 3; 7; 7; 7; ; 12; 15; 15; 16; 17; 19; 20$.*On a* $N = 12$, *alors* $M_e = \dfrac{x_6 + x_7}{2} = \dfrac{12 + 15}{2} = 13, 5$.

Exemple 16. *Dans l'exemple 9 :*

x_i	0	1	2	3	4	5	6	*total*
n_i	18	32	66	41	32	9	2	200
ECC	18	50	116	157	189	198	200	//

$\dfrac{N}{2} = 100$, *alors* $M_e = 2$.

5 Caractéristiques de dispersion

Les Caractéristiques de dispersion usuels sont l'étendue, la variance et l'écart-type.

5.1 L'étendue

L'étendue est la différence entre la plus grande valeur et la plus petite valeur du caractère, donnée

par la quantité

$$e = x_{\max} - x_{\min}$$

Remarque 6. $e > 0$.

5.2 La variance

La variance est la quantité

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

Remarque 7. $V(X) \geq 0$.

théoreme 1. *Soit* (x_i, n_i) *une série statistique de moyenne* $\bar{x}$ *et de variance* $V(X)$*. Alors*

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

5.2.1 Propriété

Si $\forall i \in \{1, 2, ..., k\}$, on a : $y_i = ax_i + b$, avec a,b$\in \mathbb{R}$, alors : $V(Y) = a^2 V(X)$.

5.3 L'écart type

L'écart type est la quantité

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

5.3.1 Propriété

Si $\forall i \in \{1, 2, ..., k\}$, on a : $y_i = ax_i + b$, avec $a, b \in \mathbb{R}$, alors :$\sigma_Y = |a| \sigma_X$.

5.4 Coefficient de Variation

Le coefficient de variation est une mesure relative de la dispersion des données autour de la moyenne. Le coefficient de variation est donné par

$$c_v = \frac{\sigma_X}{\bar{x}}$$

5.4.1 Propriété

1) Plus grand est le coefficient de variation, plus grande est la dispersion.
2) Généralement exprimé en pourcentage. Sans unité.
3) Il permet la comparaison entre 2 séries.

Références

[1] Abdennasser Chekroun, Statistiques descriptives et exercices, Université de Tlemcen, Algerie,2017/2018.

[2] Abdelkader Sahraoui, Cours de statistique descriptive, Université de Medea, Algerie, 2020