Université Abderrahmane Mira-Bejaia
Faculté des Sciences Économiques, Commerciales et des Sciences de Gestion

Département de la formation initiale SEGC-LMD

Laboratoire de Recherche en Management et Techniques Quantitatives

# Polycopié de cours

# Statistique I (Statistiques descriptives)

Cours destiné aux étudiants de Licence : Première année L1

Réalisé par Dr. BERRAH Kafia

Année universitaire : 2020/2021

## Avant-propos

Ce polycopie de cours s'adresse tout particulièrement aux étudiants de première année des sciences économiques, des sciences de gestion et des sciences commerciales. Il correspond au programme pédagogique du module statistique descriptive habituellement enseigné dans les premières années. Cependant, il peut être utile à toute personne souhaitant connaitre les notions et les techniques fondamentales dans ce domaine.

La rédaction de ce support est un fruit de plusieurs années de travail (2013-2021).

Le polycopie est organisé en cinq chapitres sur les différents aspects de la statistique descriptive : les concepts et les définitions de base utilisées dans le domaine de la statistique descriptive (la population, unité statistique et la typologie des variables,…) ; les outils de base de la statistique descriptive (les tableaux statistiques et les représentations graphiques) ; les caractéristiques de position décrivant les distributions statistiques (mode, médiane et moyennes) ; les caractéristiques de dispersion (étendue, variance, écart-type….) et enfin les caractéristiques de forme (l'asymétrie et l'aplatissement).

Chaque chapitre est associé par de nombreux exemples concrets et en contexte, afin d'éclairer la théorie par l'illustration et se conclut aussi par des exercices permettant de vérifier l'acquisition des concepts essentiels qui ont été introduits.

## Chapitre 1 : Vocabulaire et concepts de base

La statistique descriptive est un ensemble de méthodes permettant de d'écrire et d'analyser des phénomènes susceptibles d'être dénombrés et classés. Le présent chapitre a pour objet de présenter l'essentiel du vocabulaire des statistiques descriptives, comme la population, l'unité statistique, le caractère,…

### 1. Différence entre la statistique et les statistiques

On observe une différence entre la statistique qui est un outil scientifique et les statistiques qui sont des informations qualitatives ou quantitatives sur lesquelles on va appliquer cet outil pour mener une étude entre autres.

**1.1. La statistique :** une méthode scientifique qui consiste à réunir des données chiffrées sur un phénomène, puis à analyser, à commenter et à critiquer ces données. Elle permet également de tirer des conclusions logiques mais aussi de prendre des décisions à partir des analyses effectuées.

**1.2. Les statistiques :** est un ensemble de données chiffrées sur un sujet précis.

### 2. Population, individu et échantillon

**2.1. La population statistique :** on appelle population statistique un ensemble d'éléments homogènes sur lesquels porte l'étude statistique. Cet ensemble doit être bien défini. Le nombre d'individus dans la population est la taille de la population.

**Exemples :** les étudiants d'une université, les clients d'une entreprise….

**2.2. L'individu :** On appelle individu, c'est chaque élément de la population étudiée. On utilise également le terme **unité statistique** pour désigner un individu. Les individus d'une population peuvent donc être, selon les cas étudiés, des êtres humains, des objets ou des évènements.

**Exemple :** Dans les exemples précédents, l'unité statistique : un étudiant, un client.

**2.3. L'échantillon** : C'est un sous-ensemble construit et représentatif d'une population donnée.

### 3. Le caractère ou la variable statistique

Le caractère est l'aspect particulier que l'on veut étudier, pour décrire les individus qui composent une population ou un échantillon.

**Exemples :** les notes des étudiants à l'examen de statistique, les mentions qu'ils ont obtenues à leur Bac, leur Sexe, leur poids, les couleurs de leurs yeux, le chiffre d'Affaire par entreprise, le nombre d'enfants par ménage,…

**4. Les modalités de la variable statistique**

Les modalités sont les diverses situations susceptibles d'être prises par le caractère étudié. Ce dernier peut posséder une ou plusieurs modalités. Les modalités sont exhaustives et mutuellement exclusives c'est-à –dire chaque individu doit pouvoir être classé dans une et une seule modalité.

**Exemple1 :** Le sexe est un caractère qui présente deux modalités : féminin ou masculin

**Exemple 2 :** Si la variable étudiée est le nombre d'enfants par familles, les modalités de ce caractère peuvent être 0,1, 2, 3, etc.

**5. Typologie des variables statistiques**

On distingue deux types de variables statistiques : les variables qualitatives et les variables quantitatives.

**5.1. Variable qualitative :** Une variable statistique est dite de nature qualitative si ses modalités ne sont pas mesurables, c'est- à –dire ses diverses modalités sont décrites par un mot traduisant un état.

**Exemple :** La religion, le sexe et le lieu de résidence sont des caractères qualitatifs.

Les modalités d'une variable qualitative peuvent être classées en deux types de variables : variables qualitatives nominales et variables qualitatives ordinales.

**5.1.1. Variable qualitative nominale** : La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées (ne sont pas naturellement classées).

**Exemples :** la couleur des yeux (vertes, marrons, bleues, noisettes), le sexe (féminin et masculin), l'état matrimonial (célibataire, marie, divorcé et veuf).

**5.1.2. Variable qualitative ordinale** : La variable est dite qualitative ordinale si ses modalités peuvent être ordonnées.

**Exemples :** niveau d'instruction (primaires, secondaires, universitaires), classement de l'état de santé des malades (Bon, Mauvais, Très mauvais).

**5.2. Variable quantitative :** Une variable est dite de nature quantitative si toutes ses modalités sont mesurables c'est-à-dire elles s'expriment par un chiffre.

**Exemple :** Le poids, la taille et l'âge d'un groupe d'individu sont des caractères quantitatifs.

Il existe deux types de variables quantitatives : les variables discrètes et les variables continues :

**5.2.1. Variable quantitative discrète (discontinue) :** Une variable est dite discrète lorsque ses différentes modalités sont des valeurs numériques isolées, c'est-à-dire sous forme des nombres entiers : 0, 1, 2,…

**Exemple 1 :** les modalités de la variable nombre d'enfants par famille sont : 0, 1, 2, 3,…C'est une variable quantitative discrète.

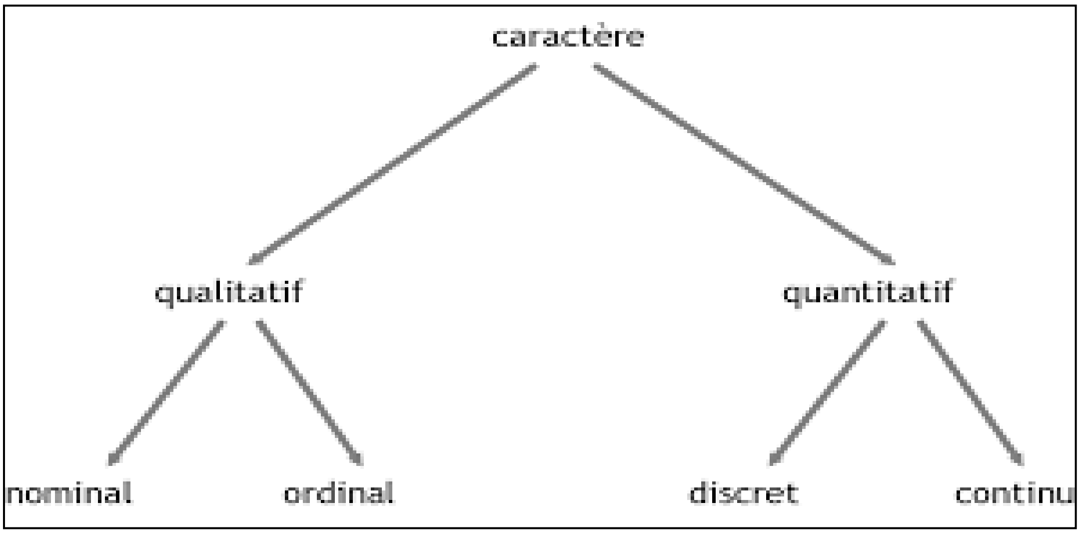
**Exemple 2** : nombre de lit dans un hôpital c'est une variable quantitative discrète.

**5.2.2. Variable quantitative continue :** une variable est dite continue lorsqu'elle peut prendre n'importe quelles valeurs d'un intervalle.

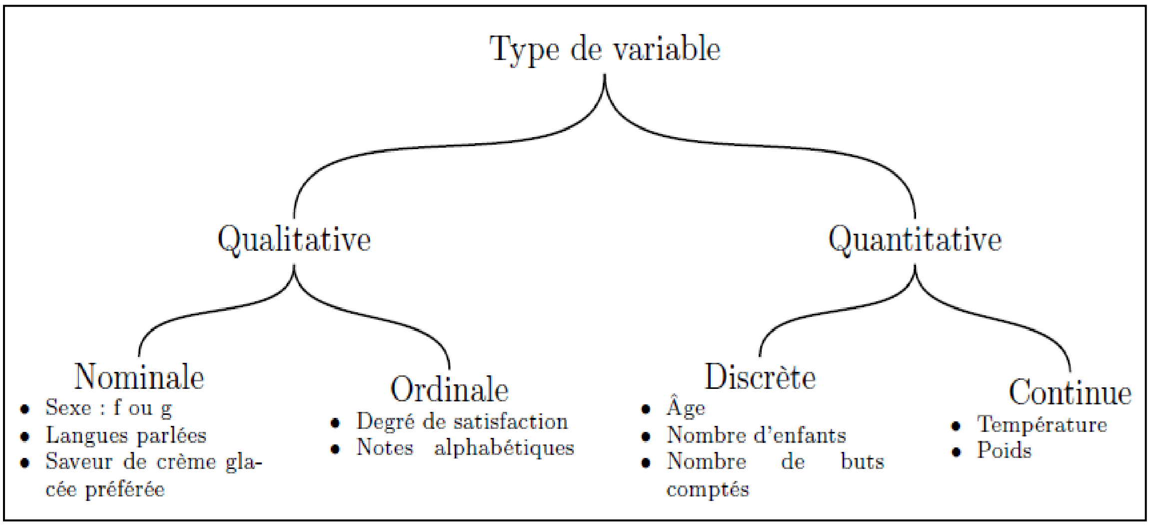
**Exemple 1 :** le poids et la taille des individus sont des caractères continus.

**Exemple 2** : Le temps pris pour effectuer un trajet en voiture est une variable continue.

Le schéma ci-après précise les types de variables statistiques.




**Exemple d'illustration :** le schéma suivant indique la classification de certains caractères en fonction de la typologie des variables.




### 6. La série statistique

Une série statistique correspond aux différentes modalités d'un caractère sur un échantillon d'individus appartenant à une population donnée.

Une série statistique est obtenue par l'observation d'une variable chez n individus ; elle correspond à la liste des valeurs ou modalités prises par la variable chez chacun des n individus.

**Exemple :** soit la série suivante qui indique les notes d'un groupe de 30 étudiants à l'examen en statistique.

2; 3; 4; 6; 5; 2; 3; 6; 4; 4; 2; 3; 3; 4; 4

5; 6; 4; 4; 3; 3; 3; 3; 4; 4; 4; 4; 5; 5; 5

On observe la variable est la note d'un étudiant : $x1 = 2$ ; $x2 = 3$ ; …

Evidemment les notes sont égales aux nombres 2, 3, 4,5 et 6. Alors le nombre des modalités du caractère est 5.

**Exemple global :** Nous résumons les différents concepts cités antérieurement dans cet exemple.

On fait une étude sur les étudiants de1ère SEGC-LMD de l'université de Bejaia. On prend comme objet d'étude la section A selon les caractères suivants : la taille, la couleur des yeux, la mention de BAC, le nombre des étudiants par groupes.

- Définir la distribution statistique.

**Solution** :

**La population statistique** : les étudiants de1ère SEGC-LMD de l'université de Bejaia.

**L'échantillon** : la section A.

**L'unité statistique** : un étudiant

**Les caractères étudiés** : taille, couleur des yeux, mention de BAC, nombre des étudiants par groupes.

**Le type des caractères étudiés** :

- Quantitatif continu pour le caractère : taille.
- Qualitatif nominal pour le caractère : couleur des yeux.
- Qualitatif ordinal pour le caractère : mention de BAC).
- Quantitatif discret pour le caractère : nombre des étudiants par groupes.

**Les modalités** : 1m60, 1m65…// marrons, bleus…// bien, très bien…. // 15, 20,21,…

**Exercices**

**Exercice 1** : Une étude est réalisée sur la pointure des chaussures des jeunes algériens de 10 à 20ans.

1- Quelle est la population et l'unité statistique étudiée ?

2- Quel est le caractère étudié et son type ?

**Exercice 2 :** Quelle est la nature des caractères ci-dessous ?

- Nombre de portable vendus chaque jour sur le marché.
- Rémunérations des enseignants d'un lycée.
- Les pays de l'Union européenne.
- Les niveaux de formation des salariés.
- Les formes de contrat de travail.
- Prix des articles achetés.
- Nombre de personnes par ménages.

**Exercice 3 :** On a relevé, pour 30 familles, le nombre d'enfants par famille :

5 ; 0 ; 3 ; 1 ; 2 ; 7 ; 1 ; 2 ; 3 ; 1 ; 0 ; 1 ; 3 ; 4 ; 0 ; 0 ; 1 ; 1 ; 2 ; 2 ; 1 ; 2 ; 0 ; 1 ; 2 ; 3 ; 0 ; 4 ; 1 ; 4.

1- Déterminer la population et l'unité statistique.
2- Quel est le caractère étudié et sa nature.
3- Iditenifier les différentes modalités de cette série statistique.

## Chapitre 2 : Présentions d'une série statistique : tableaux et graphes

A l'issue de la collecte des données (lors d'une enquête par exemple), les informations recueillies ne sont pas immédiatement exploitables. Il est alors nécessaire de les organiser, les ordonner et les présenter de façon lisible et facilement compréhensible. Pour cela la statistique descriptive offre des techniques pour la représentation des données sous forme de tableaux ou de graphes.

### 1. Présentation d'une série statistique sous forme des tableaux

Soit une population composée de $n$ individus, sur laquelle on a étudié un caractère possédant $k$ valeurs possibles $x_1, x_2, x_3, ..., x_i, ..., x_k$.

Soient : $n_1$ le nombre d'individus ayant la modalité $x_1$.

$n_2$ le nombre d'individus ayant la modalité $x_2$.

… …. …. … …. …

$n_k$ le nombre d'individus ayant la modalité $x_k$.

Alors $n_i$ indique le nombre de fois que la modalité $x_i$ a été observée. Ce nombre est appelé **l'effectif** ou **fréquence absolue** de la modalité **xi**, c'est le nombre d'individus **ni** présentant la modalité xi.

Le nombre total d'individus de la population observée est appelé **l'effectif total** (noté N), c'est-à-dire c'est la somme des effectifs. On note donc : $N = \sum_{i}^{k} n_i = n_1 + n_2 + ...n_i + ...n_k$ .

**La fréquence relative (notée $fi$)** : c'est la proportion d'individus ayant la modalité xi, est égale au rapport de l'effectif de la modalité xi du caractère sur l'effectif total. Donc $f_i = \frac{n_i}{N}$ et $\sum_{i=1} f_i = 1$.

La valeur de la fréquence relative est un nombre compris entre 0 et 1. Elle peut s'exprimer en pourcentage, elle est égale à $f_i \times 100$.

De manière générale, le regroupement des informations chiffrées dans un tableau statistique se présentent souvent sous la forme suivante :

| Nationalités ( $X_i$ ) | Effectifs Absolus ( $n_i$ ) | Fréquences relatives ( $f_i = n_i / N$ ) |
|---|---|---|
| X1<br>X2<br>…<br>…<br>$X_k$ | n1<br>n2<br>…<br>….<br>$n_k$ | f1<br>f2<br>…<br>…<br>$f_k$ |
| Total | N | 1 |

Le type du tableau statistique à utiliser pour présenter une série statistique dépend de la nature du caractère étudié.

### 1.1. Présentation des données d'un caractère qualitatif

Pour présenter les résultats observés, nous distinguons d'abord les différentes modalités du caractère étudié (les différentes valeurs de la variable considérée). On les désigne par X1, X2, X3,… Xk. Il s'agit donc d'un caractère présentant k modalités. Nous déterminons ensuite le nombre d'individus associés à chacune des modalités, ils sont identifiés par $n_1, n_2, n_3, ..., n_k$.

Nous pouvons aussi évaluer fréquence relative de chacune des modalités. Ces dernières sont représentées par $f_1, f_2, ...., f_k$ et obtenues grâce au rapport suivant : $f_i = \frac{n_i}{N}$ avec $N$ effectif total.

Nous pouvons également exprimer les fréquences relatives en pourcentage. Il suffit de multiplier la fréquence relative de chacune des modalités de la variable par **100.**

**Exemple :** une enquête dans un garage-parking concernant la couleur des automobiles a révélé les résultats suivants : verte, rouge, bleu, noire, blanche, rouge, verte, blanche, bleu, verte, rouge, verte, rouge, verte, bleu.

Pour pouvoir interpréter ces données, on doit les ranger dans un tableau de fréquences suivantes :

Tableau : répartition des automobiles selon la couleur.

| Couleur (xi) | verte | rouge | bleu | blanche | noire | total |
|---|---|---|---|---|---|---|
| Effectif (ni) | 5 | 4 | 3 | 2 | 1 | 15 |

Par exemple le nombre 5 c'est le nombre de voitures de couleur verte.

### 1.2. Présentation des données d'un caractère quantitatif discret

Les fréquences absolues (ni) correspondent à des caractères isolées (xi) de la variable statistique lesquelles sont ordonnées dans le tableau de manière croissante(ou décroissante).

**Exemple** : soit la série suivante qui indique les notes d'un groupe de 30 étudiants à l'examen en statistique.

3; 2; 4; 6; 5; 2; 3; 6; 4; 4; 2; 3; 3; 4; 4
5; 6; 4; 4; 3; 3; 3; 3; 4; 4; 4; 4; 5; 5; 5

- Construire le tableau des fréquences.
- Calculer les effectifs cumulés croissants et décroissants.

**Solution** :

Le nombre d'observations (effectif total) est N= 30.
Le nombre de modalités est 5 modalités, qui représentent les notes des étudiants : 2, 3, 4, 5 et 6. Sachant : la note 2 se rencontre 3 fois, la note 3 est 8 fois, la note 4 est 11 fois, la note 5 est 5 fois et la note 6 est 3 fois. Donc l'effectif de la valeur 2 est 3, l'effectif de la valeur 3 est 8, …..Etc.

- Le tableau des fréquences est présenté de la manière suivante

| Note ( $X_i$ ) | Effectif ( $n_i$ ) | Fréquence relative ( $f_i = \frac{ni}{N}$ ) |
|---|---|---|
| 2 | 3 | 3/30 = 0,1 |
| 3 | 8 | 8/30 = 0,2666 |
| 4 | 11 | 11/30 = 0,3666 |
| 5 | 5 | 5/30=0,1666 |
| 6 | 3 | 3/30 = 0.1 |
| Total | 30 | 1 |

- Le tableau des fréquences absolues cumulées

Les fréquences cumulées permettent de connaitre le nombre (ou la proportion) d'observations inférieures/supérieures à une valeur donnée, elles se calculent de la manière suivante :

**- Cas des fréquences cumulées croissantes** : on additionne (ou on cumule) les fréquences absolues successivement jusqu'à obtenir leur somme totale, de la manière suivante :

- L'effectif cumulé croissant de la 1ère modalité est toujours égal à l'effectif partiel de la 1ére modalité.
- L'effectif cumulé croissant d'une modalité est la somme de l'effectif partiel de cette modalité et de l'effectif cumulé croissant de la modalité précédente.
- L'effectif cumulé croisant de la dernière modalité est toujours égal à l'effectif total.

**-Cas des fréquences cumulées décroissantes** : on poursuit les étapes suivantes :

- L'effectif cumulé décroissant de la 1ère modalité est toujours égal à l'effectif total.
- L'effectif cumulé décroissant d'une modalité est la différence de l'effectif cumulé de la modalité précédente et de l'effectif partiel de la modalité précédente.
- Et ainsi de suite jusqu' à l'effectif cumulé décroisant de la dernière modalité qu'est toujours égal à l'effectif partiel de la dernière modalité.

| $x_i$ | $n_i$ | $n_i \uparrow$ | $n_i \downarrow$ |
|---|---|---|---|
| 2 | 3 | 3 | 30 |
| 3 | 8 | 11 | 27 |
| 4 | 11 | 22 | 19 |
| 5 | 5 | 27 | 8 |
| 6 | 3 | 30 | 3 |
| **Total** | 30 | - | - |

**Remarque** : Le même principe à appliquer pour les fréquences relatives cumulées.

### 1.3. Présentation des données d'un caractère quantitatif continu

Lorsque la variable statistique est continue, l'établissement du tableau de fréquences implique d'effectuer au préalable une répartition en classes des données. Cela nécessite de définir le nombre de classes attendu et donc intervalle de classe pour chaque modalité du caractère.

Par convention, l'intervalle est fermé à gauche et ouvert à droite, la ième classe est alors notée « $x_{i-1}, x_i$ » correspondant à l'intervalle [ $x_{i-1}, x_i$ [.

**L'amplitude** $a_i$ d'une classe est la largeur de l'intervalle de cette classe. La ième classe aura comme amplitude : $a_i = x_i - x_{i-1}$.

Les classes peuvent avoir toutes la même amplitude ou avoir des amplitudes différentes. Il peut également arriver que les classes extrêmes ne soient pas délimitées.

Comme la variable statistique ne prend plus une valeur précise $X_i$, mais un ensemble de valeurs possibles comprises dans un intervalle appelé classe, nous sommes tenus de calculer le **centre de classe,** noté $C_i$ qui est la valeur de la variable statistique égale à la moyenne arithmétique des valeurs des extrémités de la classe. D'où : $C_i = \frac{x_{i-1} + x_i}{2}$

**Exemple**

Une enquête sur la taille de 50 personnes donne les résultats suivant (en cm) :

158 ; 172 ; 166 ; 170 ; 168 ; 175 ; 152 ; 190 ; 191 ; 157 ; 163 ; 160 ; 149 ; 186 ;
188 ; 172 ; 173 ; 184 ; 181 ; 180 ; 170 ; 173 ; 168 ; 167 ; 169 ; 180 ; 181 ; 178 ;
166 ; 164 ; 160 ; 168 ; 166 ; 162 ; 170 ; 182 ; 183 ; 190 ; 167 ; 169.

- Classer ces données en amplitude des classes 5 cm.
- Calculer les fréquences absolues et relatives.
- Calculer les fréquences relatives cumulées croissantes.
- Calculer les centres de classe.

**Solution** : On dresse le tableau de la manière suivante.

| classes | Effectifs( $n_i$ ) | $f_i(=\frac{ni}{N})$ | $f_i \uparrow$ | $f_i \downarrow$ | Centre de classe : $C_i = \frac{x_{i-1}+x_i}{2}$ |
|---|---|---|---|---|---|
| [145 -150 [ | 1 | 0,02 | 0,02 | 1 | 147,5 |
| [150 - 155[ | 1 | 0,02 | 0,04 | 0,98 | 152,5 |
| [155-160[ | 2 | 0,04 | 0,08 | 0,96 | 157,5 |
| [160-165[ | 5 | 0,10 | 0,18 | 0,92 | 162,5 |
| [165-170[ | 12 | 0,24 | 0,42 | 0,82 | 167,5 |
| [170-175[ | 11 | 0,22 | 0,64 | 0,58 | 172,5 |
| [175-180[ | 4 | 0,08 | 0,72 | 0,36 | 177,5 |
| [180-185[ | 8 | 0,16 | 0,88 | 0,28 | 182,5 |
| [185-190[ | 2 | 0,04 | 0,92 | 0,12 | 187,5 |
| [190-195[ | 3 | 0,06 | 0,98 | 0,08 | 192,5 |
| [195-200[ | 1 | 0,02 | 1 | 0,02 | 197,5 |
| Total | 50 | 1 | Moins de la borne supérieure | Plus de la borne inférieure | - |

L'interprétation des fréquences cumulées :

- Par rapport à la colonne $f_i \uparrow$ : par exemple 64% des individus mesurent moins de 175 cm.
- Par rapport à la colonne $f_i \downarrow$ : par exemple 8% des individus mesurent plus de 190 cm.

## 2. Représentation graphique d'une série statistique

La représentation graphique d'une série statistique, permet de renseigner immédiatement sur l'allure générale de la distribution et donne une image plus claire à l'ensemble des données recueillies.

Les représentations graphiques sont diverses et dépendent principalement du type de caractère étudié :

**2.1. Cas d'un caractère qualitatif :** On distingue généralement deux types de représentations graphiques utilisées dans le caractère qualitatif : le diagramme en tuyaux d'orgue et le diagramme à secteurs circulaires.

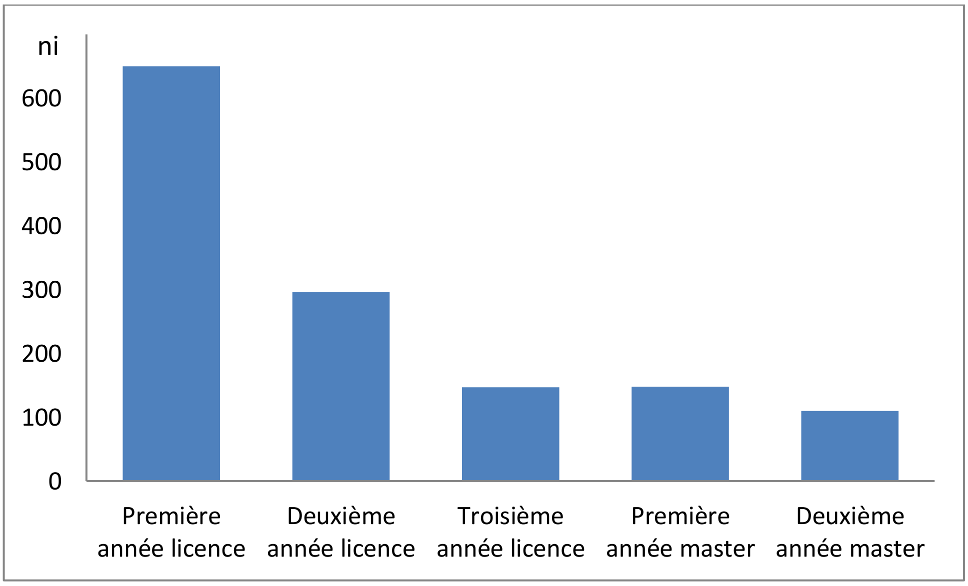
- **Le diagramme en tuyaux d'orgue** : Il s'agit de représenter les modalités de la variable par des rectangles verticaux dont les hauteurs sont proportionnelles aux effectifs des modalités (ou fréquences relatives).

**Exemple :** le tableau ci-après donne la répartition des étudiants d'une faculté en fonction de leur niveau d'étude.

| Niveau d'étude (xi) | Nombre des étudiants (ni) |
|---|---|
| Première année licence | 650 |
| Deuxième année licence | 296 |
| Troisième année licence | 147 |
| Première année master | 148 |
| Deuxième année master | 110 |

- Présenter graphiquement cette série par le diagramme en tuyaux d'orgue.

**Solution** :




- **Le diagramme circulaire :** Dans ce cas, chaque modalité est représentée par une portion de cercle proportionnelle à l'effectif de la modalité. Par conséquent chaque secteur a un angle au centre ($\alpha_i$) proportionnel à l'effectif de la modalité qu'il représente, qui s'obtient tout simplement de la règle suivante :

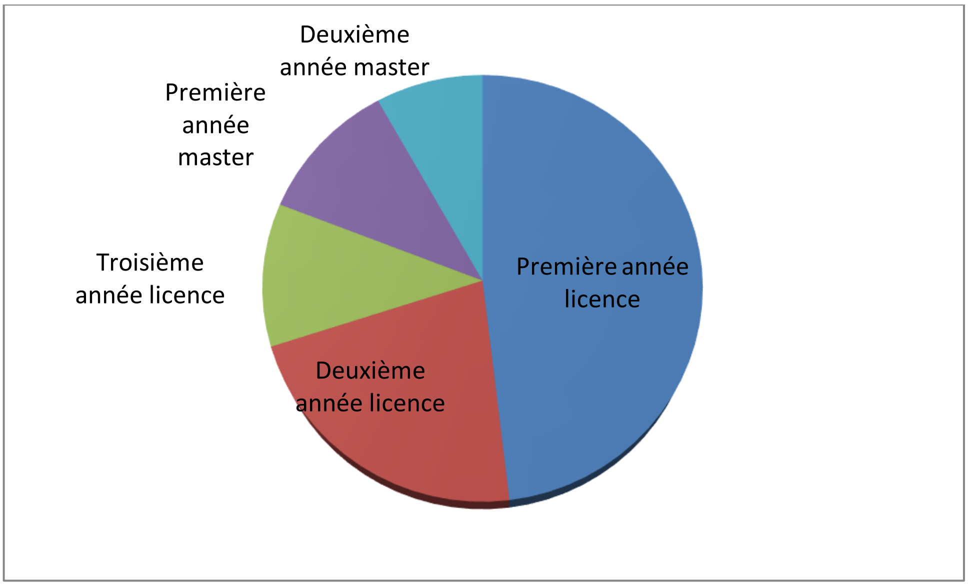
L'angle au centre ($\alpha_i$°) = $f_i \times 360^0$.

**Exemple** : le même exemple précédent, présenter graphiquement la série par le diagramme circulaire.

**Solution** : on doit construire alors le tableau de la manière suivante :

| Niveau d'étude (xi) | Nombre des étudiants (ni) | $f_i$ | $f_i \times 360^0$ |
|---|---|---|---|
| Première année licence | 650 | 0,48 | 172,8 |
| Deuxième année licence | 296 | 0,22 | 79,2 |
| Troisième année licence | 147 | 0,11 | 39,6 |
| Première année master | 148 | 0,11 | 39,6 |
| Deuxième année master | 110 | 0,08 | 28,8 |
| Total | 1351 | 1 | 360 |

**Présentation par le diagramme circulaire**




### 2.2. Cas d'un caractère quantitatif discret

Les représentations graphiques utilisées dans ce type de caractère sont les suivantes : le diagramme en bâtons et le diagramme cumulatif.

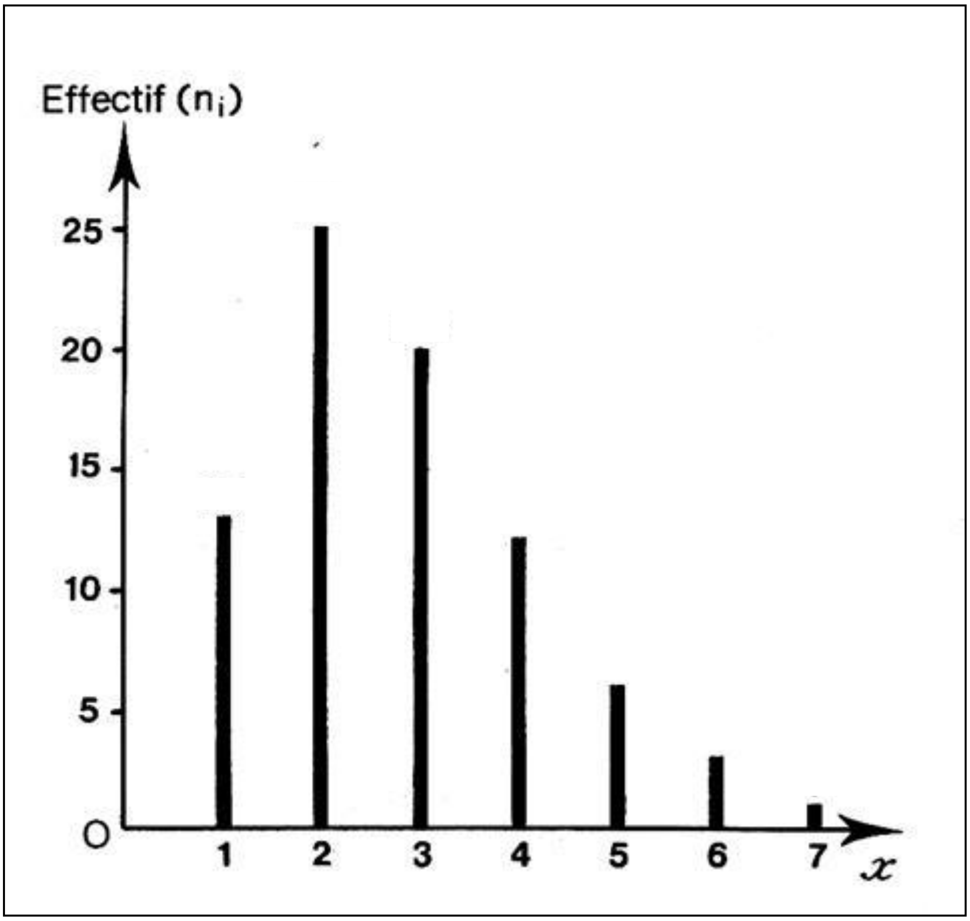
- **Diagramme en bâtons :** Est une représentation graphique d'une variable statistique dont la hauteur des bâtons correspond à l'effectif ni (ou fréquence fi) associe à chaque modalité du caractère xi. Les valeurs xi sont portées sur l'axe des abscisses et les valeurs ni (ou fi) sont portées sur l'axe des ordonnées.

**Exemple** : Le tableau suivant donne la répartition des ménages suivant le nombre de personnes possédées.

| Nombre de personnes (xi) | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Nombre de ménages (ni) | 13 | 25 | 20 | 12 | 6 | 3 | 1 |

- Présenter la série graphiquement.

**Solution :** la présentation graphique appropriée à cette série est le diagramme en bâtons.



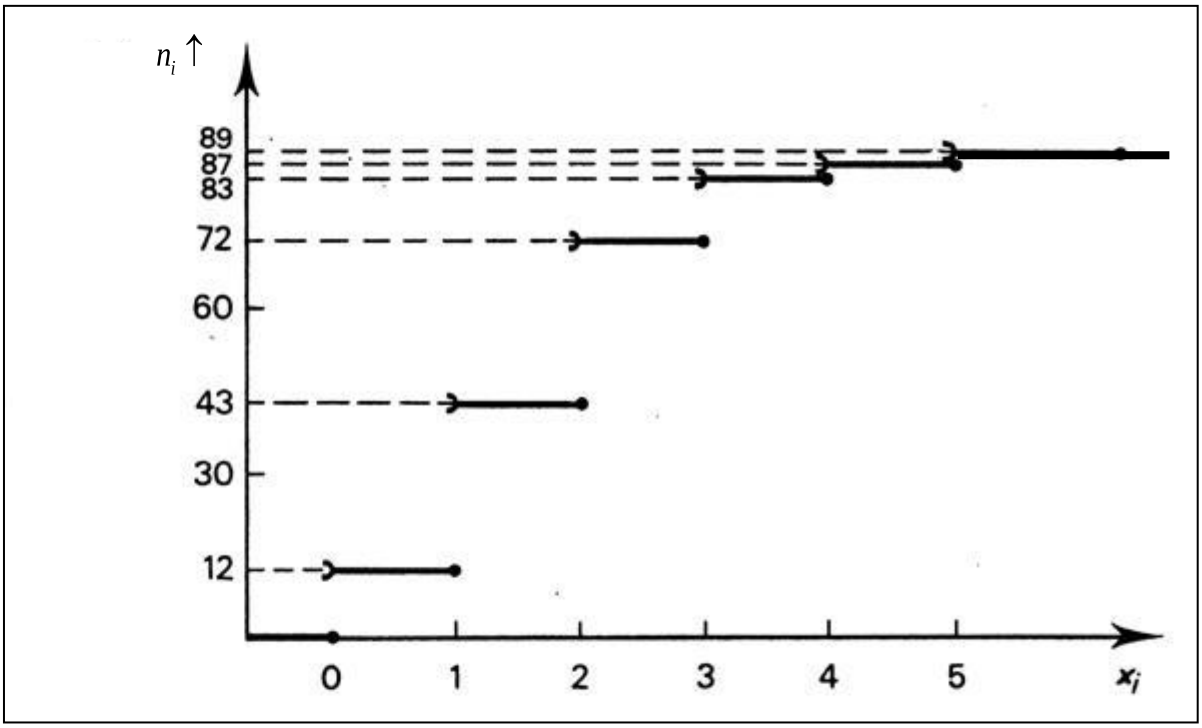

- **Diagramme cumulatif :** C'est un diagramme en escalier dont les paliers sont horizontaux, il permet de visualiser l'évolution des effectifs cumulés ou les fréquences relatives cumulées associées aux valeurs du caractère étudié.

**Exemple** : Le tableau suivant donne la répartition des foyers en fonction du nombre d'écran (téléviseur ou ordinateur).

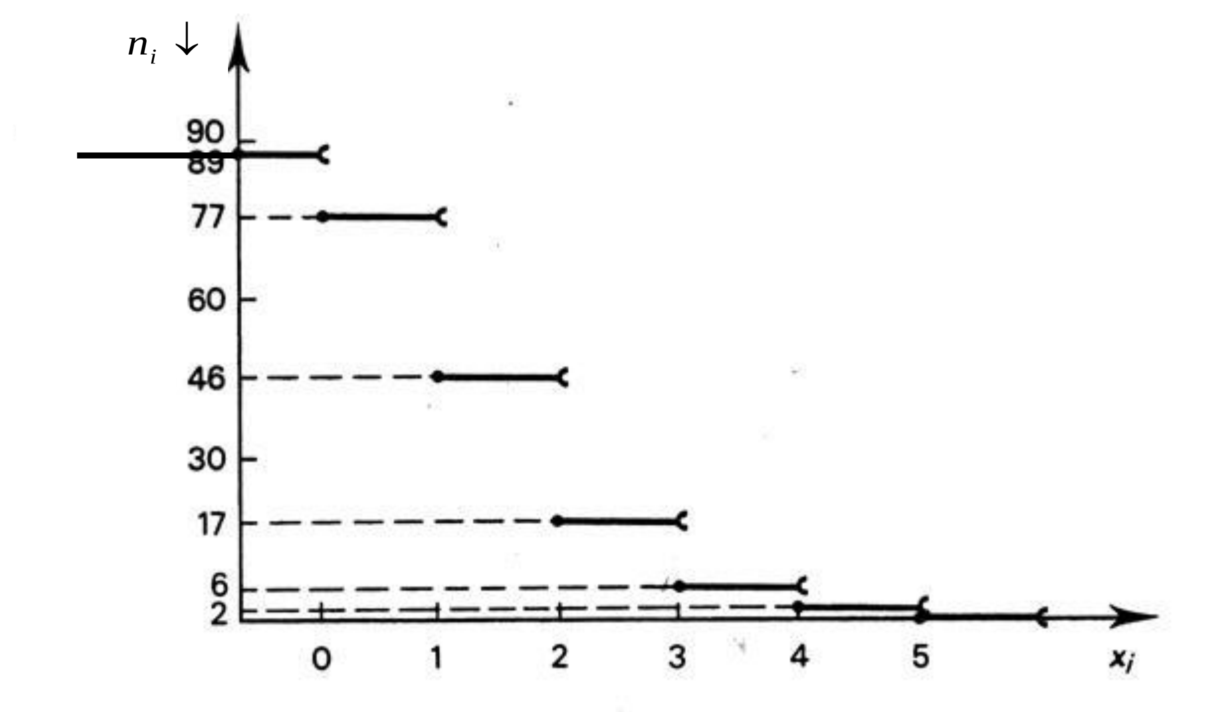
- Tracer les diagrammes cumulatifs des effectifs croissants et décroissants

| $x_i$ | $n_i$ | $n_i \uparrow$ | $n_i \downarrow$ |
|---|---|---|---|
| 0 | 12 | 12 | 89 |
| 1 | 31 | 43 | 77 |
| 2 | 29 | 72 | 46 |
| 3 | 11 | 83 | 17 |
| 4 | 4 | 87 | 6 |
| 5 | 2 | 89 | 2 |
| **Total** | 89 | - | - |

- **Le diagramme cumulatif des effectifs croissant ( $n_i \uparrow$ ).**



- **Le diagramme cumulatif des effectifs décroissants ($n_i \downarrow$).**




**2.3. Cas d'un caractère quantitatif continu**

Les principaux graphiques utilisés pour tracer le caractère quantitatif continu sont les suivants : l'histogramme et les courbes cumulatives.

- **Histogramme :** C'est un ensemble de rectangles accolés, chaque rectangle correspond à l'amplitude d'une classe et dont la hauteur est proportionnelle à l'effectif (ou fréquence relative) de la classe. Deux cas peuvent se présenter lors de la construction de l'histogramme :

- **Histogramme avec des amplitudes égales :** Dans ce cas l'histogramme se construit directement à partir les valeurs des effectifs (ni) (ou fréquence relative fi). Les hauteurs des rectangles sont proportionnelles aux effectifs (ou fréquence relative) sur l'axe des ordonnées.

- **Histogramme aves des amplitudes inégales :** Dans ce cas les classes ont des amplitudes différentes. A cet effet, il faut effectuer des corrections pour tenir compte des différences d'amplitude. Il convient de diviser les effectifs par leurs amplitudes correspondantes ( $nic = \frac{ni}{ai}$ ) (ou les fréquences relatives corrigées,

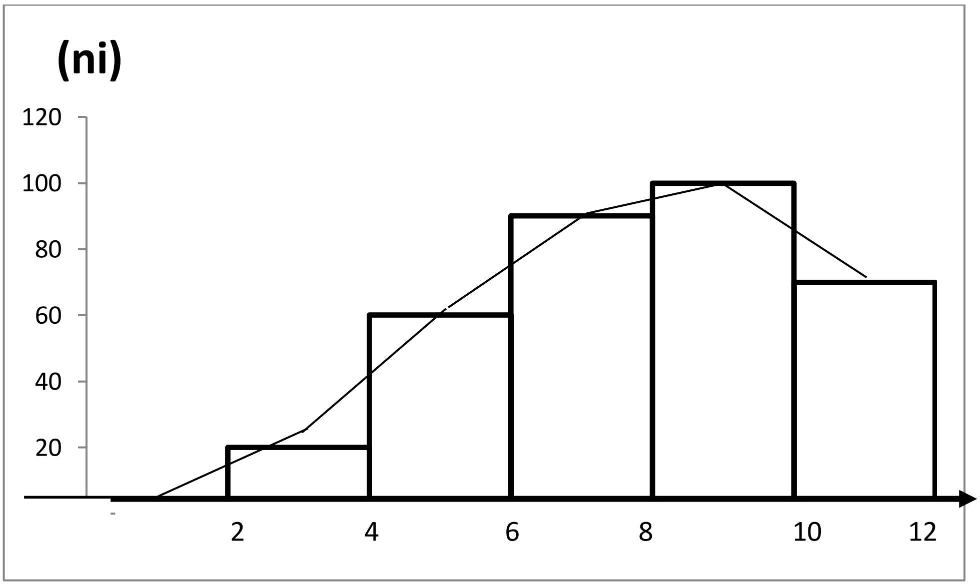
$fic = \dfrac{fi}{ai}$). Dans ce cas l'histogramme se construit à partir les effectifs corrigés (ou les fréquences relatives corrigées). Les hauteurs des rectangles sont proportionnelles aux effectifs corrigés (nic) (ou aux fréquences relatives corrigées) sur l'axe des ordonnées.

**Exemple1 (cas des amplitudes égales) :** les notes obtenues à l'examen de statistique par les étudiants sont présentées dans le tableau suivant.

| Notes ($x_i$) | Nombre d'étudiants (ni) |
|---|---|
| [2 - 4[ | 20 |
| [4 - 6[ | 60 |
| [6 - 8[ | 90 |
| [8 -10[ | 100 |
| [10 - 12[ | 70 |

- Présenter graphiquement la série statistique.

**Solution :** Présentation graphique par histogramme

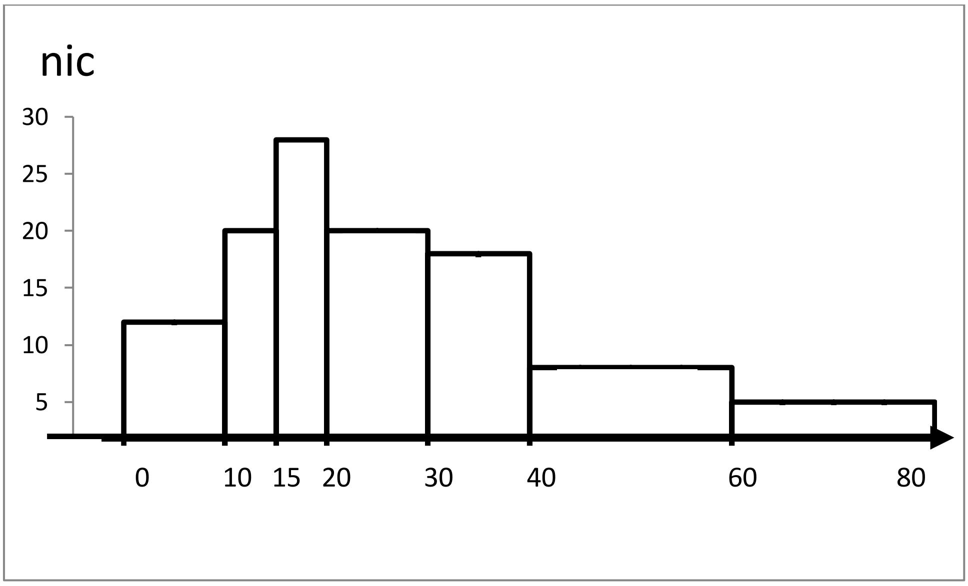



**Exemple (cas des amplitudes inégales) :** on considère une population de 1000 individus répartis en fonction de leur âge.

| Ages (xi) | Nombre ( $n_i$ ) | $a_i$ | $nic\ (=\frac{ni}{ai})$ |
|---|---|---|---|
| [0 - 10[ | 120 | 10 | 12 |
| [10 - 15[ | 100 | 5 | 20 |
| [15 - 20[ | 140 | 5 | 28 |
| [20 - 30[ | 200 | 10 | 20 |
| [30 - 40[ | 180 | 10 | 18 |
| [40-60[ | 160 | 20 | 8 |
| [60-80[ | 100 | 20 | 5 |

- Présenter graphiquement la série statistique.

**Solution** : Comme les amplitudes ( $a_i$ ) sont inégales, donc on remplace les effectifs (ni) par les effectifs corrigés (nic) pour tracer l'histogramme.




- **Les courbes cumulatives :** La courbe cumulative est une représentation graphique de la distribution des effectifs cumulés ou fréquences cumulées. Il existe deux types de courbes cumulatives :

  - Courbe cumulative croissante : elle permet de connaitre le nombre d'observations inférieures à une valeur donnée.

  - Courbe cumulative décroissante : elle permet de connaitre le nombre d'observations supérieures à une valeur donnée.

Dans ce cas, les courbes se tracent sous quelques points :

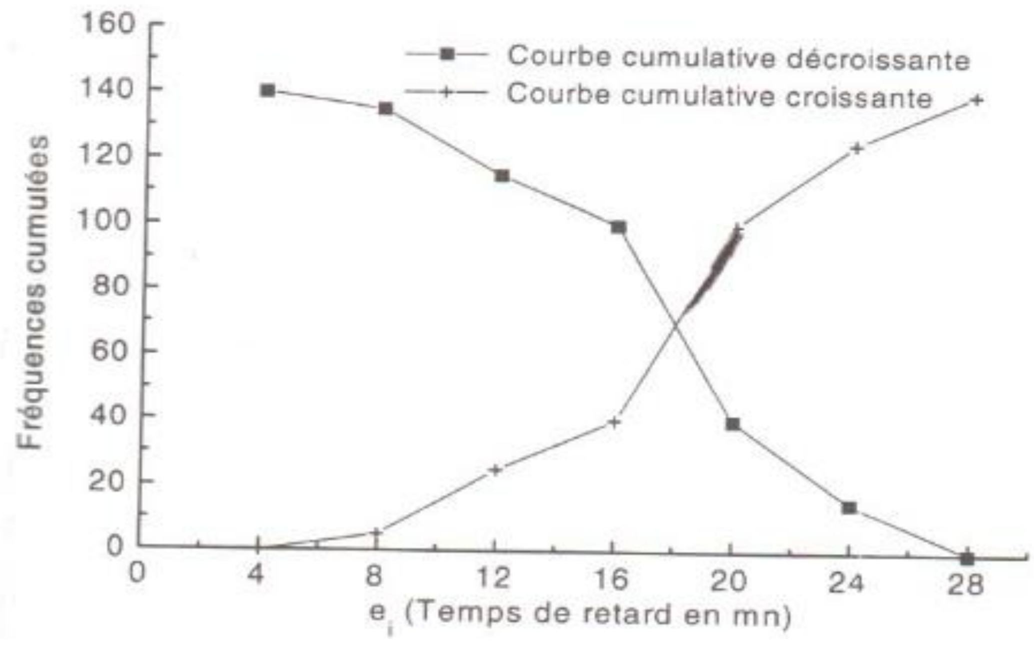
- Pour la courbe croissante, on joint les points de coordonnées $(xi,\ n_i \uparrow)$ c'est-à-dire : en abscisse, les valeurs de la borne supérieure $(xi)$ de chaque classe. En ordonnée, les valeurs de l'effectif cumulé $n_i \uparrow$ correspondent à chaque classe.
- Pour la courbe décroissante, on joint les points de coordonnées $(x_{i-1}, n_i \downarrow)$ , c'est-à-dire : en abscisse, les valeurs de la borne inférieure $x_{i-1}$ de chaque classe. En ordonnée, les valeurs de l'effectif cumulé $n_i \downarrow$ correspondent à chaque classe.

**Exemple :** Soit la distribution suivante qui indique la répartition des 140 ouvriers selon leur temps de retard (le retard est exprimé en minutes).

| classes | $n_i$ | $n_i \uparrow$ | $n_i \downarrow$ |
|---|---|---|---|
| 4-8 | 5 | 5 | 140 |
| 8-12 | 20 | 25 | 135 |
| 12-16 | 15 | 40 | 115 |
| 16-20 | 60 | 100 | 100 |
| 20-24 | 25 | 125 | 40 |
| 24-28 | 15 | 140 | 15 |
| **Total** | 140 | - | - |

- Tracer les courbes cumulatives des effectifs.

**Solution**



**Remarque :** La même chose pour le cas d'une variable dont les classes sont inégales, on trace les courbes cumulatives sans corriger les effectifs.

**Exercices**

**Exercice 1 :** La répartition des travailleurs d'une entreprise selon la qualification a été résumée dans le tableau ci-après.

| Qualification | Nombre de travailleurs |
|---|---|
| Ouvriers | 140 |
| Employés | 30 |
| Techniciens | 20 |
| Ingénieurs | 10 |

1- Quelle est la population statistique étudiée ?
2- Quelle est l'unité statistique ?
3- Quel est le caractère étudié et son type ?
4- Construire le tableau des fréquences relatives et les graphiques appropriés.

**Exercice 2 :** Un contrôle impromptu effectué par une équipe de gendarmes permet de relever de nombreuses infractions au code de la route et à la sécurité des usagers.

| Nombre d'infractions | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Nombre de conducteurs | 44 | 140 | 92 | 156 | 256 | 80 | 32 |

1- Déterminer la population statistique étudiée.
2- Quel est le caractère étudié et son type ?
3- Représenter graphiquement cette série par un diagramme en bâtons.
4- Calculer les effectifs et les fréquences cumulées.
5- Tracer le diagramme cumulatif des effectifs cumulés (croissant et décroissant.

**Exercice 3 :** On considère l'ensemble des notes obtenues, lors d'un test noté sur 20, par les candidats :10 ;8 ;3 ;12 ;13 ;9 ;12 ;9 ;12 ;11 ;11 ;8 ;5 ;13 ;14 ;14 ;6 ;12 ;7 ;11 ;10 ;15 ;12 ;10 ;1 4 ;11 ;7 ;8 ;10 ;13 ;9,7 ;13 ;11 ;9 ;4 ;10 ;8 ;9 ;6 ;7 ;14.

1- Quelle est la population statistique et le caractère étudié ?
2- Classer ces données en classes d'amplitudes 5.
3- Construire le tableau des fréquences et le représenter graphiquement.
4- Calculer les fréquences relatives cumulées.
5- Quelle est la proportion des candidats ayant une note inférieure à 16 ?
6- Quelle est la proportion des candidats ayant une note comprise entre 5 et 20 ?

## Chapitre 3 : Les paramètres de tendance centrale

Les paramètres de tendance centrale ou « mesures de tendance centrale » sont des grandeurs susceptibles de représenter au mieux un ensemble de données, ils s'expriment dans la même unité que les valeurs des variables. L'appellation «tendance centrale » vient du fait que ces paramètres donnent une idée de ce qui se passe au centre d'une distribution, d'un ensemble de données. On distingue trois mesures de tendance centrale : le mode, la médiane et la moyenne arithmétique. Tous ces trois paramètres ne décrivent par la même chose et sont, de ce fait, complémentaires dans la description et l'analyse d'une distribution.

### 1. Le Mode

Le mode, noté $\boldsymbol{Mo}$ d'une série statistique est la valeur du caractère la plus fréquente ou dominante dans l'échantillon. Autrement dit, c'est la valeur de $x_i$ qui a la fréquence (absolue ou relative) la plus grande.
Lorsque la distribution a plus d'un mode, on parle d'une distribution **« multimodale »** **(**bimodale, trimodale, etc).

Le calcul du mode d'une série statistique dépend de la nature de la variable étudiée :

**1.1. Cas d'une variable discrète :** Lorsque la variable statistique est discrète, le mode est la valeur de la variable xi possédant le plus grand effectif. Il est donc directement observable dans une série statistique.

Graphiquement le mode est la valeur de xi qui correspond au bâton le plus haut. Sa valeur est donnée par l'axe des abscisses.

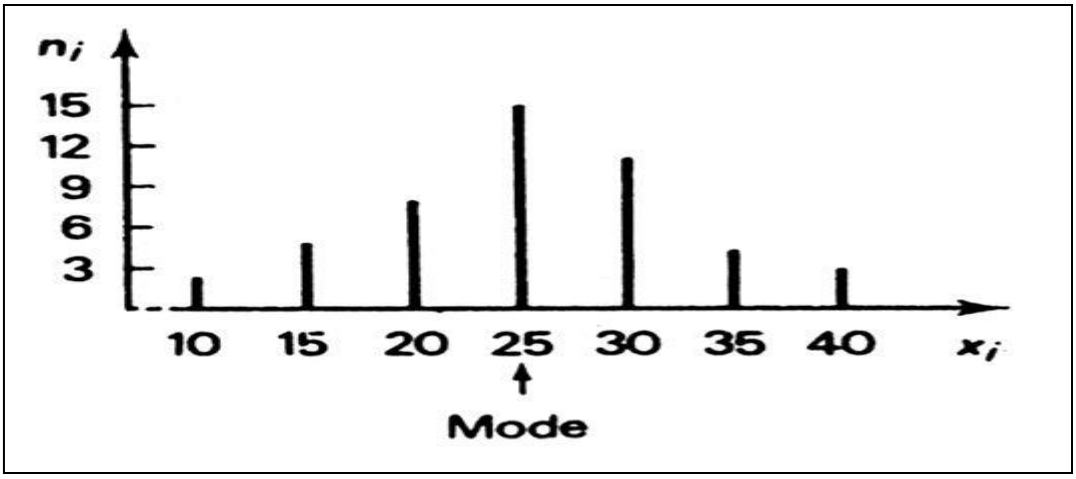
**Exemple1 :** Soit le tableau suivant.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x_i$ | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
| $n_i$ | 2 | 3 | 8 | 17 | 13 | 5 | 3 |

Le mode : $\mathbf{M_O}$**= 25,** car il correspond à la valeur maximale des effectifs (maxi $n_i$ =17).

Dans ce cas le mode signifie que la majorité des individus ont le caractère xi= 25.

Le mode graphiquement correspond au bâton le plus élevé, qui porte la valeur 25 sur l'axe des abscisses, voir le graphe suivant.




**1.2. Cas d'une variable continue :** Dans le cas continu, les valeurs prises par la variable étant regroupées en classe, on parle alors dans ce cas de la **classe modale**, qui correspond à l'effectif le plus grand. Puis le mode sera calculé par la formule suivante :

$$Mo = X_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i$$ Sachant :

$Mo$ : La valeur du mode.

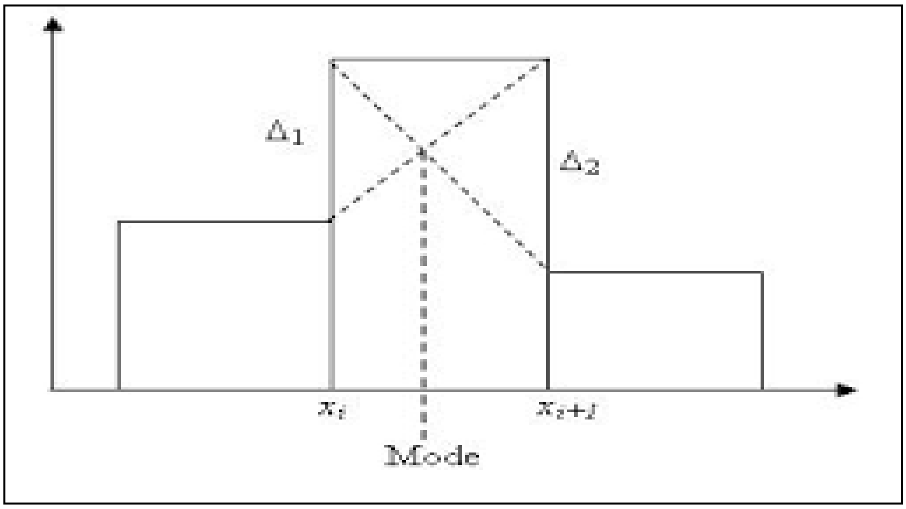
$X_{\min}$ : La valeur inférieure de la classe modale.

$\Delta_1$ : La différence entre les effectifs (ni) de la classe modale et ceux de la classe précédente.

$\Delta_2$ : La différence entre les effectifs (ni) de la classe modale et ceux de la classe suivante.

$a_i$ : L'amplitude de la classe modale.

Graphiquement la valeur du mode est déterminée sur l'histogramme, tout d'abord on retiendra le rectangle le pus haut associé à la classe modale, puis on applique la méthode des diagonales, comme on le voit sur le schéma suivant.




**Remarque importante :** Dans le cas où les amplitudes sont inégales, on remplace les effectifs (ni) par les effectifs corrigés ( $nic = \frac{ni}{ai}$ ) et on utilise les mêmes procédures précédentes pour calculer le mode. Le même cas pour la présentation graphique, on remplace les effectifs (ni) par les effectifs corrigés (nic) sur l'axe des ordonnés, et on poursuit les mêmes procédures précédentes pour déterminer le mode graphiquement.

**Exemple (cas des amplitudes égales) :** Soit le tableau suivant

| $x_i$ | **ni** |
|---|---|
| [100-110[ | 8 |
| [110-120[ | 22 |
| [120-130[ | 38 |
| [130-140[ | 12 |
| [140-150[ | 6 |
| **Total** | **86** |

**Plus grand effectif**

**Classe modale**

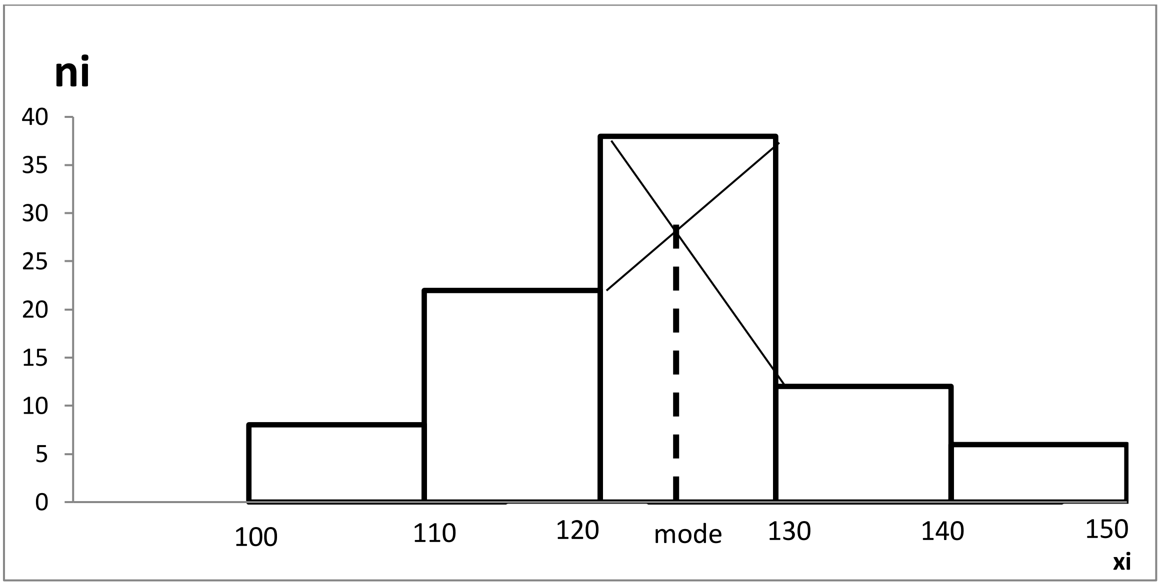
- Calculer le mode et le déterminer graphiquement.

**Solution**

- On détermine tout d'abord la classe modale, à travers le tableau on remarque que l'effectif le plus grand est 38, donc la classe modale est [120-130[.
- Maintenant on applique la formule suivante pour calculer le mode :

$$Mo = X_{min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i = 120 + \frac{38-22}{(38-22)+(38-12)} \times 10 = 123{,}81$$

Graphiquement, le mode est déterminé sur l'histogramme ci-après :




**Exemple2 (cas des amplitudes inégales) :**

La série suivante représente la répartition des exploitations agricoles selon leurs superficies évaluées en Hectare.

| **Classes** | $n_i$ | $a_i$ | $n_i$ corrigés (= ni/ai**)** |
|---|---|---|---|
| [0 - 1[ | 35 | 1 | 35 |
| [1 - 2[ | 30 | 1 | 30 |
| [2 - 3[ | 25 | 1 | 25 |
| [3 - 5[ | 54 | 2 | 27 |
| [5 - 10[ | 100 | 5 | 20 |
| [10-20[ | 80 | 10 | 8 |
| [20-22[ | 6 | 2 | 3 |
| total | 330 | - | - |

**Classe modale**

**Plus grand effectif corrigé (nic)**

**Solution**

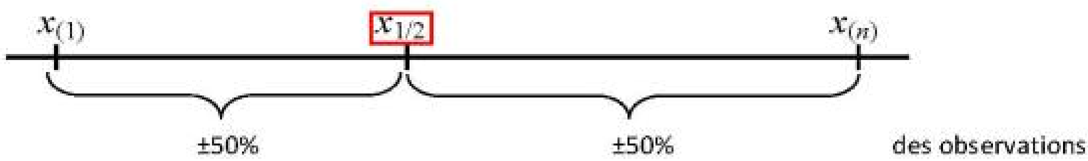
Comme les amplitudes ($a_i$) sont inégales donc la classe modale correspond à l'effectif $n_i$ corrigé (nic) le plus grand, qui est égal à 35. D'où la classe modale est : [0 - 1[

Donc on applique la formule pour calculer le mode :

$$Mo = X_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i \Rightarrow Mo = 0 + \frac{(35-0)}{(35-0)+(35-30)} \times 1 = 0{,}875 \text{ HA}$$

**2. La Médiane (notée Mé)**

La médiane (Mé) est la valeur du caractère ($x_i$), qui partage la série statistique en deux parties égales. Sachant que, les valeurs de la variable statistique étant au préalable classées dans l'ordre croissant.




Dans une distribution statistique, le calcul de la médiane se fait à partir de la colonne des effectifs cumulés croissants (ou des fréquences relatives cumulées croissantes) et en fonction de la nature de la variable statistique étudiée :

**2.1. Cas d'une variable discrète :** Dans ce cas il faut voir le nombre total des effectifs (N), deux cas se présentent[1] :

- **Si le nombre d'observations (N) est impair :**

  Dans ce cas, la médiane : $M_e = x_{\frac{n+1}{2}}$

- **Si le nombre d'observations (N) est pair :**

  Dans ce cas, la médiane : $M_e = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

Graphiquement la médiane (Mé) est déterminée sur le diagramme cumulatif croissant des effectifs(ou des fréquences relatives), c'est la valeur de xi située sur l'axe des abscisses à laquelle est associée à la valeur N/2 (ou 0,5) sur l'axe des ordonnées.

[1] Cette application correspond au sériés pondérées (les plus utilisées dans la pratique)

**Exemple** : Le tableau ci-dessous présente la distribution observée des âges des répondants à une enquête :

| Age du répondant (xi) | Effectif (ni) | $n_i \uparrow$ |
|---|---|---|
| 17 | 18 | 18 |
| 18 | 85 | 103 |
| 19 | 99 | 202 |
| 20 | 52 | 254 |
| 21 | 31 | 285 |
| 22 | 15 | 300 |
| Total | 300 | - |

- Calculer la médiane et la déterminer graphiquement.

**Solution :** N=300 est pair, donc la médiane se calcule par la formule suivante :

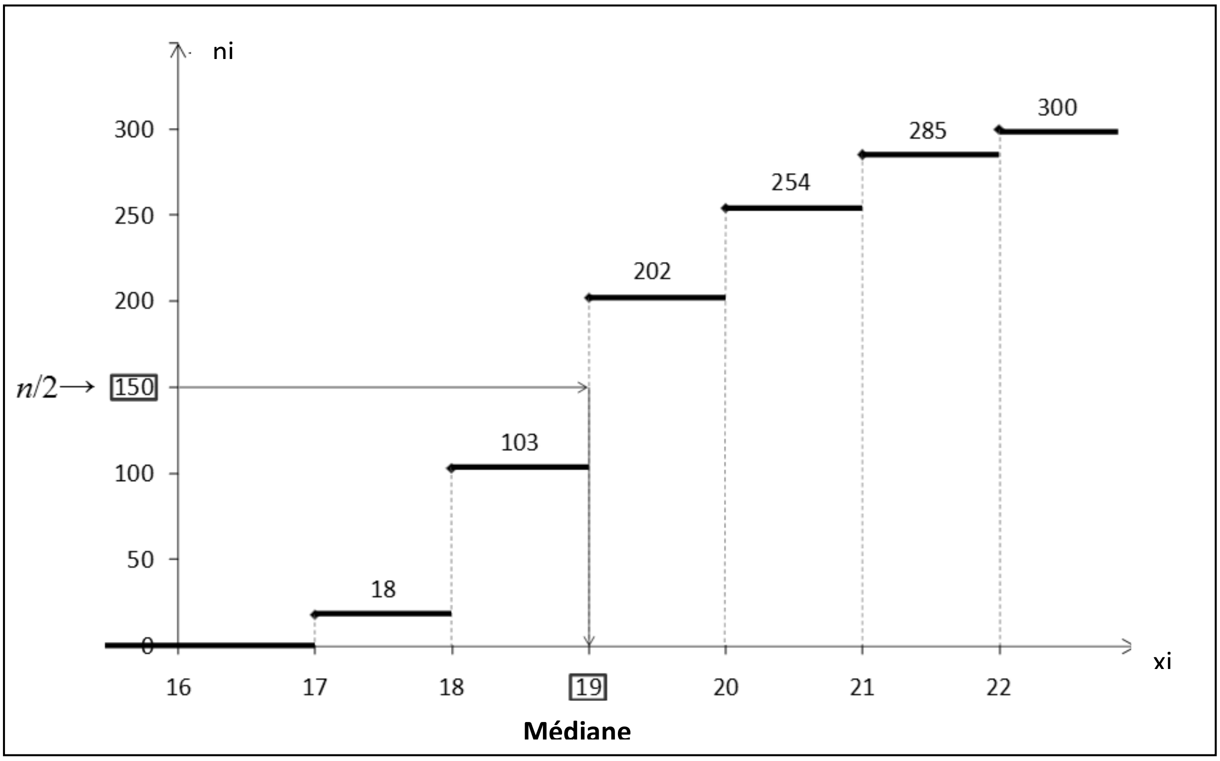
$$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \quad \Rightarrow \quad M_e = \frac{x_{\frac{300}{2}} + x_{\frac{300}{2}+1}}{2} = \frac{x_{150} + x_{151}}{2}$$

On remarque que les valeurs $x_{150}$ et $x_{151}$ sont situées dans la colonne des effectifs cumulée croissantes ( $n_i \uparrow$ ) entre les valeurs 103 et 202, cette dernière valeur correspond à la valeur 19 sur la première colonne (xi).

Donc la médiane : $M_e = \frac{19+19}{2} = 19.$

Cette valeur signifie que 50% des répondants sont âgés de moins de 19 ans.

Graphiquement la médiane se détermine sur le diagramme en escalier, voir le schéma suivant.




### 2.2. Cas d'une variable continue

Dans ce cas, pour calculer la médiane, il faut déterminer au préalable la classe médiane, c'est la classe qui correspond à la valeur N/2 sur la colonne $N\uparrow$ ou bien à la valeur 0.5 sur la colonne $F\uparrow$. Puis on applique la formule suivante pour calculer la valeur de la médiane :

$$Mé = X_{\min} + \frac{N/2 - n_{i\,précédent}\uparrow}{n_{i\,Me}} \times a_i$$

Sachant :

$Mé$ : La valeur de la médiane.

$X_{\min}$ : La borne inférieure de la classe médiane.

$n_{i\,précédent}\uparrow$ : L'effectif cumulé croissant de la classe située avant la classe médiane.

$n_{i\,Me}$ : L'effectif de la classe médiane.

$a_i$ : L'amplitude de la classe médiane.

**Remarque :** Dans le cas où nous utilisons les fréquences relatives cumulées croissantes, le calcul de la médiane se fait par la formule suivante : $Mé = X_{\min} + \frac{0,5 - F_{i\,précédent}^{\uparrow}}{fi_{Mé}} \times ai$

**Exemple :** Soit le tableau suivant, calculer la médiane.

| xi | $n_i$ | $f_i$ | $f_i \uparrow$ |
|---|---|---|---|
| [3 - 4[ | 12 | 0,024 | 0.024 |
| [4 - 6[ | 30 | 0,061 | 0.085 |
| [6 - 8[ | 120 | 0,246 | 0.331 |
| [8 - 9[ | 210 | 0,431 | 0,762 |
| [9- 10[ | 90 | 0,184 | 0,946 |
| [10- 12[ | 25 | 0,051 | $\approx 1$ |
| Total | 487 | $\approx 1$ | - |

**Solution :** Pour nos calculs, nous avons retenu les fréquences relatives cumulées croissantes.

- Mé $(\alpha = 0{,}5)$ , cette valeur est située dans la colonne des fréquences relatives cumulées croissantes ( $f_i \uparrow$ ) entre les valeurs 0,331 et 0,762, cette dernière correspond à la classe $[8;9[$. Donc $Mé \in [8;9[ \quad \Rightarrow Mé = X_{\min} + \dfrac{\alpha - F_i^{\uparrow}{}_{précédent}}{fi_{Mé}} \times ai = 8 + \dfrac{0{,}5 - 0{,}331}{0{,}431} \times 1 = 8{,}39$

### 3. La Moyenne arithmétique

La moyenne arithmétique d'une distribution statistique pour une variable X, est égale au rapport de la somme des valeurs prises par cette variable sur le nombre d'observations. On la note généralement $\overline{X}$ et se lit « X barre »

Elle peut être simple ou pondérée :

- **La moyenne arithmétique simple** : on dit la moyenne arithmétique est simple lorsque les valeurs prises par la variable X n'apparaissent qu'une seule fois. Dans ce cas chaque effectif ni est égal à 1.

$$\overline{X} = \frac{x1 + x2 + ...... + xk}{n1 + n2 + ...... + nk} \quad \Rightarrow \quad \overline{X} = \frac{\sum_{i=1}^{k} x_i}{N}$$

**Exemple :** 8 étudiants ont obtenu les notes suivantes : 3, 5, 7, 9, 10, 11, 12, 18.

La note moyenne est : $\overline{X} = \dfrac{\sum_{i=1}^{n} x_i}{N} = \dfrac{3+5+7+9+10+11+12+18}{8} = 9.37$

- **La moyenne arithmétique pondérée :** on dit la moyenne arithmétique est pondérée lorsque les valeurs prises par la variable X peuvent correspondre à plusieurs valeurs d'effectifs ni.

$$\overline{X} = \frac{n1 * x1 + n2 * x2 + ...... + nk * xk}{n1 + n2 + ...... + nk} \Rightarrow \overline{X} = \frac{\sum_{i=1}^{k} n_i x_i}{N}$$

Ou, en travaillant avec des fréquences relatives au lieu des effectifs :

$$\overline{X} = \sum_{i=1}^{n} f_i x_i \quad avec \quad f_i = \frac{n_i}{N} \quad et \quad \sum_{i}^{n} f_i = 1$$

**Exemple :** 8 étudiants ont obtenu les notes suivantes : 3, 3, 3, 5, 9, 9, 11, 11.

La note moyenne est :

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \frac{n1 * x1 + n2 * x2 + ...... + nk * xk}{n1 + n2 + ...... + nk} = \frac{3*3 + 1*5 + 2*9 + 2*11}{8} = 6.75$$

**Comment calculer la moyenne arithmétique ?**

**3.1. Cas d'une variable discrète :** Dans le cas d'une variable discrète, le calcul de la moyenne arithmétique se fait directement à partir les modalités de la variable étudiée. On ajoute au tableau statistique une nouvelle colonne contenant les produits des valeurs de la variable et celles des effectifs ($n_i \cdot x_i$). Puis on utilise la formule suivante pour calculer la moyenne arithmétique : $\overline{X} = \frac{\sum_{i=1}^{n} n_i \cdot x_i}{\sum n_i}$

**Exemple :** Soit le tableau suivant

| $X_i$ | $n_i$ | $n_i \cdot x_i$ |
|---|---|---|
| 18 | 6 | 108 |
| 19 | 10 | 190 |
| 20 | 4 | 80 |
| 21 | 2 | 42 |
| 22 | 2 | 44 |
| Total | 24 | 464 |

$$\overline{X} = \frac{\sum_{i=1}^{7} n_i \cdot x_i}{N} = \frac{464}{24} = 19{,}33$$

**3.2. Cas d'une variable continue :** Dans ce cas, les modalités de la variable sont alors des classes et par conséquent on utilise dans les calculs de la moyenne les centres de classes. Dans ce cas on ajoute au tableau statistique deux nouvelles colonnes, l'une relative aux centres de classe $c_i$ et l'autre colonne contenant les produits des valeurs de la variable et celles des effectifs ($n_i \times c_i$). Puis on utilise la formule suivante pour calculer la moyenne arithmétique :

$$\overline{X} = \frac{\sum_i n_i \times c_i}{N}$$

**Exemple :** Soit la distribution d'un échantillon de 24 étudiants selon leur taille.

| X | $n_i$ | $c_i$ (centre de classe) | $n_i \times c_i$ |
|---|---|---|---|
| [1.5 ; 1.6 [ | 6 | 1.55 | 9.3 |
| [1.6 ; 1.7 [ | 7 | 1.65 | 11.55 |
| [1.7 ; 1.8 [ | 8 | 1.75 | 14 |
| [1.8 ; 1.9 [ | 2 | 1.85 | 3.7 |
| [1.9 ; 2 [ | 1 | 1.95 | 1.95 |
| Total | 24 | - | 40.5 |

D'où la moyenne arithmétique : $\overline{X} = \frac{\sum_i n_i \times c_i}{N} = \frac{40{,}5}{24} \approx 1{,}687$

**Propriétés de la moyenne arithmétique :**

- Si on définit une nouvelle variable Z telle que : $Z_i = a\,x_i + b$ , sachant a et b sont des constantes réelles, alors : $\bar{Z} = a\,\bar{X} + b$

  Cela signifie que si on multiplie toutes les valeurs de la variable par une constante a, la moyenne sera elle aussi multipliée par a, et, si on ajoute à toutes les valeurs de la variable une constante b, la moyenne augmentera elle aussi d'une valeur égale à b.

La même chose lorsque : $Z_i = a\,x_i - b \quad alors \quad \bar{Z} = a\,\bar{X} - b$

### 4. Généralisation de la notion de moyenne : les autres types de moyennes

Nous pouvons généraliser la notion de moyenne en examinant la moyenne géométrique, harmonique et quadratique, dont les définitions précédentes du même principe.

**4.1. La moyenne géométrique (notée G) :** La moyenne géométrique d'une distribution statistique pour une variable X est égale à la racine n-ième du produit des n valeurs positives de cette variable. . On la note généralement : $G_X$. Elle est utilisée principalement dans le calcul de taux de croissance moyens. Sa formule de calcul est donnée dans les deux cas suivants :

- **Cas de la moyenne géométrique simple :**

$$G = \sqrt[N]{X_1 \times X_2 \times X_3 \times \ldots \times X_k} = \sqrt[N]{\prod_{i=1}^{k} X_i} = \left(\prod_{i=1}^{k} X_i\right)^{\frac{1}{N}} .$$

- **Cas de la moyenne géométrique pondérée :** Dans ce cas, la formule de la moyenne géométrique devient :

$$G = \sqrt[N]{X_1^{n_1} \times X_2^{n_2} \times \ldots . \times X_k^{n_k}}$$

**Remarque :** le taux de croissance moyen $Tx$ : est égal : $Tx = (G-1) \times 100$

**Exemple** :

Une société a vu son bénéfice augmenter ces trois dernières années de 10 % la première année, 21% la deuxième année et 3% la troisième année.

- Calculer la moyenne géométrique.
- Calculer le taux de croissance moyen de bénéfice.

**Solution**

- La moyenne géométrique : $G = \sqrt[N]{X_1^{n_1} \times X_2^{n_2} \times \ldots . \times X_k^{n_k}}$

On a : $X_i = 1 + t_i$, sachant $t_i$ est le taux de croissance. Donc $G = \sqrt[3]{{}_1^{n_1} \times X_2^{n_2} \times \ldots . \times X_k^{n_k}}$

$$= \sqrt[3]{(1{,}1)^1 \times (1{,}21)^1 \times (1{,}02)^1} = 1{,}1071$$

- Le taux de croissance moyen du bénéfice :

Au cours de cette période, le taux moyen de croissance *Tx* est : $Tx = (G-1)\times 100 = (1{,}1071-1)\times 100 = 10{,}72\%$

**4-2- La moyenne harmonique (notée H) :** La moyenne harmonique d'une distribution statistique pour une variable X est égale à l'inverse de la moyenne arithmétique des valeurs de cette variable. On la note généralement : $H_X$. Elle est utilisée principalement dans le calcul de moyennes de rapports, notamment de vitesses moyennes et d'autres indices statistiques (km/h, DA/minute, quintaux /hectare,… etc.). Sa formule de calcul est donnée dans les deux cas suivants :

- **Cas de la moyenne harmonique simple :** $H = \dfrac{N}{\sum_i \dfrac{1}{X_i}} = \dfrac{N}{\dfrac{1}{X_1}+\dfrac{1}{X_2}+\ldots+\dfrac{1}{X_n}}$
- **Cas de la moyenne harmonique pondérée :** $H = \dfrac{\sum n_i}{\sum_i \dfrac{n_i}{X_i}} = \dfrac{N}{\dfrac{n_1}{X_1}+\dfrac{n_2}{X_2}+\cdots+\dfrac{n_k}{X_k}}$

**Exemple :** Une voiture roule 1 heure à la vitesse de 80 km/h et parcours un tronçon de 60 km, de la suite de son trajet, à une vitesse de 120 km/h. Le trajet de retour est accompli avec une vitesse de 100km/h.

Calculer la vitesse moyenne de cette voiture.

**Solution :** La vitesse moyenne de cette voiture est une moyenne harmonique :

1- Le temps =1h et la distance (n1) = 80 km $\Rightarrow$ la vitesse (x1) = 80km/h.
2- La distance (n2) = 60 km et la vitesse (x2) = 120 km/h.
3- La distance (n3) = 80 +60 = 140 KM et la vitesse (x3) = 100km/h.

$$H = \frac{N}{\dfrac{n_1}{x_1}+\dfrac{n_2}{x_2}+\dfrac{n_3}{x_3}} \quad \text{Donc} \quad H = \frac{80+60+140}{\dfrac{80}{80}+\dfrac{60}{120}+\dfrac{140}{100}} = 96{,}55\,km/h$$

La vitesse moyenne est 96,55 km/h.

**4-3- La moyenne quadratique (notée Q) :** La moyenne quadratique d'une distribution statistique pour une variable X est égale à la racine carrée de la moyenne arithmétique des carrés des valeurs ($x_i$). On la note généralement : $Q_X$. Elle est rarement utilisée, elle trouve principalement son application dans les mesures de dispersion. Sa formule de calcul est donnée dans les deux cas suivants :

- **Cas de la moyenne quadratique simple :** $Q = \sqrt{\frac{1}{N}\sum_{i=1}^{n} X_i^2}$
- **Cas de la moyenne quadratique pondérée :** $Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i \cdot X_i^2}$

**Remarque :**

- Dans le cas continu, on retiendra dans les calculs des moyennes comme valeurs de la variable les centres de classes.
- $H \prec G \prec \overline{X} \prec Q$ : cette relation est pratiquement toujours vérifiée.

Nous pouvons vérifier cette relation en traitant l'exemple suivant

**Exemple :** soit le tableau suivant

| | | | | |
|---|---|---|---|---|
| **$x_i$** | 1 | 2 | 3 | 4 |
| **$n_i$** | 5 | 15 | 10 | 10 |

- Calculer les valeurs des moyennes arithmétique, harmonique, géométrique et quadratique.

**Solution** :

- La moyenne arithmétique : $\overline{X} = \frac{1\times 5 + 2\times 15 + 3\times 10 + 4\times 10}{40} = 2{,}62$
- La moyenne géométrique : $G = \sqrt[40]{1^5 \times 2^{15} \times 3^{10} \times 4^{10}} = 2{,}41$
- La moyenne harmonique : $H = \frac{40}{\frac{5}{1} + \frac{15}{2} + \frac{10}{3} + \frac{10}{4}} = 2{,}18$
- La moyenne quadratique : $Q = \sqrt{\frac{1^2 \times 5 + 2^2 \times 15 + 3^2 \times 10 + 4^2 \times 10}{40}} = 2{,}80$

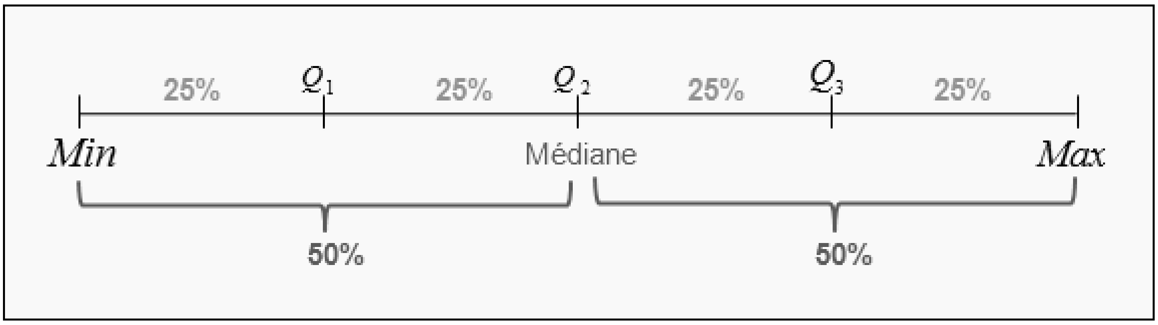
On remarque : $H \prec G \prec \overline{X} \prec Q$

## 5- Généralisation de la notion de la médiane : Les quantiles

Les quantiles constituent une généralisation de la notion de médiane qui en représente un cas particulier.

Ce sont des valeurs du caractère (xi), qui partagent la série statistique en sous groupes de même taille. Ils sont déterminés comme la médiane à partir des fréquences absolue (ou relatives) cumulées. Ils portent trois types : les quartiles, les déciles et les centiles.

**5.1. Les Quartiles :** ce sont les valeurs de xi qui partagent la série en quatre sous ensembles égaux. Ils sont de trois nombre : Q1, Q2 et Q3 et ils correspondent respectivement sur la série statistique aux valeurs : 1/4 (ou 25%), 2/4(ou 50%) et 3/4 (ou 75%), voir la figure suivante.




- le 1er quartile Q1 $\left(\alpha = \frac{1}{4}\right)$ : 25% des observations lui soient inférieures.
- le 2e quartile Q2 $\left(\alpha = \frac{2}{4}\right)$ : est la médiane, telle que 50% des observations lui soient inférieures.
- le 3e quartile Q3 $\left(\alpha = \frac{3}{4}\right)$ : telle que 75% des observations lui soient sont inférieures.

**5-2-Les Déciles :** Les déciles, sont au nombre de neuf (D1, D2,…., D9), qui divisent les effectifs de la série en 10 parties égales.

- Le 1er décile D1$\left(\alpha = \frac{1}{10}\right)$ : est la valeur de la variable telle que 10% des observations lui soient inférieures.
- ……………..
- Le 9e décile D9$\left(\alpha = \frac{9}{10}\right)$ : est la valeur de la variable telle que 90% des observations lui soient inférieures.

Le 5e décile est égal la médiane. $D_5\left(\alpha = \frac{5}{10}\right)$

**5-3-Les centiles :** Les centiles, sont au nombre de 99 (C1, C2,…., D99), qui divisent les effectifs de la série en 100 parties égales.

- Le 1er centile C1$\left(\alpha = \frac{1}{100}\right)$ : est la valeur de la variable telle que 1% des observations lui soient inférieures.
- ……………..
- Le 99e centile C99$\left(\alpha = \frac{99}{100}\right)$ : est la valeur de la variable telle que 99% des observations lui soient inférieures.

Le 50e centile $C_{50}\left(\alpha = \frac{50}{100}\right)$ correspond à la médiane. Donc : $Mé = Q_2 = D_5 = C_{50}$

**Comment calculer les quantiles ?**

Les quantiles, qu'ils soient dans le cas continu ou discret, se calculent sur le même principe que la médiane à partir des effectifs ou des fréquences relatives cumulées ; on détermine l'effectif ou la fréquence relative cumulée correspondant au quantile envisagé de paramètre $\alpha$. La suite des calculs est identique à celle de la médiane.

**Exemples d'illustration sur le calcul des quantiles**

**Exemple1 Cas d'une variable discrète) :** Soit la distribution statistique suivante qui indique la répartition du nombre d'enfants par ménage (voir le tableau ci-après).

Calculer les quantiles suivants : Q1, Q2, D5, D8, C50 et C95.

| Nombre d’enfants (xi) | Nombre de ménage (ni) | $n_i \uparrow$ |
|---|---|---|
| 0 | 20 | 20 |
| 1 | 65 | 85 |
| 2 | 70 | 155 |
| 3 | 30 | 185 |
| 4 | 10 | 195 |
| 5 | 5 | 200 |
| Total | 200 | - |

**Solution :**

- Q1 $\left(\alpha = \frac{1}{4},\ N = 200\right)$ : $Q_1 = \frac{x_{n\alpha} + x_{n\alpha+1}}{2} = \frac{x_{50} + x_{51}}{2} = \frac{1+1}{2} = 1 \text{ enfants}$

Cette valeur indique que 25% de ménages ont au plus d’un enfant et 75% ont plus d’un enfant.

- Q2 $\left(\alpha = \frac{1}{2},\ N = 200\right)$ : $Q_2 = \frac{x_{n\alpha} + x_{n\alpha+1}}{2} = \frac{x_{100} + x_{101}}{2} = \frac{2+2}{2} = 2 \text{ enfants}$

Cette valeur indique que 50% de ménages ont au plus de deux enfants et l’autre 50% de ménages ont plus de deux enfants.

- D5 $\left(\alpha = \frac{5}{10},\ N = 200\right)$ : $D_5 = \frac{x_{n\alpha} + x_{n\alpha+1}}{2} = \frac{x_{100} + x_{101}}{2} = \frac{2+2}{2} = 2 \text{ enfants}$

- D8 $\left(\alpha = \frac{8}{10},\ N = 200\right)$ : $D_8 = \frac{x_{n\alpha} + x_{n\alpha+1}}{2} = \frac{x_{160} + x_{161}}{2} = \frac{3+3}{2} = 3 \text{ enfants}$

Cette valeur indique que 80% de ménages ont au plus de 3 enfants et 20% de ménages ont plus de 3 enfants.

- C50 $\left(\alpha = \frac{50}{100},\ N = 200\right)$ : $C_{50} = \frac{x_{n\alpha} + x_{n\alpha+1}}{2} = \frac{x_{100} + x_{101}}{2} = \frac{2+2}{2} = 2 \text{ enfants}$

- C95 $\left(\alpha = \frac{95}{100},\ N = 200\right)$ : $C_{50} = \frac{x_{n\alpha} + x_{n\alpha+1}}{2} = \frac{x_{190} + x_{191}}{2} = \frac{4+4}{2} = 4 \text{ enfants}$

Cette valeur indique que 95% de ménages ont au plus de 4 enfants et 5% de ménages ont plus de 4 enfants.

**Exemple 2 (Cas d'une variable continue) :** Soit le tableau suivant, calculer Q1, D4 et C85

| xi | $n_i$ | $f_i$ | $f_i \uparrow$ |
|---|---|---|---|
| [3 - 4[ | 12 | 0,024 | 0.024 |
| [4 - 6[ | 30 | 0,061 | 0.085 |
| [6 - 8[ | 120 | 0,246 | 0.331 |
| [8 - 9[ | 210 | 0,431 | 0,762 |
| [9- 10[ | 90 | 0,184 | 0,946 |
| [10- 12[ | 25 | 0,051 | $\approx 1$ |
| Total | 487 | $\approx 1$ | - |

**Solution :** Pour nos calculs, nous avons retenu les fréquences relatives cumulées croissantes.

- $Q1(\alpha = 0{,}25)$ , cette valeur est située dans la colonne des fréquences relatives cumulées croissantes ( $f_i \uparrow$ ) entre les valeurs 0,085 et 0,331, cette dernière correspond à la classe $[6;8[$. Donc $Q1 \in [6;8[ \Rightarrow Q_1 = X_{\min} + \dfrac{\alpha - F_{i\ précédent}^{\uparrow}}{fi_{Q1}} \times ai = 6 + \dfrac{0{,}25 - 0{,}085}{0{,}246} \times 2 = 7{,}34$

- $D4(\alpha = 0{,}4{:})$ : cette valeur est située dans la colonne des fréquences relatives cumulées croissantes ( $f_i \uparrow$ ) entre les valeurs 0,331 et 0, 762, cette dernière correspond à la classe $[8;9[$. Donc $D4 \in [8\,;\ 9[ \Rightarrow D4 = X_{\min} + \dfrac{\alpha - F_{i\ précédent}^{\uparrow}}{fi_{D4}} \times ai = 8 + \dfrac{0{,}4 - 0{,}331}{0{,}431} \times 1 = 8{,}16$

- $C85(\alpha = 0{,}85{,})$ : cette valeur est située dans la colonne des fréquences relatives cumulées croissantes ( $f_i \uparrow$ ) entre les valeurs 0, 762 et 0,946, cette dernière correspond à la classe $[9;10[$.Donc $C85 \in [9;10[ \Rightarrow C85 = X_{\min} + \dfrac{\alpha - F_{i\ précédent}^{\uparrow}}{fi_{C85}} \times ai = 9 + \dfrac{0{,}85 - 0{,}762}{0{,}184} \times 1 = 9{,}48$

**Exercices**

**Exercice 1 :** Le tableau suivant indique la distribution de 100 logements selon le nombre de pièces dans quartier.

| Nombre de pièce (xi) | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Nombre de logements (ni) | 5 | 10 | 20 | 30 | 25 | 10 |

1- Quelle est la population statistique étudiée ?
2- Quel est le caractère étudié de cette distribution ? préciser sa nature.
3- Donner le mode de cette série et déterminer le graphiquement. Quelle est sa signification ?
4- Tracer le diagramme cumulatif des fréquences relatives.
5- Calculer la médiane et déterminer sa valeur graphiquement.
6- Calculer les quantiles ; Q1, Q2, Q3, D5, C50 et C75.
7- Calculer la moyenne arithmétique.

**Exercice 2 :** On considère une population de 1000 individus répartis en fonction de leur âge :

| Age | [0 - 10[ | [10 - 15[ | [15 - 20[ | [20- 30[ | [30 - 40[ | [40 - 60[ | [60 - 80[ |
|---|---|---|---|---|---|---|---|
| Nombre | 120 | 100 | 140 | 200 | 180 | 160 | 100 |

1- Déterminer la population statistique, le caractère et sa nature.
2- Déterminer le mode graphiquement et calculer sa valeur.
3- Tracer les courbes des effectifs cumulées (croissants et décroissants)
4- Déterminer la médiane graphiquement et calculer sa valeur. Quelle est sa signification ?
5- Calculer les quantiles ; Q1, Q2, Q3, D5, C50 et C75.
6- Calculer le moyen âge de ces individus.
7- Quel est le nombre d'individus ayant l'âge moins de 20 ans ?
8- Quel est le nombre d'individus ayant l'âge au moins 15 ans ?

**Exercice 3** : Les taux de croissance annuel du PIB d'un pays pendant au cours des années 2000-2003 se présentent dans le tableau suivant :

| Année | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|
| Taux | 7,2 | 6,3 | 7 | 4,3 |

- Quel est le taux de croissance annuel moyen du PIB au cours de ces quatre années ?

## Chapitre 4 : Les paramètres de dispersion

Les indices de tendance centrale vus dans le chapitre précédent, définissent le comportement général des données. Mais ils ne donnent aucune indication sur la façon dont sont groupées ces données. Ils ne renseignent donc pas sur la variabilité des données dans l'échantillon. Par conséquent, les paramètres de dispersion renseignent sur la dispersion ou l'éparpillement des données autour notamment des paramètres de tendance centrale.

Ainsi, on étudiera dans le présent chapitre les principaux paramètres de dispersion, en mettant plus particulièrement l'accent sur la variance et l'écart-type :

- L'étendue et le rapport de variation.
- L'intervalle interquartile.
- L'écart absolu moyen.
- La variance et l'écart-type.
- Le coefficient de variation.

### 1. L'étendue

L'étendue notée **E** est la différence entre la plus grande et la plus petite des valeurs observées de la variable statistique. Elle est utilisée pour obtenir très rapidement une première idée de la dispersion de la série. On note : $E = X_{max} - X_{min}$

Le principal avantage de ce paramètre est sa simplicité de calcul, mais il est, par contre, extrêmement peu robuste.

**Exemple 1 (Cas d'une variable discrète) :** Soit la distribution suivante qui indique les notes obtenues par 10 étudiants lors de deux examens.

| xi | ni |
|---|---|
| 3 | 1 |
| 4 | 1 |
| 5 | 2 |
| 7 | 2 |
| 8 | 1 |
| 9 | 2 |
| 11 | 1 |

**Solution :**

L'étendue est égale à : $E = X_{max} - X_{min}$ = (11-3) = 8.

**Exemple 2 (Cas d'une variable continue) :** soit la distribution suivante dans le tableau ci-après donne la répartition de 24 étudiants selon leur taille.

| X | $n_i$ |
|---|---|
| [1,5 ; 1,6 [ | 6 |
| [1,6 ; 1,7 [ | 7 |
| [1,7 ; 1,8 [ | 8 |
| [1,8 ; 1,9 [ | 2 |
| [1,9 ; 2 [ | 1 |
| Total | 24 |

**Solution :**

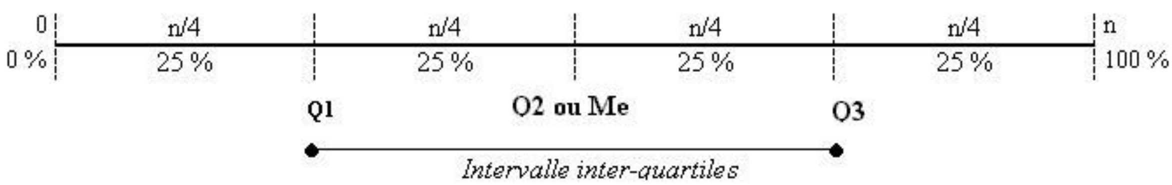
Dans ce cas, L'étendue est la différence entre la borne supérieure de la dernière classe et la borne inférieure de la première classe.

L'étendue est égale à : $E = X_{max} - X_{min}$ = (2 - 1,5) = 0.5.

**2. Les intervalles et écarts interquartiles**

Il existe trois intervalles et écarts interquartiles :

- **L'intervalle interquartile** $[Q_1; Q_3]$ est l'intervalle qui contient 50% des observations ; l'amplitude de l'intervalle est appelé **écart interquartile**, noté $EIQ$ est égale à : $EIQ = Q_3 - Q_1$




- **L'intervalle interdécile** $[D_1; D_9]$ est l'intervalle qui contient 80% des observations ; l'amplitude de l'intervalle est appelé **écart interdécile**, noté $EID$ est égale à : $EID = D_9 - D_1$

- **L'intervalle intercentile** $[C_1; C_{99}]$ est l'intervalle qui contient 98% des observations ; l'amplitude de l'intervalle est appelé **écart intercentile**, noté $EIC$ est égale à : $EIC = C_{99} - C_1$

Plus ces intervalles sont importants, plus la dispersion est forte.

### 3. Ecarts absolus moyens

L'écart absolu moyen est la moyenne arithmétique des valeurs absolues des écarts à un indicateur de tendance centrale, en général la moyenne ou la médiane.

- **Ecart absolu moyen par rapport à la moyenne arithmétique** : $e_{\bar{X}} = \frac{\sum_{i=1}^{k} n_i |X_i - \bar{X}|}{N}$
- **Ecart absolu moyen par rapport à la médiane** : $e_{Mé} = \frac{\sum_{i=1}^{k} n_i |X_i - Mé|}{N}$

**Remarque** : Lorsque la variable est continue, on remplace les valeurs de xi par les centres de classe et on applique le même principe pour calculer l'écart absolu moyen.

**Exemple 1 (cas variable discrète) :** la distribution suivante représente les notes obtenues par les étudiants à l'examen statistique.

| Notes xi | 7 | 8 | 9 | 10 | 11 | 12 | 14 |
|---|---|---|---|---|---|---|---|
| Etudiants ni | 10 | 15 | 20 | 25 | 15 | 10 | 5 |

- Calculer l'écart absolu moyen par rapport à la moyenne arithmétique.

**Solution** : Pour rendre les calculs simples et facile, on dresse le tableau suivant :

| $X_i$ | $n_i$ | $n_i \cdot X_i$ | $\lvert X_i - \bar{X} \rvert$ | $ni \cdot \lvert X_i - \bar{X} \rvert$ |
|---|---|---|---|---|
| 7 | 10 | 70 | 2,75 | 27,5 |
| 8 | 15 | 120 | 1,75 | 26,25 |
| 9 | 20 | 180 | 0,75 | 15 |
| 10 | 25 | 250 | 0,25 | 6,25 |
| 11 | 15 | 165 | 1,25 | 18,75 |
| 12 | 10 | 120 | 2,25 | 22,5 |
| 14 | 5 | 70 | 4,25 | 21,25 |
| Total | 100 | 975 | - | 137,5 |

On calcule d'abord la moyenne arithmétique :

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i X_i}{N} = \frac{975}{100} = 9{,}75$$

L'écart absolu moyen par rapport à la moyenne arithmétique =

$$e_{\overline{X}} = \frac{\sum_{i=1}^{k} n_i |X_i - \overline{X}|}{N} = \frac{137{,}5}{100} = 1{,}375$$

**Exemple 2 (cas variable continue):** on considère la distribution suivante :

| Classes | [15-25[ | [25-35[ | [35-45[ | [45-55[ | [55-65[ | [65-75[ | [75-85[ |
|---|---|---|---|---|---|---|---|
| Effectifs | 9 | 15 | 22 | 29 | 17 | 6 | 2 |

- Calculer l'écart absolu moyen par rapport à la moyenne.
- Calculer l'écart absolu moyen par rapport à la médiane.

**Solution** : on dresse le tableau suivant :

| Classes | $n_i$ | $c_i$ | $n_i \cdot c_i$ | $\lvert c_i - \overline{X}\rvert$ | $n_i \cdot \lvert c_i - \overline{X}\rvert$ | $n_i \uparrow$ | $\lvert c_i - Mé\rvert$ | $n_i \cdot \lvert c_i - Mé\rvert$ |
|---|---|---|---|---|---|---|---|---|
| [15-25[ | 9 | 20 | 180 | 25,6 | 230,4 | 9 | 26,38 | 237,42 |
| [25-35[ | 15 | 30 | 450 | 16,6 | 234,0 | 24 | 16,38 | 245,70 |
| [35-45[ | 22 | 40 | 880 | 5,6 | 123,2 | 46 | 6,38 | 104,36 |
| [45-55[ | 29 | 50 | 1450 | 4,4 | 127,6 | 75 | 3,62 | 104,98 |
| [55-65[ | 17 | 60 | 1020 | 14,4 | 244,8 | 92 | 13,62 | 231,54 |
| [65-75[ | 6 | 70 | 420 | 24,4 | 146,4 | 98 | 23,62 | 141,72 |
| [75-85[ | 2 | 80 | 160 | 34,4 | 68,8 | 100 | 33,62 | 67,24 |
| **Total** | **100** | **-** | **4560** | **-** | **1175,2** | **-** | **-** | **1168,96** |

- $\overline{X} = \frac{\sum_{i=1}^{n} n_i c_i}{N} = \frac{4560}{100} = 45{,}6 \Rightarrow e_{\overline{X}} = \frac{\sum_{i=1}^{k} n_i |c_i - \overline{X}|}{N} = \frac{1175{,}2}{100} = 11{,}75$

- $Mé \in [45;55[ \;\Rightarrow Mé = X_{\min} + \frac{n/2 - N_i^{\uparrow}{}_{précédent}}{ni_{Mé}} \times ai = 45 + \frac{50-46}{29} \times 10 = 46{,}38$

$$\Rightarrow \; e_{Mé} = \frac{\sum_{i=1}^{k} n_i |X_i - Mé|}{N} = \frac{1168{,}96}{100} = 11{,}69$$

## 4. La variance

La variance de la variable X est la moyenne des carrés des écarts des valeurs observées par rapport à la moyenne *arithmétique*. Elle signifie la variation des valeurs observées autours de leur moyenne. On la note généralement V(X). Sa formule correspondant à sa définition :

$$V(X) = \frac{\sum_{i=1}^{n} n_i (X_i - \bar{X})^2}{N} \quad ou \quad V(X) = \sum f_i (X_i - \bar{X})^2 \text{..............} \textit{ formule de définition}$$

Elle peut aussi se calculer à partir d'une formule développée déduite de la formule précédente :

$$V(X) = \frac{\sum_{i=1}^{n} n_i x_i^2}{N} - \bar{X}^2 \quad ou \; V(X) = \sum_{i=1}^{n} fi \; x_i^2 - \bar{X}^2 \text{ ..............} \textit{ formule développée}$$

**Propriétés de la variance**

1. $V(X) \geq 0$
2. Soit Y une variable telle que : $y_i = a\, x_i + b$ , sachant a et b sont des constantes réelles, alors : $V(Y) = a^2 V(X)$

**Remarque :** Pour calculer la variance dans le cas continu, on remplace les valeurs (xi) par les centres de classe ci, comme suite :

$$V(X) = \frac{\sum_{i=1}^{n} n_i (c_i - \bar{X})^2}{N} \quad \Rightarrow \quad V(X) = \sum f_i (c_i - \bar{X})^2$$

O u $V(X) = \frac{\sum_{i=1}^{n} n_i c_i^2}{N} - \bar{X}^2 \quad \Rightarrow \quad V(X) = \sum_{i=1}^{n} fi \; ci^2 - \bar{X}^2$

## 5. L'écart type

L'écart type est un indicateur de dispersion. Il est défini comme la racine carrée de la variance, il s'exprime dans la même unité que la variable xi. On le note généralement : $\sigma(X)$ . Donc : $\sigma(X) = \sqrt{v(x)}$ .

Cet indicateur mesure la distance moyenne entre $\overline{X}$ et les valeurs de X. Il sert à mesurer la dispersion d'une série statistique autour de sa moyenne.

- Plus il est petit, plus les caractères sont concentrés autour de la moyenne (on dit que la série est homogène).
- Plus il est grand, plus les caractères sont dispersés autour de la moyenne (on dit que la série est hétérogène).

**6. Le coefficient de variation (coefficient de dispersion relative)**

Malgré la pertinence de l'écart-type dans la mesure de la dispersion d'une distribution, il possède un inconvénient majeur : il est exprimé dans la même unité de la variable à laquelle il se rapporte, il est alors impossible de comparer les dispersions de deux distributions ayant un lien entre elles et dont les valeurs s'expriment dans des unités différentes.
Pour comparer la dispersion de deux séries qui ne sont pas exprimées dans les mêmes unités, on utilise le coefficient de variation (CV). Cette statistique est une mesure sans unité, qui s'exprime la plupart du temps en pourcentage. Il se calcule en divisant l'écart-type par la moyenne arithmétique et s'écrit donc : $CV = \frac{\sigma}{\overline{X}}$
Plus la valeur de ce coefficient est grande, plus la dispersion est grande.

**Exemple1 (cas d'une variable discrète)**

Le tableau suivant indique la distribution de notes obtenues par les étudiants à l'examen de statistique.

| Notes | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Etudiants | 2 | 9 | 14 | 20 | 18 | 15 | 9 | 6 | 4 | 2 | 1 |

- Calculer la variance, l'écart type et le coefficient de variation.

**Solution**

On calcule la variance par la formule de définition, donc on dresse le tableau de cette manière :

| $X_i$ | $n_i$ | $n_i X_i$ | $X_i - \overline{X}$ | $(X_i - \overline{X})^2$ | $n_i (X_i - \overline{X})^2$ |
|---|---|---|---|---|---|
| 5 | 2 | 10 | -4 | 16 | 32 |
| 6 | 9 | 54 | -3 | 9 | 81 |
| 7 | 14 | 98 | -2 | 4 | 56 |
| 8 | 20 | 160 | -1 | 1 | 20 |
| 9 | 18 | 162 | 0 | 0 | 0 |
| 10 | 15 | 150 | 1 | 1 | 15 |
| 11 | 9 | 99 | 2 | 4 | 36 |
| 12 | 6 | 72 | 3 | 9 | 54 |
| 13 | 4 | 52 | 4 | 16 | 64 |
| 14 | 2 | 28 | 5 | 25 | 50 |
| 15 | 1 | 15 | 6 | 36 | 36 |
| **Total** | 100 | 900 | - | - | 444 |

- **La moyenne arithmétique :**

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \frac{900}{100} = 9 .$$

- **La variance :**

$$V(X) = \frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^2}{N} = \frac{444}{100} = 4,44 \ .$$

- **L'écart-type :**

$$\sigma(X) = \sqrt{V(X)} = \sqrt{4,44} = 2,107 .$$

- **Le coefficient de variation :**

$$CV = \frac{\sigma}{\overline{X}} = \frac{2,107}{9} = 0,23$$

**Exemple 2 (Cas d'une variable continue)**

Le tableau ci-dessous indique la distribution des salaires mensuels des travailleurs d'une entreprise.

| Classes | [50-60[ | [60-70[ | [70-80[ | [80-90[ | [90-100[ |
|---|---|---|---|---|---|
| Effectifs | 25 | 40 | 20 | 10 | 5 |

- Calculer la variance, l'écart-type et le coefficient de variation.

**Solution**

Dans ce cas, on calcule la variance par la formule développée, on doit construire alors le tableau de la manière suivante :

| Classes | $n_i$ | $c_i$ | $n_i \times c_i$ | $n_i \cdot c_i^2$ |
|---|---|---|---|---|
| [50-60[ | 25 | 55 | 1375 | 75625 |
| [60-70[ | 40 | 65 | 2600 | 169000 |
| [70-80[ | 20 | 75 | 1500 | 112500 |
| [80-90[ | 10 | 85 | 850 | 72250 |
| [90-100[ | 5 | 95 | 475 | 45125 |
| Total | 100 | - | 6800 | 474500 |

On calcule d'abord **la moyenne arithmétique** :

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i c_i}{N} = \frac{6800}{100} = 68$$

**La variance** est égale à :

$$V(X) = \frac{\sum_{i=1}^{n} n_i c_i^2}{N} - \overline{X}^2 = \frac{474500}{100} - (68)^2 = 121$$

**L'écart – type :**

$$\sigma(X) = \sqrt{V(X)} = \sqrt{121} = 11.$$

**Le coefficient de variation** :

$$CV = \frac{\sigma}{\overline{X}} = \frac{11}{68} = 0{,}16$$

## Exercices

### Exercice 1

Le tableau suivant indique la distribution d'une population de 16 femmes d'un village selon leur nombre d'enfants.

| Nombre d'enfants par femme (xi) | Nombre de femmes (ni) |
|---|---|
| 0 | 1 |
| 1 | 3 |
| 2 | 8 |
| 3 | 4 |

1- Déterminer la population statistique.
2- Quel est le caractère de cette distribution ? préciser sa nature.
3- Calculer le nombre moyen d'enfants par femme.
4- Calculer la variance ainsi que l'écart type.
5- Calculer le coefficient de variation.

### Exercice 2

Une enquête auprès de 500 visiteurs d'un musée âgés d'au moins 15 ans permet d'obtenir la distribution statistique ci-après :

| Ages (en année) | [15 - 25[ | [25 - 35[ | [35 - 50[ | [50- 65[ | [65 - 75[ |
|---|---|---|---|---|---|
| Nombre de visiteurs | 96 | 118 | 138 | 101 | 47 |

1- Déterminer la population statistique, le caractère et sa nature.
2- Présenter graphiquement cette distribution.
3- Calculer la moyenne arithmétique.
4- Calculer les caractéristiques de dispersion suivantes : l'étendue, l'écart absolu moyen, l'écart type et coefficient de variation.

## Chapitre 5 : Les moments et les paramètres de forme

Les paramètres de forme sont des nombres sans unités, ils permettent de préciser l'allure d'une distribution statistique sans avoir besoin de la tracer. On repère généralement deux mesures de la forme celle de l'asymétrie et celle de l'aplatissement. Ainsi, ces paramètres sont établis à l'aide des moments de la distribution.

### 1. Les moments

Les moments sont utilisés principalement pour déterminer les différentes formules relatives aux caractéristiques de forme. On définit le moment d'ordre k par rapport à une valeur quelconque $X_0$ comme suit : $M_k = \dfrac{\sum_{i=1}^{n} n_i (X_i - X_0)^k}{N} = \sum_{i=1}^{n} fi(X_i - X_0)^k$

Avec k : ordre du moment, prenant les valeurs : $\{0,1,2,...,n\}$ et $X_0$ : origine du moment.

#### 1.1. Moments non centrés (notés : $m_k$)

Lorsque la valeur $X_0$ est nulle, on définit les moments non centrés, sous la formule suivante :

$$m_k = \frac{\sum_{i=1}^{n} n_i x_i^k}{N} = \sum_{i=1}^{k} f_i x_i^k$$

On peut déterminer quelques moments non centrés en fonction de la valeur de k:

- Si k=0 $\Rightarrow m_0 = 1$.
- Si k=1 $\Rightarrow m_1 = \dfrac{\sum_{i=1}^{n} n_i X_i}{N} = \overline{X}$ : Moyenne arithmétique, moment du 1er ordre.
- Si k=2 $\Rightarrow m_2 = \dfrac{\sum_{i=1}^{n} n_i X_i^2}{N}$ : Carré de la moyenne quadratique, moment du 2ème ordre.

#### 1.2. Moments centrés (notés : $\mu_k$)

Lorsque la valeur $X_0$ est égale à la moyenne arithmétique ( $X_0 = \overline{X}$ ), on définit les moments centrés. Son expression devient alors :

$$\mu_k = \frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^k}{N} = \sum_{i=1}^{n} fi(X_i - \overline{X})^k$$

On peut déterminer quelques moments centrés en fonction de la valeur de k:

- Si k=0 $\Rightarrow \mu_0 = 1$.
- Si k=1 $\Rightarrow \mu_1 = 0$.
- Si k=2 $\Rightarrow \mu_2 = \frac{\sum_{i=1}^{n} n_i (X_i - \overline{X})^2}{N} = \sum_{i=1}^{n} fi(X_i - \overline{X})^2 = V(X)$ : la variance.

**2. Les caractéristiques de forme**

Il s'agit de savoir si telle distribution est symétrique et moins aplatie ou non.

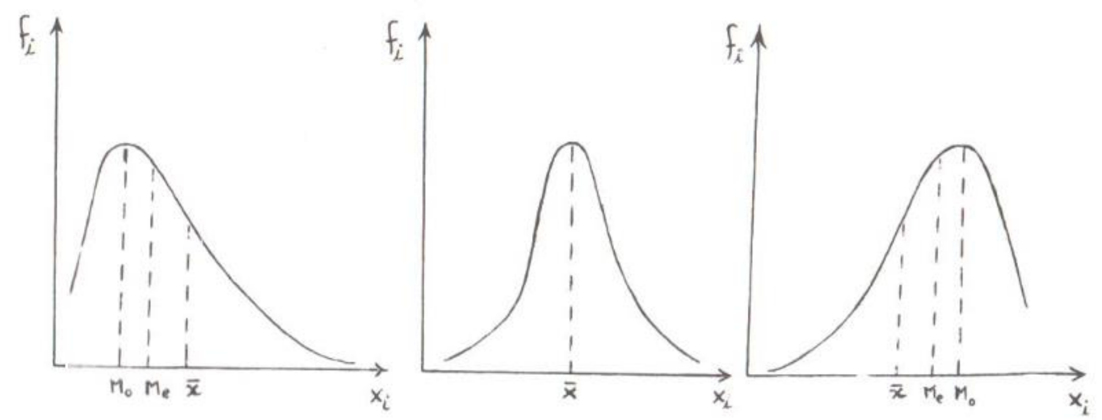
**2.1. La mesure de l'asymétrie**

L'asymétrie a pour objet de nous renseigner sur la façon régulière ou non dont les observations selon leurs fréquences se répartissent de part et d'autre part d'une valeur centrale.

Dans une distribution symétrique, les trois caractéristiques de tendance centrale sont confondues : Moyenne = Médiane = Mode.

Ainsi Q1 et Q3, D1 et D9 et C1 et C99 sont équidistants par rapport à la médiane. C'est-à-dire : $\begin{cases} Q3 - Mé = Mé - Q1 \\ D9 - Mé = Mé - D1 \\ C99 - Mé = Mé - C1 \end{cases}$

En général, l'asymétrie des courbes de fréquences des distributions statistiques est étudiée par rapport aux courbes de fréquence suivantes :




Assymétrie : obliquité à gauche $M_o < M_e < \bar{x}$ | Symétrie $M_o = M_e = \bar{x}$ | Assymétrie : Obliquité à droite $M_o > M_e > \bar{x}$

Les principaux coefficients utilisés pour mesurer l'asymétrie d'une distribution statistique sont les suivants :

**1. Coefficient de Yule**

Le coefficient de Yule sert à mesurer l'asymétrie de la distribution en tenant compte des positions relatives des quartiles par rapport à la médiane, il est définit par la formule suivante :

$$C_y = \frac{Q_3 - M_é - (M_é - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2M_é + Q_1}{Q_3 - Q_1}$$

Ce coefficient est toujours compris entre -1 et 1, lorsque :

$C_y = 0 \rightarrow$ La courbe est symétrique.

$C_y \succ 0 \rightarrow$ La courbe est étalée à droite.

$C_y \prec 0 \rightarrow$ La courbe est étalée à gauche.

**2. Coefficients de Pearson :** Pearson propose deux coefficients pour mesurer l'asymétrie :

**-Le premier coefficient :** consiste à mesurer l'asymétrie d'une distribution par comparaison entre les valeurs de la moyenne et du mode. Il se note : $\beta_1 = \dfrac{\overline{X} - M_0}{\sigma(X)}$

$\beta_1 = 0 \rightarrow$ La distribution est symétrique.

$\beta_1 \succ 0 \rightarrow$ La distribution est étalée à droite.

$\beta_1 \prec 0 \rightarrow$ La distribution est étalée à gauche.

**-Le deuxième coefficient :** Il est calculé à partir des moments centrés d'ordre 3 et d'ordre 2 de la série statistique. Il est définit par **:** $\beta_2 = \frac{\mu_3^2(X)}{\mu_2^3(X)}$

$\beta_2 = 0 \rightarrow$ La distribution est symétrique

$\beta_2 \succ 0 \rightarrow$ La distribution est étalée à droite pour $\mu_3(X) \succ 0$ .

$\beta_2 \prec 0 \rightarrow$ La distribution est étalée à gauche pour $\mu_3(X) \prec 0$ .

**3. Coefficient de Fisher :** il est définit par : $\gamma_1 = \frac{\mu_3(X)}{\sigma^3(X)}$

Nous constatons que le coefficient de Fisher n'est autre que la racine carrée du coefficient $\beta_2$ de Pearson.

Les conclusions sont les mêmes que précédemment : si $\gamma_1 = 0$ , La distribution est symétrique ; si $\gamma_1$ est positif, la distribution est étalée à droite , si $\gamma_1$ est négatif, la distribution est étalée à gauche.

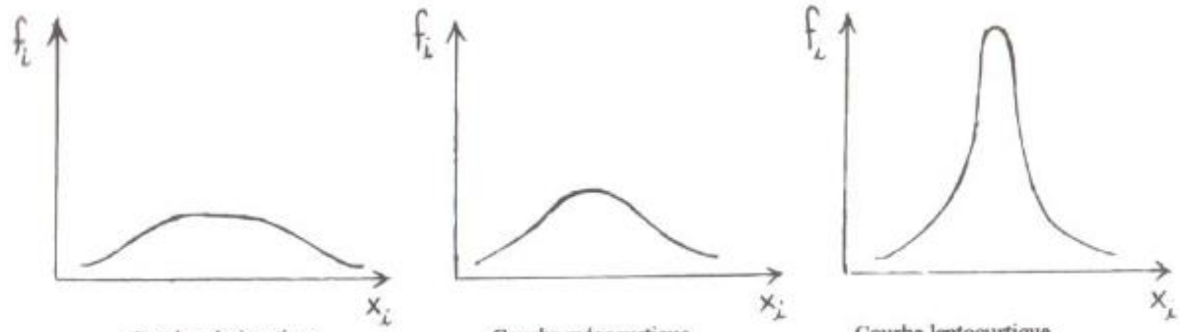
### 2.2. La mesure de l'aplatissement

L'aplatissement a pour objet de faire apparaitre si une faible variation de la variable entraine ou non une forte variation des fréquences relatives.

Une courbe de fréquences est considérée plus ou moins aplatie par rapport à la courbe des fréquences de la loi normale.

Ainsi, une distribution est dite aplatie si une forte variation de la variable entraine une faible variation de la fréquence relative (et inversement).

En général, les formes d'aplatissement de références sont les suivantes :



Courbe platicurtique — Courbe mésocurtique (normale ou enclôche) — Courbe leptocurtique

Les principaux coefficients utilisés pour mesurer le degré d'aplatissement sont les suivants :

**1. Coefficient de Pearson**

Il est de la forme : $\beta_3 = \dfrac{\mu_4(X)}{\mu_2^2(X)} = \dfrac{\mu_4(X)}{\sigma^4(X)}$

En général, nous savons toujours l'intégralité : $\mu_4 \succ \mu_2^2$

Le coefficient $\beta_3$ est toujours supérieur à 1, car $\mu_4 \succ \mu_2^2$.

Si $\beta_3 = 3$ , la courbe de la distribution est normale.

Si $\beta_3 \succ 3$ , la courbe de la distribution est leptocurtique (moins aplatie que la courbe normale).

Si $\beta_3 \prec 3$ , la courbe de la distribution est platicurtique (plus aplatie que la courbe normale).

**2. Coefficient de Fisher**

Il est donné par la formule suivante :

$$\gamma_2 = \beta_3 - 3 = \frac{\mu_4(X)}{\sigma^4(X)} - 3$$

Sens de l'aplatissement :

Si $\gamma_2 = 0$ , la courbe de la distribution est normale.

Si $\gamma_2 \succ 0$ , la courbe de la distribution est leptocurtique.

Si $\gamma_2 \prec 0$ , la courbe de la distribution est platicurtique.

**Exemple d'application** : soit la série suivante. Calculez les coefficients d'asymétrie et d'aplatissement de Pearson.

| xi | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| fi | 0,216 | 0,432 | 0,288 | 0,064 |

**Solution :**

Le tableau s'écrit de la manière suivante :

| xi | fi | fi.xi | $X_i - \overline{X}$ | $f_i(X_i - \overline{X})^2$ | $f_i(X_i - \overline{X})^4$ |
|---|---|---|---|---|---|
| 0 | 0,216 | 0 | -1,2 | 0,311 | 0,448 |
| 1 | 0,432 | 0,432 | -0,2 | 0,017 | 0,00069 |
| 2 | 0,288 | 0,576 | 0,8 | 0,184 | 0,11796 |
| 3 | 0,064 | 0,192 | 1,8 | 0,207 | 0,6718 |
| Total | 1 | 1,2 | - | 0,72 | 1,238 |

1- Le coefficient d'asymétrie de Pearson.

On utilise le premier coefficient de Pearson : $\beta_1 = \dfrac{\overline{X} - M_0}{\sigma(X)}$

On calcule d'abord, $\overline{X}$, $M_O$ et $\sigma(X)$ :

La moyenne arithmétique : $\overline{X} = \dfrac{\sum_{i=1}^{n} n_i \cdot x_i}{N} = \sum_{i=1}^{n} fi.xi = 1{,}2$.

Le mode : $M_O = 1$.

L'écart-type : $\sigma(X) = \sqrt{V(X)} = \sqrt{f_i(X_i - \overline{X})^2} = \sqrt{0{,}72}$.

$\Rightarrow \beta_1 = \dfrac{1{,}2 - 1}{\sqrt{0{,}72}} = 0;23 \succ 0$, donc la distribution est étalée à droit.

2- Le coefficient d'aplatissement de Pearson : $\beta_3 = \dfrac{\mu_4(X)}{\mu_2^2(X)} = \dfrac{\mu_4(X)}{\sigma^4(X)}$

$\Rightarrow \beta_3 = \dfrac{f_i(X_i - \overline{X})^4}{\left(f_i(X_i - \overline{X})^2\right)^2} = \dfrac{1{,}238}{\left(\sqrt{0{,}72}\right)^4} = 2{,}38 \prec 3$ , donc la courbe des fréquences est platicurtique.

## Exercices

**Exercice 1 :** Soit la distribution suivante :

| Classes | [15 - 25[ | [25 - 35[ | [35 - 45[ | [45- 55[ | [55 - 65[ | [65 - 75[ | [75 - 85[ |
|---|---|---|---|---|---|---|---|
| Effectifs | 5 | 15 | 32 | 40 | 66 | 25 | 17 |

1- Calculer ses moments centrés d'ordre 2,3 et 4.

2- Calculer le coefficient d'asymétrie de Ficher.

3- Calculer le coefficient d'aplatissement de Pearson.

**Exercice 2 :** Soit la distribution de lots fabriqués par une usine selon le nombre de pièces défectueuses :

| Nombre de pièces défectueuses | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|
| Nombre de lots | 5 | 7 | 8 | 11 | 16 | 13 |

1- Calculer la valeur des indicateurs d'asymétrie et commenter les résultats.

2- Calculer la valeur des indicateurs d'aplatissement et commenter les résultats.

## Conclusion générale

Ce présent document constitue un support de cours préliminaire de la statistique descriptive, il donne à l'étudiant les outils nécessaires à la compréhension et à l'analyse de ce module selon le programme officiel de la matière. Il s'agit d'une initiation simple et claire aux étudiants, qui ne nécessite pas de connaissances spécifiques et compliquées dans le domaine ; nous avons utilisé tout au long de ce document des expressions et des instruments très simples.

Nous avons établi ce polycopié selon un enchainement logique de plusieurs chapitres, dont le premier a été consacré à la définition de la statistique et sa composante descriptive, il explique la signification des différents concepts et notions adoptées dans la démarche statistique, suivi d'un détail sur la nature des caractères étudiées en statistique et les différentes techniques d'organisation et de structuration d'une série de données. Ce polycopié a exposé également les différents paramètres descriptifs d'une série statistique simple à savoir les paramètres de position, les paramètres de dispersion et les paramètres de forme d'une distribution de fréquences.

Et nous avons accompagné chaque chapitre par un ensemble d'exemples explicatifs et une série des exercices.La fin de ce document a été réservée à un sujet d'examen avec correction, qui permet à l'étudiant d'évaluer ses connaissances en la matière.

Université Abderrahmane Mira Bejaia Avril 2018

Département Tronc Commun SEGC-LMD

Première année SEGC-LMD

**Examen de Rattrapage de STATI**

**(Durée 1h30mn)**

**Exercice n°1 (08 points).**

Le responsable des ventes d'un fournisseur de matériel informatique a note le niveau de la demande journalière pour un produit pendant 100 jours ouvrables consécutifs.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Nombre d'unités demandées | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Nombre de jours | 5 | 15 | 23 | 22 | 16 | 12 | 7 |

1. Quel est le caractère étudié ? Déterminer son type.
2. Représenter cette série par un diagramme en bâtons.
3. Donner le mode de cette série et déterminer le graphiquement.
4. Calculer la médiane et les quantiles : Q1 ; Q2 et Q3.
5. Calculer le nombre moyen d'unités demandées par jour.
6. Quel est le nombre de jours dans lesquels le nombre d'unités demandées est au plus 3.
7. Quel est le pourcentage de jours dans lesquels le nombre d'unités demandées dépasse 3.

**Exercice n°2 (12 points).**

Un fabricant de tubes métalliques pour voitures effectue périodiquement une étude statistique de longueur de ces tubes en sortie de chaine de fabrication. Les résultats de la dernière étude sont donnés dans le tableau suivant.

| | | | | | |
|---|---|---|---|---|---|
| Longueur des tubes (mm) | [1095, 1100[ | [1100, 1110[ | [1110, 1120[ | [1120, 1125[ | [1125, 1130[ |
| Nombre de tubes | 4 | 12 | 24 | 8 | 2 |

1. Déterminer la population statistique, le caractère et sa nature.
2. Déterminer le mode graphiquement et calculer sa valeur.
3. Calculer la médiane, Q1 et Q3.
4. Quel est le pourcentage de tubes ayant au moins 1110 mm de longueur.
5. Calculer la longueur moyenne des tubes.
6. Calculer la variance ainsi que l'écart type.
7. Déterminer le nombre de tubes dont la longueur appartient à l'intervalle [ $\bar{X}$- σ, $\bar{X}$ +σ [.

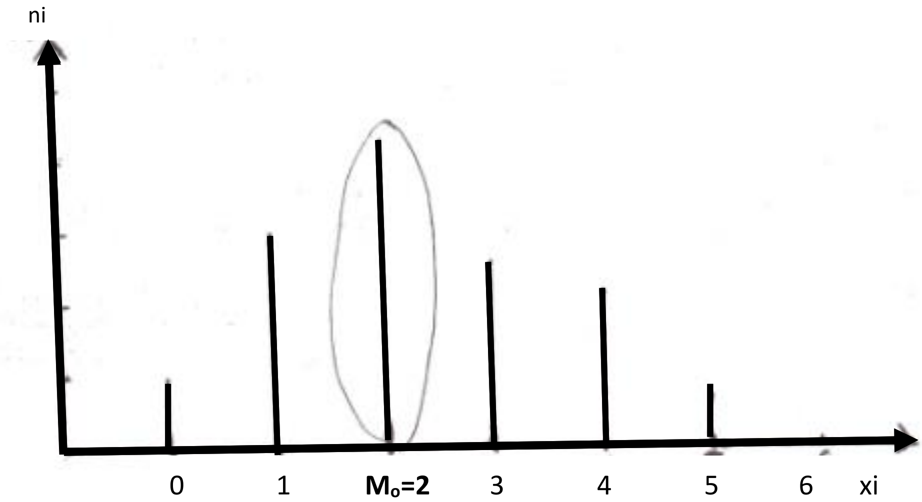
*Corrigé de rattrapage de stat I*

**Solution de l'exercice n° 01 :**

| $x_i$ | $n_i$ | $n_i \uparrow$ | $n_i x_i$ |
|---|---|---|---|
| 0 | 5 | 5 | 0 |
| 1 | 15 | 20 | 15 |
| 2 | 23 | 43 | 46 |
| 3 | 22 | 65 | 66 |
| 4 | 16 | 81 | 64 |
| 5 | 12 | 93 | 60 |
| 6 | 7 | 100 | 42 |
| Total | 100 | | 293 |

**1) Le caractère étudié** est le niveau de la demande journalière pour un produit (nombre d'unités demandées). **Sa nature** est **quantitatif discret**

**2)** La représentation graphique (**diagramme en bâtons)**




**3) Le mode de la série est Mo=** 2 unités par jour

**4) Calcul de la médiane** : $M_e = \frac{X_{50+}X_{51}}{2} = \frac{3+3}{2} = 3\,\text{unités par jour}$

**5) Calcul des quantiles** $Q_1, Q_2, Q_3$

$$Q_1 = \frac{X_{25+}X_{26}}{2} = \frac{2+2}{2} = 2\,\text{unités par jour}$$

$$Q_2 = M_e = 2\,\text{unités par jour}$$

$$Q_3 = \frac{X_{75+}X_{76}}{2} = \frac{4+4}{2} = 4\,\text{unités par jour}$$

**6) Calcul de la moyenne :** $\overline{X} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{293}{100} = 2{,}93 \approx 3\text{unités par jour}$

7) Le nombre de jours dans lesquels le nombre d'unités demandées est au plus 3 unités est **65jours**

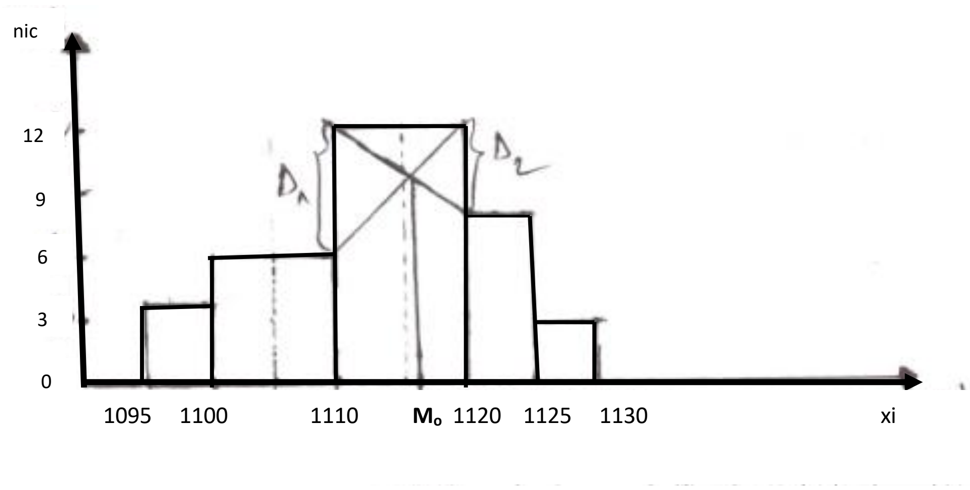
8) Le pourcentage de jours dans lesquels le nombre d'unités demandées dépasse 3 unités est **(16+12+7)/100=0,35= 35%**

**Solution de l'exercice n° 02 :**

| Classes | $n_i$ | $a_i$ | $n_{ic}$ | $n_i \uparrow$ | $c_i$ | $n_i c_i$ | $n_i c_i^2$ |
|---|---|---|---|---|---|---|---|
| [1095-1100 [ | 4 | 5 | 4 | 4 | 1097,5 | 4390 | 4818025 |
| [1100-1110[ | 12 | 10 | 6 | 16 | 1105 | 13260 | 14652300 |
| [1110-1120 [ | 24 | 10 | 12 | 40 | 1115 | 26760 | 29837400 |
| [1120-1125[ | 8 | 5 | 8 | 48 | 1122,5 | 8980 | 10080050 |
| [1125-1130[ | 2 | 5 | 2 | 50 | 1127,5 | 2255 | 2542512,5 |
| Total | 50 | - | - | - | - | 55645 | 61930288 |

**1) La population statistique** est les tubes en sortie de chaine de fabrication (50 tubes métalliques). **Le caractère** étudié est **la longueur des tubes,** sa nature est **quantitatif continu.**

**2) Le mode graphiquement (histogramme)**




La classe modale correspond à la classe de plus grand effectif, c'à d à la classe [1110-1120[.

On a la relation : $M_0 = X_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i$

$$M_0 = 1110 + \frac{(12-6)}{(12-6)+(12-8)} \times 10 = 1116mm$$

**3) Calcul de la médiane et** $Q_1$ et $Q_3$

La classe médiane correspond à la classe [1110-1120 [ $\Rightarrow M_e = X_{\min} + \frac{\frac{N}{2} - ni^{\uparrow}_{Mé-1}}{n_{iMé}} \times a_i$

Donc on a : $M_e = 1110 + \frac{25-16}{24} \times 10 = 1113{,}75mm$

**Calcul des quartiles** $Q_1$ et $Q_3$

Le premier quartile est dans la classe [1100-1110[. On a la relation :

$$Q_1 = X_{\min} + \frac{\frac{N}{4} - ni^{\uparrow}_{Q_1-1}}{n_{iQ_1}} \times a_i \quad \Rightarrow \quad Q_1 = 1100 + \frac{12.5-4}{12} \times 10 = 1107{,}08mm$$

Le troisième quartile est dans la classe [1110-1120 [. On a la relation :

$$Q_3 = 1110 + \frac{37.5 - 16}{24} \times 10 = 1118{,}95$$

**4)** Le pourcentage de tubes ayant au moins une longueur égale à 1110 mm est égale à 34/50=0,68=68%.

**5) La longueur moyenne est :** $\overline{X} = \frac{\sum_i n_i c_i}{N} = \frac{55645}{50} = 1112{,}9$

**6) Calcul de la variance et de l'écart type**

$$V(X) = \frac{\sum n_i c_i^2}{N} - \overline{X}^2 = \frac{61930288}{50} - (1112{,}9)^2 = 59{,}34$$

**L'écart type :** $\sigma_{(X)} = \sqrt{V(X)} = \sqrt{59{,}34} = 7{,}70mm$

**7) Le nombre de tubes dont la longueur appartient à l'intervalle** $[\overline{X} - \sigma, \quad \overline{X} + \sigma[$

$$[\overline{X} - \sigma, \quad \overline{X} + \sigma[ = [1112{,}9 - 7{,}70 \ , \ 1112{,}9 + 7{,}70[ = [1105{,}2 \ , \ 1120{,}9[ \approx [1105 \ , \ 1121[$$

$$n_{[1105 \ , \ 1121[} = n_{[1105 \ , \ 1110[} + n_{[1110 \ , \ 1120[} + n_{[1120 \ , \ 1121[}$$

On doit trouver la valeur $n_{[1105 \ , \ 1110[}$

$$n_{[1100 \ , \ 1110[} \to a = 10 \to n = 12$$

$$n_{[1105 \ , \ 1110[} \to a = 5 \to n_x \Rightarrow n_x = \frac{5 \times 12}{10} = 6$$

**et**

$$n_{[1120 \ , \ 1125[} \to a = 5 \to n = 8$$

$$n_{[1120 \ , \ 1121[} \to a = 1 \to n_x \Rightarrow n_x = \frac{1 \times 8}{5} \approx 2$$

**Donc :** $n_{[1105 \ , \ 1121[} = 6 + 24 + 2 = 32$

## Références bibliographiques


- Ayache.A & Hamonier.J (2014). Cours : **Statistique Descriptive et Calcul de Probabilités**. Université de Lille. France. Disponible sur le site : http://math.univ-lille1.fr › cours_SDCP_Janv19.
- Baccini A. (2010) : **Statistique descriptive élémentaire**. Publications de l'Institut de Mathématiques de Toulouse, UMR CNRS.
- Chaumont L(**2010)** : **Statistique descriptive et prévision**. *Cours*.
- Chauvat.G & REAU.J-P (2008) : **Statistiques descriptives**. Edition Armand colin, 5ème édition. Paris.
- Dakhmouche.M (2010) : cours : **Introduction à la Statistique Descriptive**. Ecole Préparatoire en Sciences Economiques Commerciales et des Sciences de Gestion de Constantine. disponible sur le site : https://www.cours-gratuit.com › manuel.
- Dubos.J(1984) : **Statistique descriptive en science économique**. Edition Dunod, 2eme édition. Paris.
- Duthil, G& et Vanhaecke. D (1998) : **Initiation à la statistique descriptive**. Edition Ellipses.
- Olivier, E. (2008) : **L’essentiel de statistique descriptive**. Edition Gualino. ISBN : 978-2- 297-01103-7.
- Grais.B (1998) : **Exercices corrigés, statistique descriptive avec rappels de cours**. Edition Dunod. Paris.
- Hamdani.H (2001) : **Statistique descriptive avec initiation aux méthodes d’analyse de l’information économique.** Edition OPU. Algérie.
- Hubler.J (2011) : **Statistique descriptive appliquée à la gestion et à l’économie**. Edition Bréal, 3éme édition. Paris.
- Leboucher.L & Voisin.M-J (2015) : **Introduction à la statistique descriptive**. Edition Cépadués, 3eme édition. Paris
- Monino.J-L, Kosianski.J.M & Cornu.F (2010) : **Statistique descriptive**. Edition Dunod, 4ème édition. Paris.
- **Séries de TD de Stat I** proposées par l’équipe pédagogique de la faculté SECG de l’université de Bejaia, composé de : Bouakline S, Meziani N., Yahi Z, Samoune T, Bouaissaoui S, Chalane T, Berrah K, Hamoudi Z, Amrani.S et Mousli A.

## Notations et abréviations

Les déférentes abréviations et notations utilisées tout au long de ce polycopié sont expliquées ci-dessous.

| | |
|---|---|
| $\sum_{i=1}^{k}$ | Somme pour i variant de 1 à k. |
| $\prod_{i=1}^{k}$ | Produit pour i variant de 1 à k. |
| $\lvert\ \rvert$ | Valeur absolue |
| $x_i$ | iéme modalité |
| X | Variable statistique |
| N | Effectif total |
| $n_i$ | Effectif ou fréquence absolue |
| $f_i$ | Fréquence relative |
| $n_i \uparrow$ | Effectif cumulé croissant. |
| $f_i \uparrow$ | Fréquence relative cumulée croissante |
| $a_i$ | Amplitude de la classe |
| $c_i$ | Centre de classe |
| $M_o$ | Mode |
| $M_é$ | Médiane |
| $\overline{X}$ | Moyenne arithmétique d'une série statistique X |
| G | Moyenne géométrique |
| H | Moyenne harmonique |
| Q | Moyenne quadratique |
| $t_X$ | Taux de croissance moyen |
| $q_x$ | Quartile |
| $D_X$ | Décile |
| $C_X$ | Centile |
| Var (x) | Variance de X |
| $\sigma(X)$ | Ecart-type |
| CV | Coefficient de variation |
| $m_k$ | Moments non centrés |
| $\mu_k$ | Moments centrés |

## Table des matières