Université Abderrahmane Mira-Bejaia
Faculté des Sciences Économiques, Commerciales et des Sciences de Gestion
Département des enseignements de base pour le domaine SEGC-LMD
Laboratoire d'Economie et Développement

Polycopié pédagogique

Dossier numéro (à remplir par l'administration) : ....

Titre

Statistique Descriptive

Cours destiné aux étudiants de

Première année SEGC-LMD

Année : 2020/2021

Avant propos

Ce présent support de cours de statistique descriptive est destiné aux étudiants de première année des Sciences Economiques, des Sciences de Gestion et des Sciences Commerciales (SEGC). Son objectif est de mettre à la disposition des étudiants un manuel qui leur permet d'acquérir des bonnes connaissances sur les statistiques. Le présent manuel est structuré et enrichie selon le programme national du module statistique descriptive.

Le support de cours aidera l'étudiant à comprendre comment étudier et analyser plusieurs phénomènes économiques. Nous avons explicité chaque élément traité par un exercice corrigé, pour faciliter à l'étudiant de comprendre comment appliquer l'ensemble des formules présentée. Cela permet à l'étudiant d'analyser lui-même plusieurs d'autres exercices. A la fin de chaque chapitre, nous avons proposé une série d'exercices.

Ce manuel est composé de quatre chapitres selon le programme officiel. Le premier chapitre traite des définitions et les généralités sur les statistiques descriptives. Il définit les notions de base que nous avons utilisé tous au long de ce manuel. Le deuxième chapitre porte sur la manière de présenter une série statistique dans un tableau et graphe, pour les deux types du caractère, qualitatif et quantitatif. Le troisième chapitre présente les paramètres de tendance centrale, ce qu'on appel les paramètres de position ; tels que le mode, la médiane, la moyenne, ainsi que les quantiles. Le dernier chapitre aborde les paramètres de dispersion et de forme.

Pour chaque chapitre, nous avons proposé une série d'exercices. L'ordre des exercices respect l'ordre des éléments traité dans chaque section de chaque chapitre. Nous avons proposés également, un sujet d'examen, avec un corrigé type, pour permettre à l'étudiant de se préparer pour l'examen final.

Chapitre I : Définitions et généralités

Introduction

Dans ce premier chapitre qui représente l'ensemble des notions de base de statistique descriptive, nous allons définir les principaux concepts utilisés dans le vocabulaire statistique, tels que ; la population, l'unité statistique, le caractère, ainsi que les types du caractère (Caractère qualitatif et caractère quantitatif). On termine par la définition d'une série statistique.

1. **Les statistiques et la statistique**

1.1.Les statistiques : Les statistiques représentent l'ensemble des données ou d'observations numérique sur un sujet bien déterminé. Par exemple : enquête sur la croissance démographique en Algérie, évolution d'un chiffre d'affaire dans une entreprise, taux de croissance de l'économie algérienne, nombre d'étudiants ayant le baccalauréat en 2020,etc.

1.2.La statistique : La statistique est une science qui consiste à traiter, à étudier et interpréter les données d'observation dans le but de tirer des conclusions et faire progresser convenablement la connaissance des phénomènes. Donc la statistique sert à étudier l'ensemble des statistiques recueillie.

2. **La population, l'unité statistique et l'échantillon :** Dans cet élément, on présente la population, l'unité statistique et l'échantillon.

2.1.La population statistique : C'est un ensemble d'unité statistique qui fera l'objet de l'étude.

2.2.L'unité statistique : On l'appelle aussi l'individu statistique, c'est chaque élément de la population statistique étudiée.

Exemple 1 : étude réalisée sur le nombre des maladies enregistré durant la période 1990 à 2000, selon les régions.

- La population statistique : Les maladies.
- L'unité statistique : Une maladie.

Exemple 2 : étude sur le nombre des entreprises qui ont certifiés les normes ISO, selon leurs types.

- La population étudiée : Les entreprises.
- L'unité statistique : Une entreprise.

Exemple 3 : Etude réalisée sur la quantité des produits électroménagers vendus par une entreprise, dans des marchés nationaux et internationaux.

- La population étudiée : Les produits électroménagers.
- L'unité statistique : Un produit électroménager.

2.3.L'échantillon : Est un sous-ensemble de la population considérée. Le nombre d'individus dans l'échantillon est la taille de l'échantillon.

3. Le Caractère ou la variable

C'est l'aspect particulier de l'unité statistique étudiée. C'est un caractère commun à tous les individus de la population ou l'échantillon étudier. Il permet de décrire les individus les uns des autres. Toute unité statistique peut être décrite selon un ou plusieurs caractères. Par exemple, si on souhaite étudier l'ensemble des salariés d'une entreprise, le caractère ça peut être leur âge, leur sexe, leur niveau socioprofessionnels, leurs nationalité, etc.

On prend les mêmes exemples de l'élément précédent, et on détermine pour chacun le caractère étudié.

Exemple 1 : Le caractère étudié est la région des malades.

Exemple 2 : Le caractère étudié est le type de l'entreprise.

Exemple 3 : Le caractère étudié est le marché destiné pour la vente des produits.

4. Les modalités du caractère

Les modalités du caractère, ce sont les différentes situations susceptibles d'être prise par le caractère. On prend toujours les mêmes exemples précédents pour déterminer les modalités possibles pour chaque caractère.

Exemple 1 : Les modalités possibles pour étudier les régions des malades sont :

- Modalité 1 : Est.
- Modalité 2 : Sud
- Modalité 3 : Ouest.
- Modalité 4 : Nord.

Exemple 2 : Les modalités possibles pour étudier le type des entreprises sont :

- Modalité 1 : SARL.
- Modalité 2 : SNC.
- Modalité 3 : SPA.
- Modalité 4 : EURL.

Exemple 3 : Les modalités possibles pour étudier les marchés destinés pour la vente des produits d'électroménagers sont :

- Modalité 1 : marchés nationaux.
- Modalité 2 : Marchés internationaux.

5. Les types du caractère

On trouve deux types du caractère, un caractère qualitatif et un caractère quantitatif.

5.1. Le caractère qualitatif : Un caractère il est dit qualitatif lorsque ses modalités ne sont pas mesurable (non mesuré par un nombre). Ses différentes modalités sont interprétées par des mots qui traduisent un état. Par exemple :

- La couleur : Rouge, noire, blanche, etc.
- La nationalité : algérienne, tunisienne, marocaine, etc.
- Le sexe, féminin et masculin.
- Le niveau d'instruction : Secondaire, primaire, universitaire.
- L'état matrimonial : célibataire, marié, divorcé et veuf.
- La mention : Excellente, très bien, passable et bien.

On trouve deux types du caractère qualitatif, un caractère qualitatif nominal et un caractère qualitatif ordinal.

5.1.1. Le caractère qualitatif nominal : le caractère est dit qualitatif nominal lorsque ses modalités ne peuvent pas être ordonnées (non classée par ordre). Si on prend les exemples de l'élément précédent, on trouve la couleur, la nationalité, le sexe, l'état matrimonial ce sont des caractères qualitatifs nominaux.

5.1.2. Le caractère qualitatif ordinal : le caractère est dit qualitatif ordinal lorsque ses modalités peuvent être ordonnées (classées par ordre). De l'exemple précédent on trouve le niveau d'instruction et la mention ce sont des caractères qualitatifs ordinaux

Exemple les modalités du niveau d'instruction peuvent être ordonnées comme suit : primaire, secondaire puis universitaire. Comme aussi, les modalités de la mention peuvent être ordonnées comme suit : passable, bien, très bien et en fin excellent.

5.2. Le caractère quantitatif

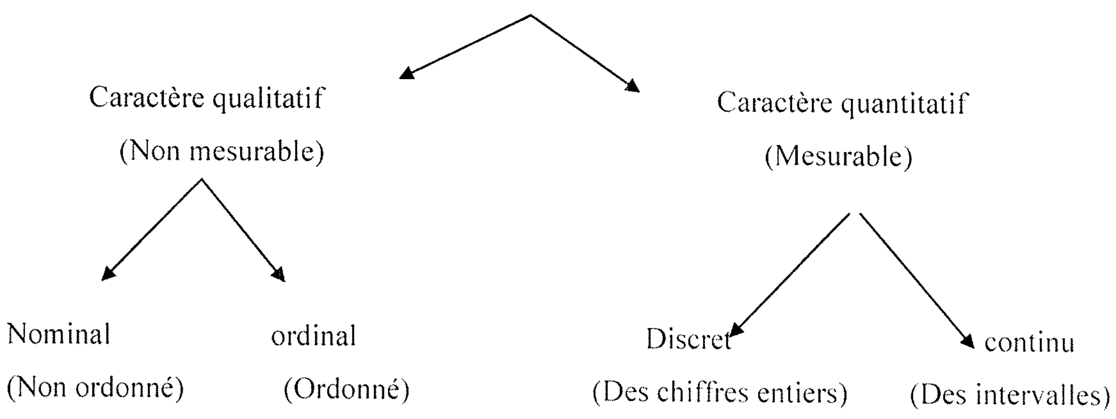
Le caractère est dit quantitatif lorsque ses modalités sont mesurables (mesuré par un nombre). Par exemple : l'âge est mesuré par jour, par mois et par an. La taille est mesurée par centimes, par mètre. Le chiffre d'affaire est mesuré en dinars, le nombre d'enfants est mesuré en chiffre.

On distingue deux caractères quantitatifs, un caractère quantitatif discret et un caractère quantitatif continu.

5.2.1. Le caractère quantitatif discret : le caractère est dit quantitatif discret s'il est mesuré par des nombres entier. par exemple : le nombre d'enfant par ménage. Ici le caractère qui est le nombre d'enfants se représente soit par 0, 1, 2, 3, on ne pourra pas diviser l'unité de la modalité (on ne peut pas dire un enfant et demi).

5.2.2. Le caractère quantitatif continu : Le caractère est dit quantitatif continu s'il est mesuré par des nombres indéterminés (des intervalles). Par exemple ; la répartition des salaries d'une entreprise selon leur âges. ici le caractère qui est l'âge peut être mesuré en année, en moins et en jour. Alors on présente les modalités dans des intervalles par tranche d'âge. Ce point sera bien détailler dans le prochain chapitre.

Les types du caractère



6. **La série statistique** (Les données statistiques)

On appel une série statistique (les données statistiques) la suite des valeurs prises pour une variable X sur les unités d'observation. Le nombre d'unité d'observation est noté par N et les valeurs de la variable X sont notées par :

$$x1, x2, x3, \dots\dots\dots. xi, \dots\dots xp$$; tels que le P est le nombre du modalité.

Exemple d'illustration : la série suivante représente les températures du mois de juin

32, 32, 30, 25, 24, 24, 25, 25, 27, 27, 28, 28, 28, 30,

30, 32, 33, 33, 33, 30, 33, 32, 33, 32, 32, 32, 30.

N = 30 (Le total des d'observation enregistré).

P= 07 (Nombre de modalité : 24, 25, 27, 28, 30, 32, 33).

Conclusion

Après la définition de la pluparts des concepts utilisés dans la statistique descriptive, que ce sont nécessaire pour avoir une idée générale sur l'ensemble des notions de base de cette matière. Le prochain chapitre sera consacré pour présenter une série statistique dans un tableau et de montrer comment la présenter graphiquement.

Exercices sur le chapitre I

Exercice n° 1 :

Déterminer parmi ces différentes études ; la population statistique étudiée, l'unité statistique, le caractère étudié et son type (sa nature).

1. La durée de vie des 50 lampes économiques fabriquées par une usine.
2. Le salaire annuel des employés d'une banque.
3. La mention du bac des étudiants de première année.
4. La couleur des stylos fabriqué par une usine.
5. La nationalité des 2000 touristes visiter le site Gouraya durant la période estival.
6. Nombre de voiture vendue par une entreprise.

Exercice n° 02 :

Un vendeur de radio TV à relever au cours du mois d'octobre le nombre d'appareil du TV vendues chaque jour, les observations obtenues sont les suivantes : (3, 5, 4, 2, 6, 4, 4, 2, 3, 4, 3, 5, 2, 3, 4, 5, 3, 2, 4, 5, 3, 5, 4, 6, 3, 6, 5, 3, 2, 1, 1).

1. Quelle est la population statistique étudiée, le caractère et sa nature ?
2. Donner le total de la population (N).
3. Déterminer les différentes modalités de la série statistique.

Exercice n° 03 :

La série suivante présente des résultats sur le nombre du logement vendus par jour par une agence immobilière, durant le mois de décembre.

1, 2, 2, 2, 0, 3, 4, 1, 1, 1, 0, 0, 4, 0, 3, 3, 1, 2, 2, 2, 0, 0, 4, 5, 3, 5, 2, 1, 1, 3, 1.

1. Déterminer la population statistique étudiée, le caractère et sa nature.
2. Donner le total de la population.
3. Déterminer les différentes modalités de la série statistique.

Chapitre II : Présentation d'une série statistique : Tableaux et graphes

Introduction

1. Notion d'effectif et de fréquence
2. Présentation d'une série statistique sous forme des tableaux.
   - 2.1.Présentation d'un caractère qualitatif.
   - 2.2.Présentation d'un caractère quantitatif.
     - 2.2.1. Caractère quantitatif discret.
     - 2.2.2. Caractère quantitatif continu.
3. Présentation graphique d'une série statistique.
   - 3.1.Présentation graphique d'un caractère qualitatif.
   - 3.2.Présentation graphique d'un caractère quantitatif.
     - 3.2.1. Caractère quantitatif discret.
     - 3.2.2. Caractère quantitatif continu.

Conclusion

Chapitre II : Présentation d'une série statistique : Tableaux et graphes

Introduction

Les observations statistiques recueillies dans une étude représentent une série statistique. Pour étudier et traiter l'ensemble de ses observations, il faut, tout d'abord les ordonnés et les présenter sous forme de tableaux et graphes. Cela permet de traduire concrètement les observations. Ce présent chapitre sera consacré à la présentation d'une série statistique sous forme de tableaux et graphes. Mais, on met l'accent, tout d'abord, sur la définition de la notion d'effectif et de fréquence.

1. Notion d'effectif et de fréquence

Soit la série statistique suivante, qui représente un caractère X à k modalités.

$$Xi = x1, x2, x3, \dots\dots . xk$$

Une série statistique se représente dans un tableau, par deux colonnes, la première colonne correspond aux caractères (noté xi) et la deuxième colonne correspond aux effectifs (noté ni) ou fréquences (notée f i).

- ✓ On appel xi, le nombre des modalités qui représente une série.
- ✓ On appel l'effectif (ni) d'une modalité (xi), le nombre de fois que cette modalité apparaisse dans une série statistique. On note (ni) l'effectif de (xi)

On a la sommes des effectifs ni égale au total de la population.

On note : $\left(N = \sum_{i=1}^{k} n_i = n_1 + n_2 + n_3 + \dots + n_k \right)$ (Hamdani, 2001)

- ✓ On appel la fréquence relative (f i), l'effectif ni divisé sur le total d'effectif (N).

$$fi_i = \frac{ni}{N} \quad \text{(Barthe, 1989)}$$

2. Présentation d'une série statistique sous forme des tableaux

Pour présenter une série statistique, on distingue tout d'abord les différentes modalités du caractère étudié notés par xi. Ensuite, nous nous déterminons pour chaque modalité le nombre d'individus observés. Il s'agit donc de l'effectif ni. Sur le tableau, nous pouvons déterminer aussi les fréquences relatives fi.

Présentation des effectifs (ni) et les fréquences relatives (fi) en tableau (Cas général)

Modalités xi	Effectifs ni	Fréquences Relatifs $f_i = n_i / N$
x_1	n_1	f_1
x_2	n_2	f_2
x_3	n_3	f_3
…	…	…
…	…	…
x_p	n_p	f_k
Total	N	1

Dans l'élément suivant on va présenter une série statistique dans un tableau, mais on prend en considération les types du caractère, tels que le caractère qualitatif et le caractère quantitatif.

2.1. Présentation d'un caractère qualitatif

Nous avons déjà défini le caractère qualitatif qui est lié à des observations qui ne feront pas l'objet d'une mesure. Pour présenter un caractère qualitatif en tableau on propose l'exemple suivant :

Exemple : Une enquête réalisée par nous même, pour différencier le type de 30 entreprises enquêtées. Les résultats sont présentés comme suit : 19 entreprises industrielles, 2 entreprises commerciales et 9 entreprises de service.

- Présenter la série statistique dans un tableau.

Solution : Nous avons dans ce cas 3 modalités du caractère, qui sont le type des entreprises : industrielle, commerciale et de service.

Les résultats se présentent dans un tableau comme suit :

Type de l'entreprise	Effectif ni	Fréquence relative fi	Pourcentage fi %
- Industrielle	19	0,633	63,3
- Commerciale	2	0,067	6,7
- De service	9	0,30	30
Total	$N = 30$	$Fi = 1$	100

Commentaire : Le tableau montre que la majorité des entreprises étudiées ce sont des entreprises industrielles, soit 63,3%.

2.2. Présentation d'un caractère quantitatif

Il existe deux caractères quantitatifs, un caractère quantitatif discret et un caractère quantitatif continu. On expose comment présenter les deux cas dans des tableaux avec des exemples concrets.

2.2.1. Caractère quantitatif discret

Exemple : soit la série suivante qui représente la répartition de 30 ménages selon le nombre d'enfants.

0, 1, 1, 3, 3, 2,4, 3, 1, 0, 2, 4, 1,5, 2, 2, 1, 1, 2, 1, 3, 4, 2, 4, 5, 0, 5, 3, 2, 0.

a. Construire le tableau d'effectifs et de fréquences.
b. Calculer les effectifs cumulés croissants et décroissants.
c. Calculer les fréquences cumulées croissantes et décroissantes.

Solution :

Nous avons le nombre d'observations égale à 30, donc le total d'effectif $N = 30$. Nous avons 6 modalités qui représentent le nombre d'enfants pour chaque ménage : 0, 1, 2, 3, 4, 5 telles que ;

- La modalité $x_1 = 0$ est répétée 4 fois, donc $n_1 = 4$.
- La modalité $x_2 = 1$ est répétée 7 fois, donc $n_2 = 7$.
- La modalité $x_3 = 2$ est répétée 7 fois, donc $n_3 = 7$.

- La modalité $x_4 = 3$ est répétée 5 fois, donc $n_4 = 5$.
- La modalité $x_5 = 4$ est répétée 4 fois, donc $n_5 = 4$.
- La modalité $x_6 = 5$ est répétée 3 fois, donc $n_6 = 3$.

a. Le tableau d'effectifs et de fréquences :

Nombre d'enfants xi	Effectifs ni	Effectifs cumulés croissants $n_i \uparrow$	Effectif cumulés décroissants $n_i \downarrow$	Fréquences relatives fi	Fréquences relatives croissantes $fi \uparrow$	Fréquences relatives décroissante $f_i \downarrow$
0	4	4	30	0,13	0,13	1
1	7	11	26	0,23	0,36	0,87
2	7	18	19	0,23	0,59	0,64
3	5	23	12	0.17	0.76	0.41
4	4	27	7	0,14	0,9	0.24
5	3	30	3	0,1	1	0,1
total	$N = 30$	-	-	1	-	-

b. Calcule des effectifs cumulés croissants et décroissants :

✓ On appel effectif cumulés croissants (Noté $n_i \uparrow$) associés à une valeur, la somme des effectifs des valeurs inférieurs. Dans notre exercice, la première valeur de ($n_i \uparrow$) est 4, puis, (4 + 7) = 11. (11 + 7) = 18. (18 + 5) = 23. (23 + 4) = 27. (27 + 3) = 30. (Les résultats sont présentés dans le tableau précédent).

✓ On appel les effectifs cumulés décroissants (Noté $n_i \downarrow$) associés à une valeur, la somme des effectifs des valeurs inférieurs. Dans notre exercice, la première valeur de ($n_i \downarrow$) est 30, puis (30 − 4) = 26. (26 − 7) = 19. (19 − 7) = 12. (12 − 5) = 7. (7 − 4) = 3. (Les résultats sont présentés dans le tableau précédent).

C. Calcule des fréquences cumulées croissantes et décroissantes :

On appel les fréquences cumulées croissantes (Notée $f_i \downarrow$) et décroissantes (Notée) se calculent comme les effectifs cumulés croissants et décroissants :

- ✓ Dans notre exercice, la première valeur de ($f_i \uparrow$) est 0,13, puis, (0,13 + 0,23) = 0,36. (0,36 + 0,23) = 0,59. (0,59 + 0,17) = 0,76. (0,76 + 0,14) = 0,9. (0,9 + 0,1) = 1. (Les résultats sont présentés dans le tableau précédent).
- ✓ Même chose pour le calcule des valeurs de ($f_i \downarrow$), la première valeur est 1, puis (1 – 0,13) = 0,87. (0,87 – 0,23) = 0,64. (0,64 – 0,23) = 0,41. (0,41 – 0,17) = 0,24. (0,24 – 0,14) = 0,1. (Les résultats sont présentés dans le tableau précédent).

2.2.2. Caractère quantitatif continu

Dans la présentation d'un caractère quantitatif continu, on distingue deux cas : cas d'amplitude égale et cas d'amplitude inégale.

- ➢ **Cas d'amplitude égale** : le tableau suivant représente la répartition des ouvriers d'une entreprise selon leur salaire horaire.

Salaire	[20 - 25[	[25 - 30[	[30 - 35[	[35 - 40[	Total
Effectif	23	33	25	19	100

a. Calculer l'amplitude de chaque classe.
b. Calculer le centre de chaque classe.
c. Calculer les effectifs cumulés croissants et décroissants, ainsi que les fréquences cumulées croissantes et décroissantes.

Solution

a. **Calcule de l'amplitude (Notée ai)** : l'amplitude d'une classe est la largeur de l'intervalle de cette même classe. On la calcule comme suit :

$$a_i = e_i - e_{i-1}$$

e_i : S'appelle la borne inférieure.

e_{i-1} : S'appelle la borne supérieure.

b. **Le centre de classe (Notée Ci)** : C'est la valeur moyenne des valeurs des extrémités de la classe. On le calcule comme suit :

$$C_i = \frac{e_{i-1} + e_i}{2}$$ (Chauvat, 2002)

c. Tableaux des effectifs et des fréquences : On calcule les effectifs cumulés croissants et décroissants ainsi que les fréquences cumulées croissantes et décroissantes de la même manière que le cas discret. (Les résultats sont présentés dans le tableau suivant).

Salaire	Effectif ni	Amplitude ai	Centre de classe ci	Eff. Cum. Crois. $n_i \uparrow$	Eff. Cum. Décr. $n_i \downarrow$	Fréq. relat. fi	Fréq. Relat. Crois. $f_i \uparrow$	Fréq. Relat. Crois. $f_i \downarrow$
[20-25[	23	5	22,5	23	100	0,23	0,23	1
[25-30[	33	5	27,5	56	77	0,33	0,56	0,77
[30-35[	25	5	32,5	81	44	0,25	0,81	0,44
[35-40[	19	5	37,5	100	19	0,19	1	0,19
Total	100	-	-	-	-	1	-	-

- **Cas d'amplitudes inégales** : Dans le cas des amplitudes inégales, on doit corriger les effectifs, soit par le calcule des effectifs corrigés (n_{ic}), soit par le calcule de la densité (d_i).

$$n_{ic} = n_i \times \frac{\text{amplitude la plus petite}}{\text{amplitude de la classse considérée}}$$

$$d_i = \frac{\text{l'effectif de la classe considéré}}{\text{l'amplitude de la classse considérée}}$$ (Chauvat, 2002)

Exemple : soit la distribution suivante qui représente la répartition de 50 diplômés selon la durée du chômage.

- On calcule l'amplitude et le centre de classe pour chacune des classes, ainsi que les ni corrigés et les densités.

Durée du chômage	Effectif ni	ai	Centre de classe ci	ni corrigé	Densité $di=ni/ai$
[0-1[	02	1-0 = 1	0+ 1 /2 = 0,5	2	2
[1-3[	04	3-1 = 2	1+3 /2 =1,5	2	2
[3-6[	09	6-3 = 3	3+6 /2 = 4,5	3	3
[6-12[	21	12-6 = 6	6+12 /2 = 9	3,5	3,5
[12-18[	06	12-18 = 6	12+18 /2 = 15	1	1
[18-24[	08	18-24 = 6	18+24 /2 =21	1,33	1,33
Total	50	-	-	-	-

- Pour calculer les n_i corrigés on applique la formule précédente :

$n_{1c} = 2 \times \frac{1}{1} = 1$, ensuite $n_{2c} = 4 \times \frac{1}{2} = 2$, $n_{3c} = 9 \times \frac{1}{1} = 3$, $n_{4c} = 21 \times \frac{1}{1} = 3{,}5$, $n_{5c} = 6 \times \frac{1}{1} = 1$, $n_{6c} = 8 \times \frac{1}{1} = 1{,}33$.

Remarque : Pour le calcule des effectifs cumulés croissants et décroissants et les fréquences cumulées croissantes et décroissantes se fait de la même façon que le cas des amplitudes égales.

3. Présentation graphique d'une série statistique

La présentation graphique pour une série statistique est très importante, elle donne une image plus claire et facile à commenter. Elle permet de traduire concrètement les donnés. La présentation graphique d'une série se diffère selon le caractère étudié. Dans cet élément on essaye du présenter graphiquement les deux caractères qualitatif et quantitatif.

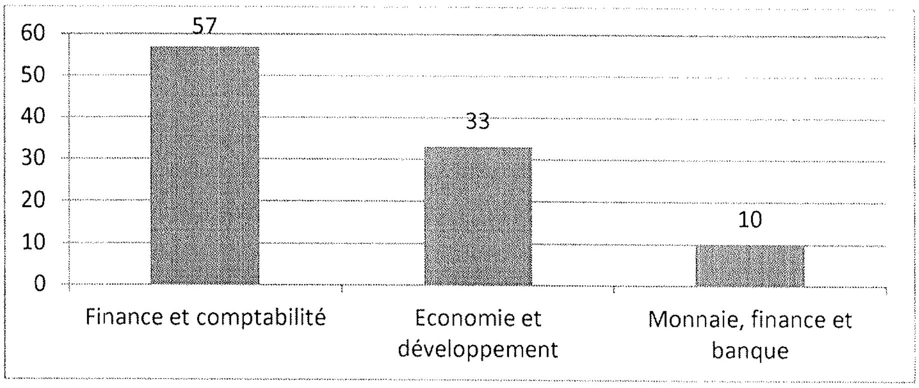
3.1.Présentation graphique d'un caractère qualitatif :

On présente un caractère qualitatif graphiquement par deux types, soit par un diagramme en colonne (dit aussi diagramme en tuyaux d'orgue), soit par un diagramme circulaire.

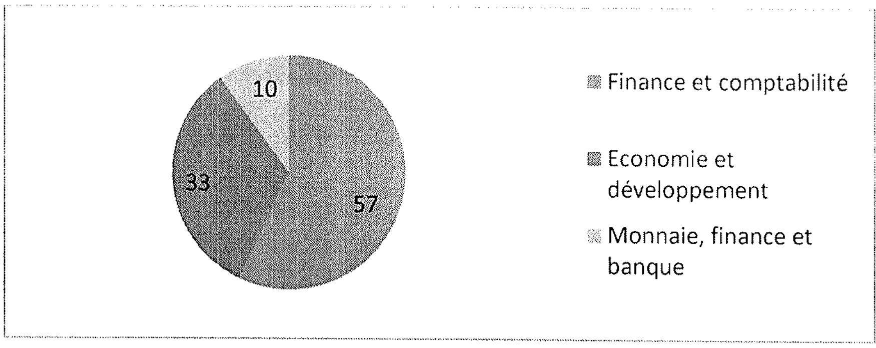
Exemple : La série suivante représente la spécialité choisie par 100 étudiants de première année, présenter la série graphiquement.

Choix des étudiants xi	**Effectif** ni	fi	fi **360°**
Finance et comptabilité	57	0,57	205,2
Economie et développement	33	0,33	118,8
Monnaie, finance et banque	10	0,1	36
Total	**100**	**1**	**360**

Solution : Présentation par le diagramme en colonne (Tuyaux d'orgue)



Présentation par le diagramme en secteur



3.2.Présentation graphique d'un caractère quantitatif

La présentation graphique d'un caractère quantitatif se fait selon le type du caractère, discret (discontinu) ou continu.

3.2.1. Caractère quantitatif discret

On présente graphiquement un caractère quantitatif discret par deux types du graphe :

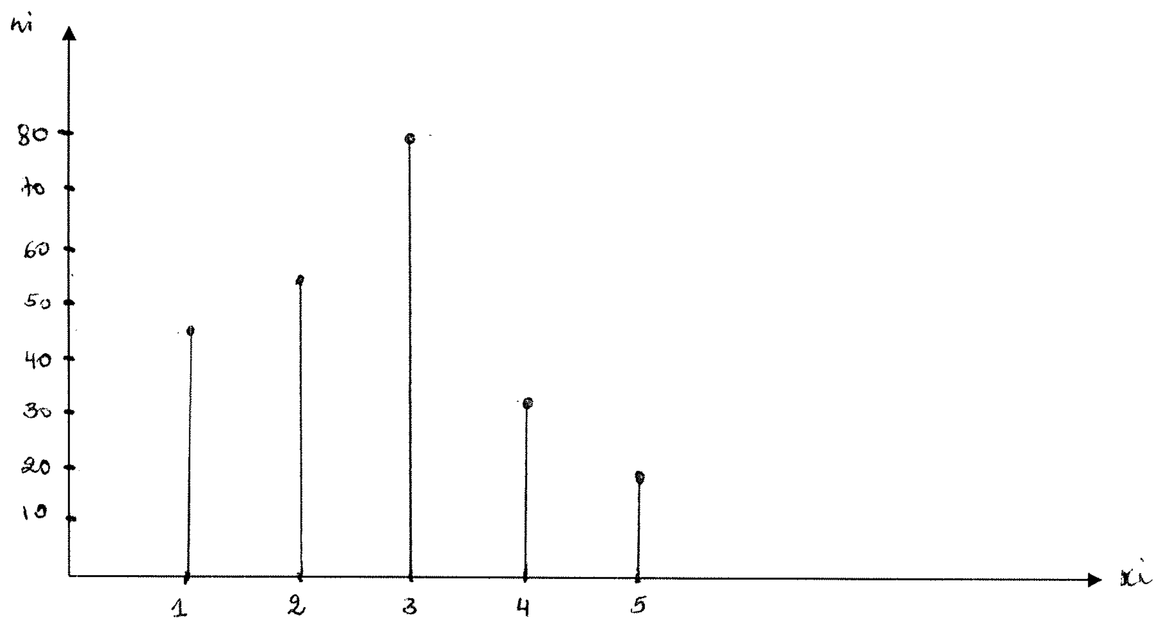
- **Un diagramme en bâton :** Pour présenter les effectifs absolu (ni), ou les fréquences relatifs (fi).
- **Un diagramme en escalier** (Dite aussi, les courbes en escalier) : Pour présenter les effectifs cumulés croissants et décroissants ($n_i \uparrow$), ($n_i \downarrow$), et les fréquences cumulées croissantes et décroissantes ($f_i \uparrow$), ($f_i \downarrow$).

Exemple : Une étude sur la population d'une cité résidentielle, selon le nombre de personnes par ménage, fait sortir les résultats présentés dans le tableau suivant :

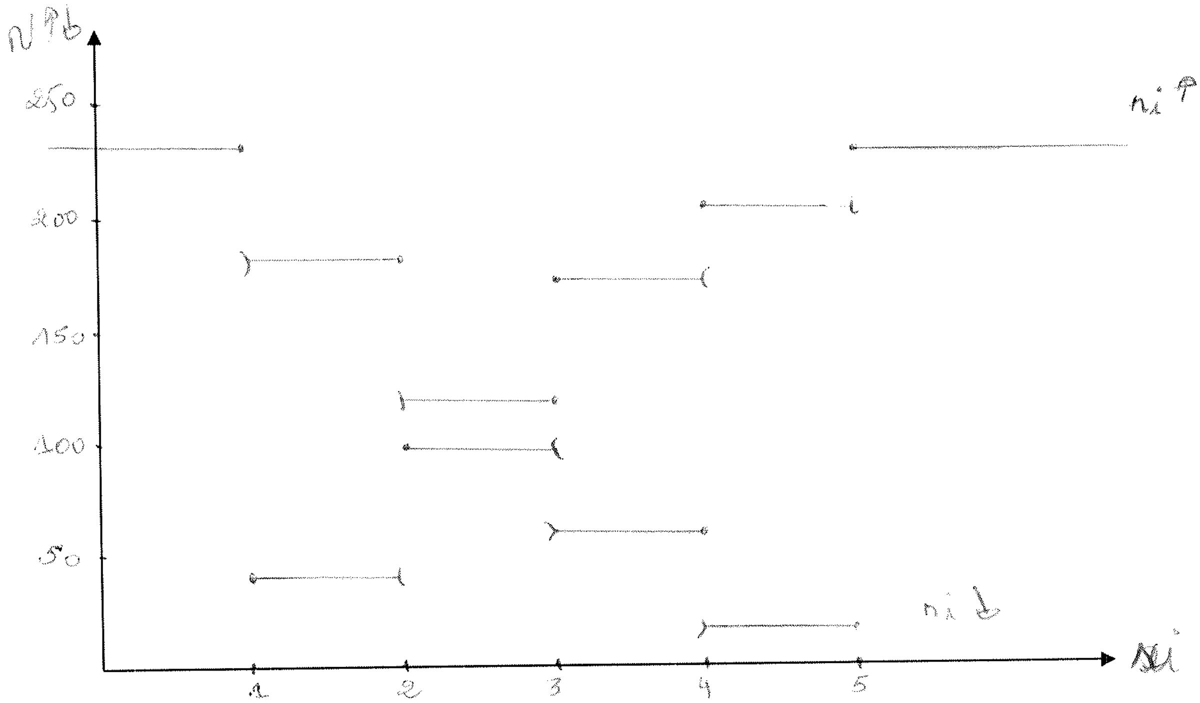
a. Présenter la série graphiquement.
b. Présenter les effectifs cumulés croissants et décroissants.

Nombre de personnes xi	Nombre de ménages ni	$n_i \uparrow$	$n_i \downarrow$
1	45	45	233
2	55	100	188
3	80	180	133
4	33	213	53
5	20	233	20
Total	233	-	-

a. **Présentation graphique de la série statistique par le diagramme en bâton**



b. Présentation des effectifs cumulés croissants et décroissants par le diagramme en escalier



3.2.2. Caractère quantitatif continu

On présente graphiquement un caractère quantitatif continu par deux types du graphe :

- **Un histogramme :** Pour présenter les effectifs absolus (ni), ou les fréquences relatifs (fi).
- **Les courbes cumulatives :** Pour présenter les effectifs cumulés croissants et décroissants ($n_i \uparrow$), ($n_i \downarrow$), et les fréquences cumulées croissantes et décroissantes ($f_i \uparrow$), ($f_i \downarrow$).

Remarque : Dans le cas des amplitudes inégales, on présente l'histogramme par les valeurs des effectifs corrigés (ou par les valeurs de la densité).

Exemple 1 (Cas des amplitudes égales) :

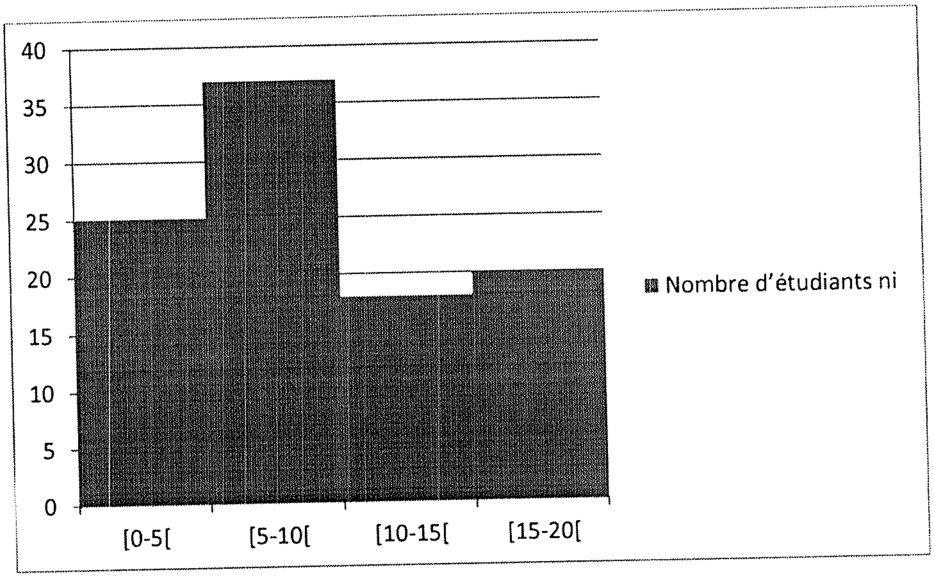
Soit le tableau suivant qui représente la répartition du 50 étudiant, selon les notes obtenues dans le module statistique descriptive.

a. Représenter la série graphiquement.

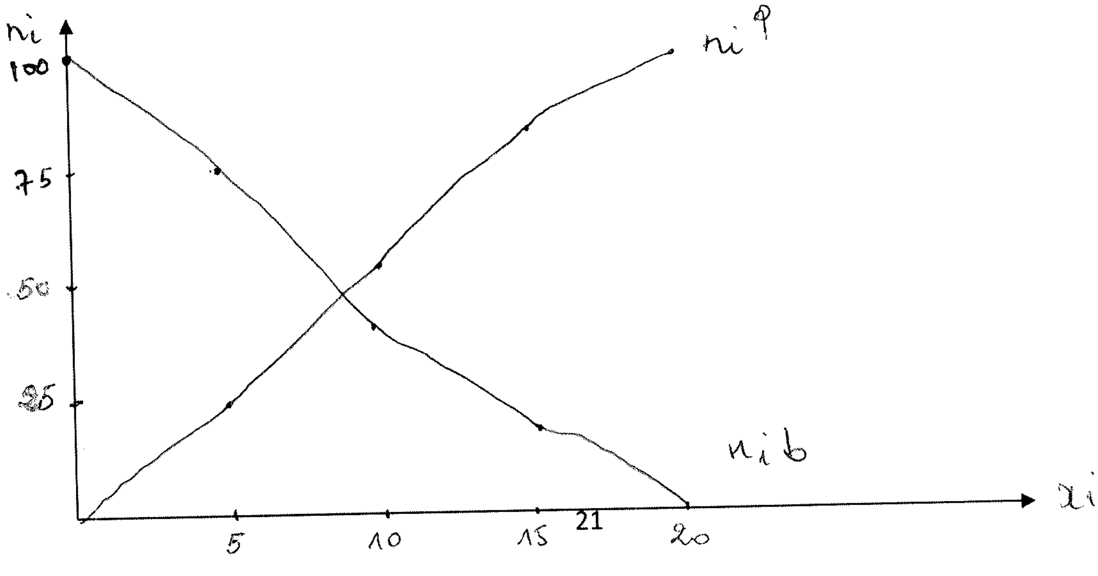
b. Représenter les effectifs cumulés croissants et décroissants.

Les notes xi	Nombre d'étudiants ni	ai	$n_i \uparrow$	$n_i \downarrow$
[0-5[	25	5	25	100
[5-10[	37	5	62	75
[10-15[	18	5	80	38
[15-20[	20	5	100	20
Total	100	-	-	-

a. Présentation graphique par un histogramme



b. Présentation des effectifs cumulés croissants et décroissants

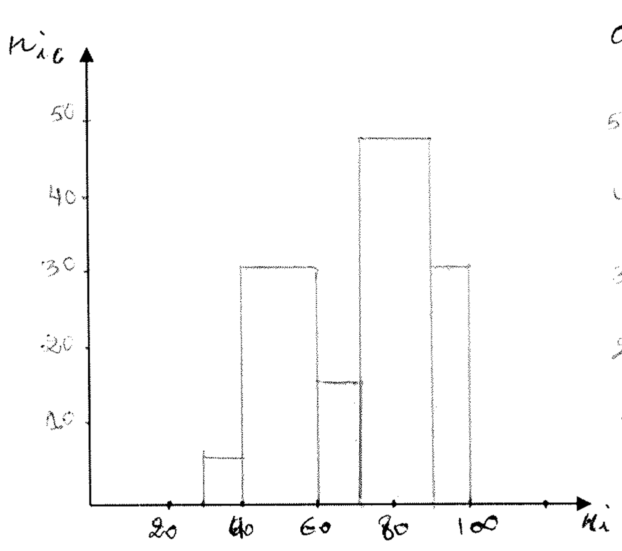


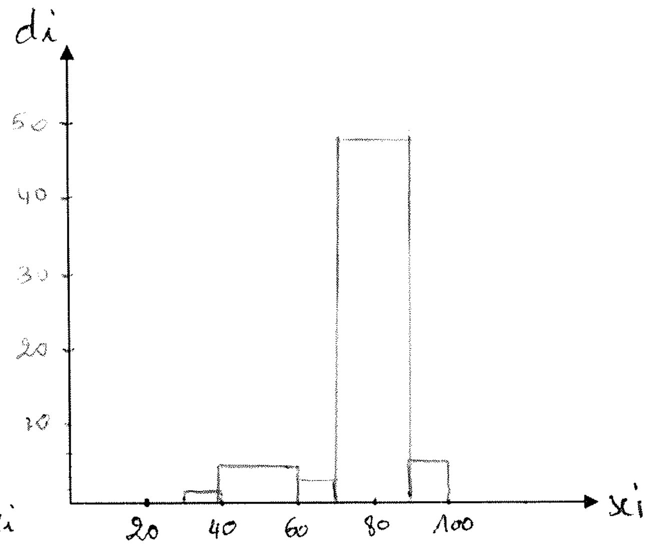
Exemple 2 : (Cas des amplitudes inégales)

Ce tableau représente la répartition du nombre de familles ayant un enfant étudiant en première année à l'université, en fonction des dépenses annuelles.

Dépenses annuelles	Effectif ni	ai	n_{ic}	di
[30-40[	5	10	5	0,5
[40-60[	60	20	30	2
[60-70[	15	10	15	1,5
[70-90[	95	20	47,5	47,5
[90-100[	30	10	30	3
Total	205	-	-	-

Pour présenter la série graphiquement, par un histogramme, on prend, soit, les valeurs du ni corrigés, soit, les valeurs de la densité.





- **Remarque** : pour la présentation des effectifs cumulés croissants et décroissants et les fréquences cumulées croissantes et décroissantes on utilise les valeurs de $(n_i \uparrow)$, $(n_i \downarrow)$, $(f_i \uparrow)$, $(f_i \downarrow)$. Comme le cas des amplitudes inégales.

Conclusion :

La présentation d'une série statistique dans un tableau et sa représentation par le graphe approprié est une phase très importante pour faciliter l'interprétation des observations. Mais, quand il s'agit de plusieurs observations, l'interprétation sera un peu compliquer. Pour cela il existe d'autres utiles du traitement tels que le mode, la médiane, la moyenne, etc. ce qu'on appel les paramètres de position, qui sera l'objet de notre prochain chapitre.

Exercice sur le chapitre II

Exercice n°01 :

Les résultats récapitulés ci-dessous représentent le nombre des étudiants de la faculté SEGC de l'université de Bejaia pour l'année 2019/2020, selon les spécialités.

Les spécialités xi	**Le nombre des étudiants** ni
Formation initial SECG-LMD.	2441
Science économique.	1216
Science commerciale.	1177
Science de gestion.	2600
Total	**7434**

1. Déterminer la population statistique étudiée, le caractère et sa nature.
2. Représenter la série graphiquement par le diagramme en colonne.
3. Représenter la série graphiquement par le diagramme circulaire.

Exercice n°2

Soit la série suivante qui représente le nombre d'accident de la circulation durant une période de 100 jours.

Nombre d'accidents par jour	0	1	2	3	4	5	6	7	Total
Nombre des jours	13	27	27	19	9	3	1	1	100

1. Déterminé la population statistique étudiée, le caractère et son type.
2. Représenter la série graphiquement.
3. Calculer les fréquences relatives (fi), et les fréquences relatives croissantes et décroissantes.
4. Calculer les effectifs cumulés croissants et décroissants et représenter-les graphiquement.
5. Quel est le nombre de jour ou ils ont enregistré plus de 4 accidents ?
6. Quel est le nombre de jour ou ils ont enregistré mois de 3 accidents ?

Exercice n°03 :

Une enquête réalisée sur le nombre des personnes qui ont visités un certain musé, selon leur âge, fait sortir les résultats suivants :

Age xi	**Effectif** ni
[15 - 25[	96
[25 - 35[	118
[35 -50[	138
[50 -65[	101
[65- 80[	47
Total	**500**

1. Déterminé la population étudiée, le caractère et son type.
2. Calculer l'amplitude et le centre de chaque classe.
3. Est-il nécessaire de corrigé l'effectif ?
4. Représenter la série graphiquement.
5. Calculer les effectifs cumulés croissants et décroissants et représenter-les graphiquement.
6. Calculer les fréquences et les fréquences cumulées croissantes et décroissantes.

Chapitre III : Les paramètres de tendance centrale

Introduction

1. Le mode
   1.1.Cas d'une variable discrète.
   1.2.Cas d'une variable continu.
2. La médiane
   2.1.Cas d'une variable discrète.
   2.2.Cas d'une variable continu.
3. La moyenne arithmétique.
   3.1.Cas d'une variable discrète.
   3.2.Cas d'une variable continu.
4. Autres types de moyenne.
   4.1.La moyenne géométrique.
   4.2.La moyenne harmonique.
   4.3.La moyenne quadratique.
5. Les quantiles
   5.1.Les quartiles.
   5.2.Les déciles.
   5.3.Les centiles.

Conclusion

Chapitre III : Les paramètres de tendance centrale

Introduction

La présentation d'une série statistique par les tableaux et graphes, que nous avons présentée dans les deux premiers chapitres, sont considéré comme une étape essentiel pour prendre une idée globale du phénomène étudié, ainsi pour approfondir l'analyse et rendre l'interprétation des informations renquillés plus objective, surtout quand la série statistique comporte un grand nombre d'observations, on utilise ce qu'on appelle les paramètres de tendance centrale (les paramètres de positions).

Nous avons trois caractéristique de tendance centrale utilisé fréquemment ; le mode, la médiane et la moyenne, ainsi que les quantiles.

1. Le mode (Noté Mo)

Le mode d'une série statistique est la valeur de la variable X qui correspond au plus grand effectif ou fréquence.

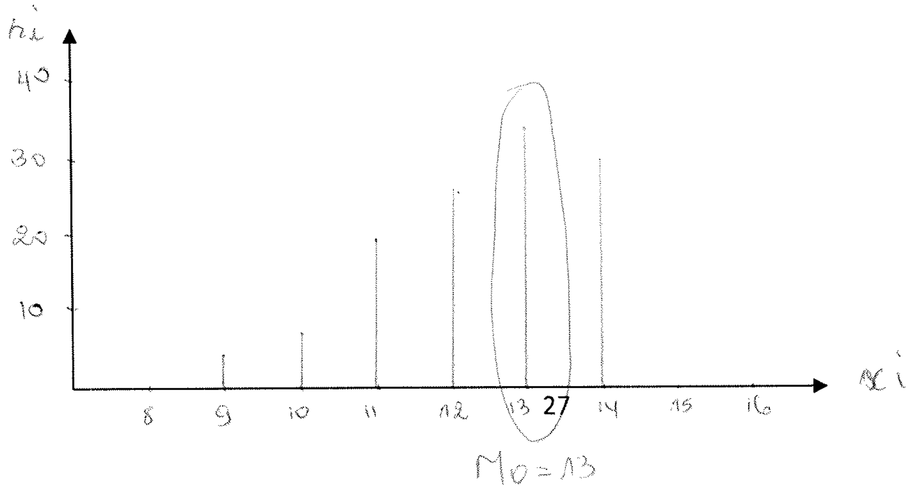
1.1.Cas d'une variable discrète : Lorsque la variable statistique est discrète, le mode correspond à la variable de Xi exacte.

Exemple : Les notes des étudiants

Les notes xi	09	10	11	12	**13**	14	Total
Nombre d'étudiants ni	4	6	20	15	**34**	28	107

Le mode = 13 (il correspond à la valeur de xi qui a l'effectif le plus grand ni = 34).

Le mode graphiquement :



Remarque : La distribution statistique est dite Bimodale lorsque correspond deux valeurs modales de la variable Xi deux grands effectifs. Dans ce cas on aura deux modes M_{O1} et M_{O2}.

Exemple : La série suivante représente les notes des étudiants de première année de la section H.

Les notes xi	09	10	**11**	12	**13**	14	Total
Nombre d'étudiants ni	05	13	**28**	18	**28**	15	107

Dans ce cas nous avons deux valeurs du mode : M_{O1} = 11 et M_{O2} = 13.

1.2.Cas d'une variable continue : Lorsque la variable est continue, la détermination de la valeur du mode n'est pas la valeur de xi exacte. Puisque les valeurs du xi sont présentées par une classe. Alors, dans ce cas on parle de la classe modale qui correspond à l'effectif le plus grand. Pour calculer le mode on utilise la formule suivante :

$$Mo = e_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i , \qquad \text{(Bahouayila, 2016)}$$

tels que ;

M_O : La valeur du mode

e_{min} : La valeur inférieure de la classe modale.

ai : L'amplitude de la classe modale.

Δ_1 : La différence entre la fréquence absolue (ou relative) de la classe modale et celle de la précédente.

Δ_2 : La différence entre la fréquence absolue (ou relative) de la classe modale et celle de la suivante.

Exemple n° 1 (Cas des amplitudes égales) :

Lorsque les amplitudes sont égales, la classe modale correspond à la fréquence absolue ni (ou relative fi) la plus élevée.

Le tableau suivant représente la répartition des salaires de 100 ouvriers d'une banque.

Salaire *xi*	Ouvrier *ni*
[10-20[	15
[20-30[	20
[30-40[	**30**
[40-50[	15
[50-60[	12
[60-70[	08
Total	100

$\Delta_1 = (30-20) = 10$

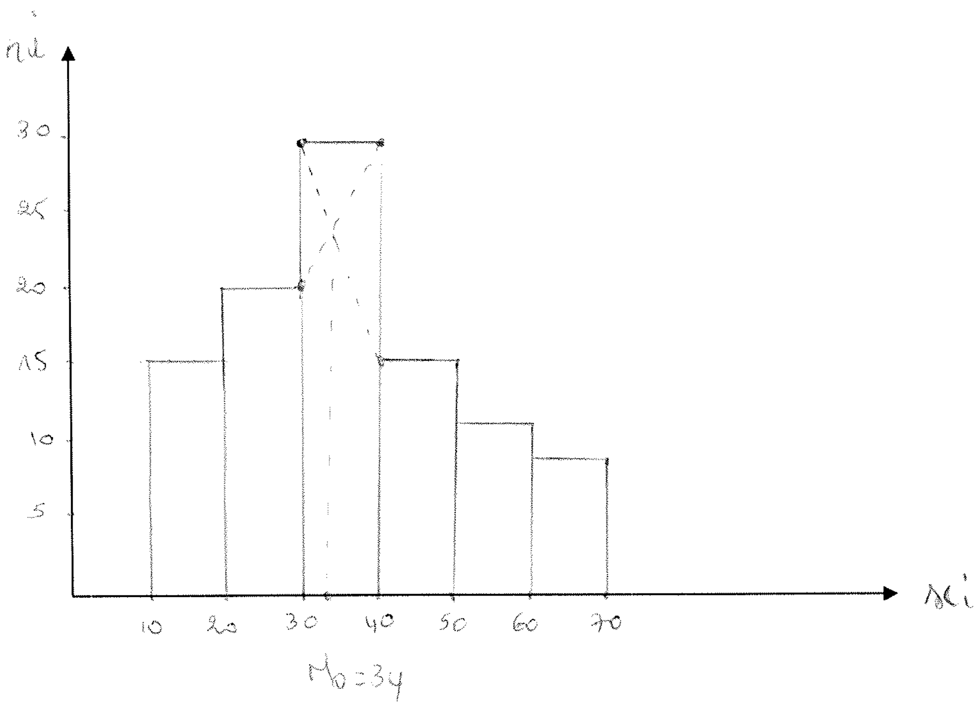
$\Delta_2 = (30-15) = 15$

On trouve tout d'abord la classe modale. L'effectif le plus grand est 30, donc la classe modale est [30-40[. On applique la formule précédente pour calculer la valeur du mode :

$$Mo = e_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i ,$$

$$Mo = 30 + \frac{30-20}{(30-20)+(30-15)} \times 10 = 34Da$$

Le mode graphiquement : Dans le cas continu on détermine le mode sur l'histogramme



Lorsque les amplitudes sont inégales, la classe modale correspond à la fréquence absolue corrigées n_{ic} (ou relative fi) la plus élevée.

Le tableau suivant représente le salaire des 119 ouvriers par heur.

Le salaire xi	Les ouvriers ni	ai	ni corrigé n_{ic}
[10 - 20[	12	10	12
[20 - 30[	13	10	13
[30 – 50[	**32**	**20**	**16**
[50 - 80[	47	30	15,66
[80 -90[	15	10	15
Total	119	-	

La classe modale est [30 – 50[, elle correspond à l'effectif corrigé le plus élevé $n_{ic} = 16$. Pour calculer le mode, on applique la formule du mode, on aura :

$$Mo = e_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i$$

$$Mo = 30 + \frac{16 - 13}{(16 - 13) + (16 - 15{,}66)} \times 20 = 47{,}64 / h$$

2. La médiane (Notée Me)

Est la valeur de la variable statistique qui partage la série statistique en deux sous-ensembles égaux, représentant chacun 50% des effectifs. Sachant que, les valeurs de la variable doivent être rangées par un ordre croissant.

2.1.Cas d'une variable discrète :

Exemple 1 : On considère la série simple suivante qui représente les notes de 9 étudiants

7, 8, 9, 10, **13**, 14, 15, 17, 19.

- La médiane Me = 13 (est la valeur qui partage la série en deux)
- Interprétation : Au moins 50% des étudiants ont une note plus que 13.

Exemple 2 : On considère la série simple suivante qui représente les notes de 8 étudiants

7, 8, 9, **10, 11**, 12, 13, 14.

- La médiane Me = $\frac{10+11}{2}$ = 10,5.
- Interprétation : Au moins 50% des étudiants ont une note plus que 10,5.

NB : une série simple veut dire le nombre d'observation est faible et que chaque modalité n'est prise qu'une seule fois.

Exemple 3 : on considère une série statistique pondéré, dans ce cas la médiane se calcule comme suit :

- $M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$ si N est pair
- $M_e = X\frac{N+1}{2}$ si N est impair (Salmon, 2014)

Cas de N impair : Nous avons la série suivante qui représente les températures de la première quinzaine du mois d'octobre.

Les températures xi	Les jours ni	$n_i \uparrow$
8	2	2
9	3	5
13	3	8
17	5	13
18	2	15
Total	15	-

N = 15 impair, donc la médiane se calcule par la formule : $M_e = X\frac{N+1}{2}$

$$M_e = X\frac{15+1}{2} + X\ 8 = 13°$$

- **Interprétation** : 50% du jour ou nous avons enregistré une température plus que 13°, et 50% du jour nous avons enregistré une température moins que 13°.

Cas de N pair : nous avons la série suivante qui représente les températures des 30 jours du mois d'octobre.

Les températures xi	Les jours ni	$n_i \uparrow$	f_i	$f_i \uparrow$
16	3	3	0,1	0,1
17	5	8	0,17	0,27
18	4	12	0,13	0,4
19	5	17	0,17	0,57
20	7	24	0,23	0,8
21	4	28	0,13	0,93
22	2	30	0,07	1
Total	30	-	1	-

$N = 30$, pair, donc la médiane se calcule comme suit :

$$M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \qquad M_e = \frac{x_{\frac{30}{2}} + x_{\frac{30}{2}+1}}{2} \qquad M_e = \frac{x16 + x17}{2}$$

x_{16} : On cherche la valeur de x qui correspond à l'effectif cumulé n supérieur ou égale à 16.

x_{17} : On chercher la valeur de x qui correspond à l'effectif cumulé n supérieure ou égale à 17.

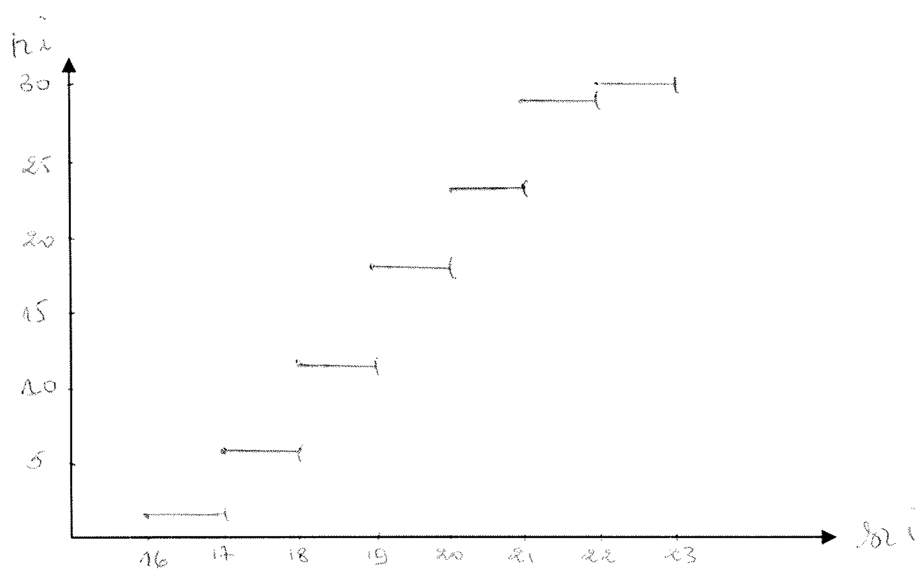
$$M_e = \frac{19+19}{2} = 19°$$

- **Interprétation :** au moins 50% des jours du moins d'octobre, les températures enregistré dépassent 19°.

Remarque : Nous pouvons utiliser les valeurs des fréquences cumulées croissantes pour déterminer la médiane, il suffit de prendre la valeur de x qui correspond à la valeur qui est supérieur ou égale à 0,5 ($f_i \uparrow$).

Si on prend l'exemple précédent on trouve, Me = 19°

La médiane graphiquement : Dans le cas discret, la médiane se détermine graphiquement sur le diagramme en escalier.



2.2.Cas d'une variable continue

Pour calculer la médiane dans le cas continu, on doit tout d'abord déterminer la classe médiane qui correspond soit à :

- ✓ La valeur de $n_i \uparrow$ supérieur ou égale à N/2.
- ✓ La valeur de $f_i \uparrow$ supérieur ou égale à 0,5.

On calcule la médiane par cette formule :

$$Me = e_{\min} + \frac{N/2 - n_{iM_{e-1}}\uparrow}{n_{iMe}} \times a_i$$

Dans le cas où nous utilisons les fréquences relatives cumulées croissantes, le calcul de la médiane se fait, par la formule suivante:

$$Me = e_{\min} + \frac{0{,}5 - f_{iM_{e-1}}\uparrow}{f_{iMe}} \times a_i \qquad \text{(Bahouayila, 2016)}$$

tel que ;

Me : La valeur de la médiane.

e_{min} : la borne inférieur de la classe médiane.

$n_{i\,me-1}$ cumulé : l'effectif cumulé croissant de la classe située avant la classe médiane.

n_{ime} : la valeur de l'effectif de la classe médiane.

ai : L'amplitude de la classe médiane.

N : Le total de la population.

3. La Moyenne arithmétique (Notée $\overline{X}$)

La moyenne arithmétique, notée $\overline{X}$ et se lit X barre, d'une variable statistique est la somme des valeurs prises par la variable statistique, divisée par le nombre d'observations.

Cette moyenne est dite simple par opposition à la moyenne pondérée par les effectifs correspondant à chaque valeur de la variable statistique.

3.1.Cas d'une variable discrète :

La moyenne arithmétique simple : la moyenne arithmétique est dite simple lorsque les valeurs prises par la variable X n'apparaissent qu'une seule fois :

$$\overline{X} = \frac{x1 + x2 + x3 + ... + xp}{n1 + n2 + n3 + ... + np} \qquad \overline{X} = \frac{\sum_{i=1}^{p} x_i}{N}$$ (Chakroun, 2018)

Exemple : soit la série suivante qui représente les salaires horaires des ouvriers. Calculer le salaire moyen de chaque ouvrier

8, 10, 12, 14, 15, 16.19.

$$\overline{X} = \frac{8 + 10 + 12 + 14 + 15 + 16 + 19}{6} = \frac{94}{7} = 13,4$$

Le salaire moyen de chaque ouvrier est de 13,4 Da.

La moyenne arithmétique pondérée : la moyenne arithmétique est dite pondérée lorsque les valeurs prises de Xi ont plusieurs valeurs d'effectifs Ni, on la calcule comme suit :

$$\overline{X} = \frac{\sum_{i=1}^{p} n_i \cdot x_i}{\sum n_i}$$ (Chakroun, 2018)

Exemple : Supposons que les 7 ouvriers ont reçu les primes suivantes : 1070 830 830 960 960 960 1070

Répartition des ouvriers selon leur prime

X_i	n_i	$n_i \cdot x_i$
830	2	1660
960	3	2880
1070	2	2140
Total	7	6680

$$\overline{X} = \frac{\sum_{i=1}^{7} n_i \cdot x_i}{N} = \frac{6680}{7} = 954{,}3 \text{ DA}$$

3.2.Cas d'une variable continue : Dans le cas continu où, à chaque classe, e_{i-1}, e_i, de la variable statistique correspond un effectif (ni), nous utilisons la formule de la moyenne pondérée. Seulement, à chaque classe, on associe, pour les calculs, le centre de classe c_i au lieu de xi.

Exemple : La série suivante dans le tableau ci-dessous donne la répartition des salariés d'une entreprise selon leur salaire mensuel

xi	n_i	c_i **(centre de classe)**	$n_i \times c_i$
[4-5[	12	4,5	54
[5-6[	23	5,5	126,5
[6-7[	42	6,5	273
[7-9[	56	8	448
[9-11[	34	10	340
[11-15[	32	13	416
[15-23[	16	19	304
[23-31[	4	27	108
$\sum$	219	-	2069,5

D'où : $\overline{X} = \frac{\sum_i n_i \times c_i}{N} = \frac{2069{,}5}{219} = 9{,}45$

Le salaire mensuel moyen est 9450 DA.

Propriétés de la moyenne :

- Si nous avons deux caractères tel que Y = a X + b, alors = a $\overline{X}$ + b. $\overline{Y}$
- Soit une population de taille N composée de deux sous-population

  P1 : de taille N1 et de moyenne $\overline{X}$ 1.

  P2 : De taille N2 et de moyenne $\overline{X}$ 2.

Donc la moyenne de cette série est : $\overline{X} = \frac{1}{N1+N2}(N_1\ \overline{X}_1 + N_2\ \overline{X}_2)$.

4. Autres types de moyennes

4.1. La moyenne géométrique (Notée G) : elle est utilisée pour calculer les taux de croissance moyen et les variations moyennes. elle est, cependant, recommandée lorsque les données sont présentées en chiffres relatifs, pourcentage ou proportion.

Moyenne géométrique simple : Soit une population composée de N individus dont les valeurs sont $X_1, X_2, \ldots, X_n$. La moyenne géométrique est dite simple lorsque dans le calcul de la moyenne chaque valeur x_i n'intervient qu'une seule fois. Son calcul est donné par la formule suivante :

$$G = \sqrt[N]{X_1 \times X_2 \times X_3 \times \ldots \times X_n} = \sqrt[N]{\prod_{i=1}^{n} X_i} = \left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{N}}, \qquad \text{(monino, 2010)}$$

Remarque: lorsque le nombre des observations est grand, le calcul de la moyenne G sera difficile. Pour cela on fait appel à la fonction logarithmique:

$$\log G = \log\left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{N}} = \frac{1}{N}\log\left(\prod_{i=1}^{n} X_i\right) = \frac{1}{N}\sum_i (\log X_i) = \frac{1}{N}(\log X_1 + \log X_2 + \ldots \log X_n)$$

Exemple : au cours des 4 dernières années, les taux de croissance annuels de la production intérieure brute (PIB) ont été les suivants :

$1^{ère}$ année	$2^{ème}$ année	$3^{ème}$ année	$4^{ème}$ année
$r_1 = +7{,}2\%$	$r_2 = +6{,}3\%$	$r_3 = +7\%$	$r_4 = +4{,}8\%$

Quels est le taux moyen de croissance de la PIB au cours de ces années ?

On a : $X_i = 1 + r_i$

$$G = \sqrt[4]{1{,}072 \times 1{,}063 \times 1{,}07 \times 1{,}048} = \sqrt[4]{1{,}278} = (1{,}278)^{\frac{1}{4}} = 1{,}063$$

$$G = (1.072 \times 1{,}063 \times 1{,}07 \times 1{,}048)^{\frac{1}{4}}$$

$$\log G = \frac{1}{4}(\log 1{,}072 + \log 1{,}063 + \log 1{,}07 + \log 1{,}048) = 0.0266$$

$$G = 10^{0.0266} = 1{,}063$$

Au cours de cette période, le taux moyen de croissance Tx est :

$$Tx = (G-1)\times 100 = (1{,}063-1)\times 100 = +6{,}3\%$$

Moyenne géométrique pondérée

Dans ce cas, la formule de la moyenne géométrique devient :

$$G = \sqrt[\sum n]{X_1^{n_1} \times X_2^{n_2} \times \dots \times X_k^{n_k}}$$

$$\log G = \frac{1}{\sum_i n_i} \sum_i n_i \log X_i = \frac{1}{\sum_i n_i} [n_1 \log X_1 + n_2 \log X_2 \dots + n_k \log X_k]$$

Exemple : Trois équipes se sont succédé à la direction d'une entreprise. Pendant la première période qui a durée quatre ans, les bénéfices réalisés ont augmenté de 50% par an pendant la seconde période, de trois ans, 17% par an et pendant la dernière période, de deux ans, les bénéfices ont enregistré une baisse de 30% par an.

Quel est le taux de croissance annuel moyen des bénéfices réalisés au cours de ces années.

$$G=\sqrt[9]{(1{,}50)^4\times(1{,}17)^3\times(0{,}70)^2}$$

$$\log G=\frac{1}{9}(4\log 1{,}50+3\log 1{,}17+2\log 0{,}70)=0.059$$

D'où : $G=10^{0{,}059}=1{,}145\%$

L'augmentation moyenne annuelle est donc de 14,5%

4.2.La moyenne harmonique (Notée H) : La moyenne harmonique est utilisée quand les unités du caractère sont définit en rapport, c'à d les valeurs d'une variable sont données par le calcul d'une unité constante sur une autre variable (km/h, DA/minute, quintaux /hectare,... etc.).

La moyenne harmonique est l'inverse de la moyenne arithmétique des inverses des valeurs observés.

Moyenne harmonique simple : En considérant N observations $X_1, X_2 \ldots X_n$, les valeurs de la variable, la moyenne harmonique notée **H** est calculée à partir de la formule suivante :

$$H=\frac{N}{\sum_i \frac{1}{X_i}}=\frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\ldots+\frac{1}{X_n}}$$ (Ayache, 2019)

Exemple : une entreprise de transport possède 3camions qui effectuent des rotations entre Bejaia et Bouira. Au cours d'une de celles-ci, le trajet Bejaia – Bouira a été couvert aux vitesses moyennes suivantes : 40 km/h, 60km/h, 80km/h

Quelle est la vitesse moyenne ?

$$H=\frac{3}{\frac{1}{40}+\frac{1}{60}+\frac{1}{80}}=56{,}6km/h$$

Moyenne harmonique pondérée : Dans ce cas, la moyenne harmonique est obtenue à partir de la forme suivante :

$$H = \frac{\sum n_i}{\sum_i \frac{n_i}{X_i}} = \frac{N}{\frac{n_1}{X_1} + \frac{n_2}{X_2} + \cdots + \frac{n_k}{X_k}}$$

Exemple : une entreprise de transport a 10 camions qui font des rotations entre Bejaia et Sétif. Au cours d'une de celles-ci, le trajet a été couvert par ces véhicules aux vitesses moyennes présentées dans le tableau suivant :

Vitesse moyenne (km/h)	40	60	70
Nombres de camions	4	4	2

Quelle est la vitesse moyenne globale ?

$$H = \frac{10}{\frac{4}{40} + \frac{4}{60} + \frac{2}{70}} = 51{,}5 km/h$$

4.3.La moyenne quadratique : Elle est utilisée quand on 'a des valeurs négatives qui donnent une moyenne négative difficile à interpréter, et aussi utilisée quand on 'a des unités en puissance comme par exemple mettre carré (m^2), mettre cube (m^3).

La moyenne quadratique simple

Exemple : Soit la série de chiffres {-4, -2, 0, 2, 4}. Si l'on calcule la moyenne arithmétique simple on obtient zéro.

Parfois, on souhaite obtenir une caractéristique de tendance centrale ayant une valeur positive là où le calcul de la moyenne arithmétique simple aurait donné zéro. On calcule alors **la moyenne quadratique simple** en additionnant le carré de toutes les valeurs de la série et en prenant la racine carrée du total. Autrement dit, dans notre exemple :

$$Q = \sqrt{\frac{(-4)^2 + (-2)^2 + (0)^2 + (2)^2 + (4)^2}{5}} = \sqrt{\frac{40}{5}} = 2{,}83$$

Formule générale de la moyenne quadratique simple : Soient $\{X_1, X_2, ..., X_n\}$ une série de chiffres. La formule de la moyenne quadratique simple de cette série est donnée par : $Q = \sqrt{\frac{1}{N}\sum_{i=1}^{n} X_i^2}$

La moyenne quadratique pondérée : Soient $\{X_1, X_2, ..., X_k\}$ une série de chiffres et $\{n_1, n_2, ..., n_k\}$ les effectifs correspondants. La formule de la moyenne quadratique pondérée de cette série est donnée par : $Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i \cdot X_i^2}$ (Meunier, 2008)

Exemple : soit le tableau suivant :

x_i	25	8	4	12	Total
n_i	10	16	25	20	71
$n_i \times x_i^2$	6250	1024	400	2880	10554

$$Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i \times X_i^2} = \sqrt{\frac{10554}{71}} = 12{,}1921$$

Remarque : les moyennes, qu'elles soient simples ou pondérées sont toujours ordonnées de la même façon : H < G < $\overline{X}$ < Q

Nous pouvons vérifier cette relation en traitant l'exemple suivant

Exemple : un étudiant en sciences commerciales a obtenu à l'examen du rattrapage les notes suivantes :

Statistique descriptive 10/10 coefficient .3

Histoire de la pensée économique = 4/10 coefficient.2

Micro-économie = 3/10 coefficient.1

Calculer les quatres moyennes pondérées ?

Alors nous avons : (x_1 =10, n_1 = 3), (x2 = 4, n2 = 2), (x3 = 3, n3 = 1).

- $\overline{X} = \dfrac{(7\times 3)+(4\times 2)+(3\times 1)}{6} = \dfrac{32}{6} = 5{,}33/10$

- $G = \sqrt[6]{7^3 \times 4^2 \times 3^1} = (343\times 16\times 3)^{\frac{1}{6}} = (16464)^{\frac{1}{6}} = 5{,}01$

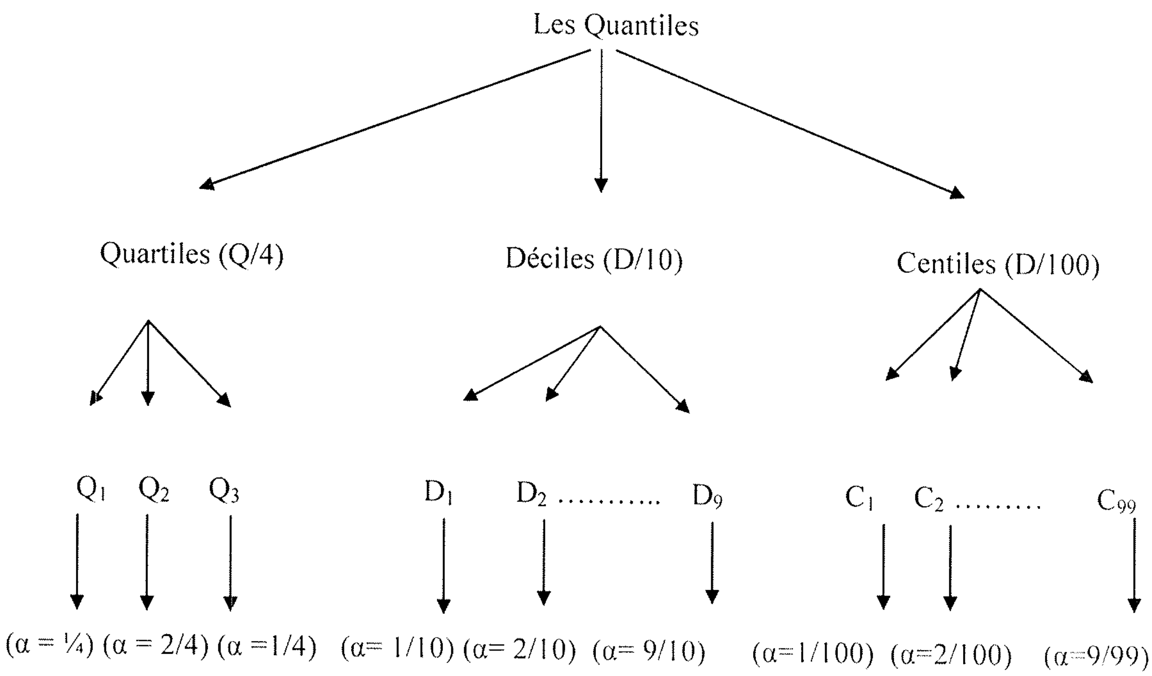
- $H = \dfrac{6}{\frac{3}{7}+\frac{2}{4}+\frac{1}{3}} = \dfrac{6}{1{,}26} = 4{,}76/10$

- $Q = \sqrt{\dfrac{7^2\times 3+4^2\times 2+3^2\times 1}{6}} = \sqrt{\dfrac{147+32+9}{6}} = \sqrt{\dfrac{188}{6}} = 5{,}597/10$

4,76 $<$ 5,01 $<$ 5,33 $<$ 5,59 La relation $H < G < \overline{X} < Q$ est vérifiée

5. Les quantiles

Les quantiles sont déterminés comme la médiane à partir des fréquences absolue (ou relatives) cumulées. Lorsqu'on divise la population totale par quatre, par dix et par cent on obtient respectivement : des Quartiles, des Déciles et des Centiles. (Chauvat, 2002)



5.1.Les Quartiles : Les quartiles (Q_1, Q_2, Q_3) divisent la population de la série en quatre groupes de même effectif.

% des observations ayant des valeurs de xi	Q_1	Q_2	Q_3
- Inférieurs aux quartiles - Supérieures aux quartiles	25% 75%	50% 50%	75% 25%

Le premier quartile Q_1 : 25% des observations sont inférieurs à Q_1.

75% des observations sont supérieurs à Q_1.

Le deuxième quartile Q_2 : 50% des observations sont inférieurs à Q_2 = Me

50% des observations sont supérieurs à Q_2 = Me

Le troisième quartile Q_3 : 75% des observations sont inférieurs à Q_3.

25% des observations sont supérieurs à Q_3.

5.2.Les Déciles : Les déciles, au nombre de neuf (D_1, D_2, D_3, ...,D_9), divisent la population de la série en 10 groupes de même effectif.

% des observations ayant des valeurs de xi	D_1	D_5	D_9
- Inférieurs aux déciles - Supérieures aux déciles	10% 90%	50% 50%	90% 10%

Le premier décile D_1 : 10% des observations sont inférieurs à D_1.

90% des observations sont supérieurs à D1.

Le cinquième décile D_5 : 50% des observations sont inférieurs à D_5 = Me

50% des observations sont supérieurs à D_5 = Me

Le neuvième décile D_9 : 90% des observations sont inférieurs à D_9.

10 % des observations sont supérieurs à D_9.

5.3.Les centiles : Les centiles, au nombre de 100 (C_1, C_2, C_3, ...,C_{99}), divisent la population de la série en 100 groupes de même effectif.

% des observations ayant des valeurs de xi	C_1	C_{50}	C_{99}
- Inférieurs aux centiles	1%	50%	99%
- Supérieures aux centiles	99%	50%	1%

Le premier centile C_1 : 1% des observations sont inférieurs à C_1.

99% des observations sont supérieurs à C_1.

Le 50eme centile C_{50} : 50% des observations sont inférieurs à C_{50} = Me

50% des observations sont supérieurs à C_{50} = Me

Le 99eme centile D_{99} : 99% des observations sont inférieurs à C_{99}.

1% des observations sont supérieurs à C_{99}.

Comment calculer les quantiles ?

Dans le cas discret : On calcule tout d'abord N α, si le résultat de la multiplication de :

- N α appartient aux nombres naturels donc le quantile considéré égale à $\mathbf{X}_{n\alpha}$
- N α n'appartient pas aux nombres naturels donc le quantile considéré égale à $\mathbf{X}_{n\alpha+1}$

Sachant que : - N est le totale de la population.

-α est le paramètre de quantile considéré exemple le paramètre de (Q1, α= ¼), (Q_3, α= ¾), (D_5, α=1/10), (C_{99}, α=99/100), etc.

Dans le cas continu : dans le cas continu, les quantiles se calculent de la même manière que la médiane.

$$Q\alpha = e_i + \frac{N\alpha - N^{\uparrow}_{i\ Q_{\alpha-1}}}{ni_{Q_1}} \times a_i$$

Exercices d'illustration sur le calcule des quantiles

Cas d'une variable discrète

Exemple : soit la série suivante qui représente la répartition des 200 jours de l'hiver selon la température.

- Calculer les quantiles suivants : Q_1, Q_2, Q_3, D_1, D_5, D_9, C_5, C_{50} et C_{75}.
- Commenter Q_1, Q_3, D_5 et C_{75}.

La durée xi	Nombre de logements ni	$n_i \uparrow$	fi	$f_i \uparrow$
2	10	10	0,05	0,05
4	15	25	0,075	0,125
6	25	50	0,125	0,25
8	30	80	0,15	0,4
10	40	120	0,2	0,6
12	30	150	0,15	0,75
14	30	180	0,15	0,9
16	20	200	0,1	1
Total	200	-	1	-

Solution :

- **Q_1**$= q_{1/4} =$? On calcule tout d'abord $N\alpha = 200 \cdot ¼ = 50 \in$ IN, donc $Q_1 = X_{N\alpha}$

$Q_1 = X_{50} = 6°$ (c'est la valeur de X qui correspond à la valeur $\geq n_i \uparrow$ qui est égale à 50).

- **Q_2**$= q_{2/4} = q_{1/2} =$? On calcule $N\alpha = 200 \cdot ½ = 100 \in$ IN, donc $Q_2 = X_{N\alpha}$,

$Q_2 = X_{100} = 10°$ (c'est la valeur de X qui correspond à la valeur $\geq n_i \uparrow$ qui est égale à 10).

- **Q_3**$=$ q3/4= ? On calcule tout d'abord $N\alpha = 200 \cdot 3/4 = 150 \in$ IN, donc $Q3/4 = X_{N\alpha}$

$Q_3 = X_{150} = 12°$ (c'est la valeur de X qui correspond à la valeur $\geq n_i \uparrow$ qui est égale à 150).

- **D_1**$=$ q1/4= d1/10 ? On calcule tout d'abord $N\alpha = 200 \cdot 1/10 = 20 \in$ IN, donc $D_1 = X_{N\alpha}$

$D_1 = X_{20} = 4°$ (c'est la valeur de X qui correspond à la valeur $\geq n_i \uparrow$ qui est égale à 20).

- **D_5**= d5/10= d1/2 = q1/2 = 10° .
- **D_9**= $d_{9/10}$= ? On calcule tout d'abord Nα = 200 9/10= 180 ϵ IN, donc $D_1 = X_{N\alpha}$.

$D_9 = X_{180} = 14°$ (c'est la valeur de X qui correspond à la valeur $\geq n_i \uparrow$ qui est égale à 180).

- **C_5**= $c_{5/100}$= ? On calcule tout d'abord Nα = 200 5/100= 10 ϵ IN, donc $C_5 = X_{N\alpha}$

$C_5 = X_{10} = 2°$ (c'est la valeur de X qui correspond à la valeur $\geq n_i \uparrow$ qui est égale à 10).

- **C50**= $c_{50/100}$= $c_{1/2}$ = d ½ = q ½ = 10°.
- **C 75** = $c_{75/100}$ = c ¾ = q ¾ = 12°.

Commentaire :

<u>Q1 = 6°.</u>

- ✓ 25% des jours (soit 50 jours) de la période d'hiver ont enregistré une température moins que 6°.
- ✓ 75% des jours (soit 150 jours) de la période d'hiver ont enregistré une température plus que 6°.

<u>Q3 = 12° :</u>

- ✓ 75% des jours (soit 150 jours) de la période d'hiver ont enregistré une température moins que 12°.
- ✓ 25% des jours (soit 50 jours) de la période d'hiver ont enregistré une température plus que 12°.

<u>D5 = 10° :</u>

- ✓ 50% des jours (soit 100 jours) de la période d'hiver ont enregistré une température moins que 6°.
- ✓ 50% des jours (soit 100 jours) de la période d'hiver ont enregistré une température plus que 10°.

<u>C75 = 12° :</u>

- ✓ 75% des jours (soit 150 jours) de la période d'hiver ont enregistré une température moins que 12°.
- ✓ 25% des jours (soit 50 jours) de la période d'hiver ont enregistré une température plus que 12°.

Remarque : Nous pouvons calculer les valeurs des quantiles à travers les valeurs de $f_i \uparrow$

Par exemple la valeur de Q1 égale à la valeur de Xi qui correspond à la fréquence cumulée croissante 0,25 (≥ 0,25). Q2 égale à la valeur de Xi qui correspond à la fréquence cumulée croissante 0,50 (≥ 0,5).

Cas d'une variable continue

Exemple : le tableau ci-dessous représente la répartition de 51 branches d'arbre selon leur longueur :

- Calculer Q1, Q3 et D9.

Classes	n_i	$n_i \uparrow$	f_i	$f_i \uparrow$
[2,3 - 4,7[	7	7	0,13	0,13
[4,7 - 7,1[	15	22	0,29	0,42
[7,1 - 9,5[	9	31	0,17	0.59
[9,5 - 11,9[	12	43	0,23	0,82
[11,9- 14,3[	5	48	0,1	0,92
[14,3- 16,7[	2	50	0,04	0.96
[16,7- 19,1[	2	52	0.04	1
Total	52	-	1	-

- **Calcule de Q1, Q3 et D9** :

$$Q_1 = e_{\min} + \frac{\frac{N}{4} - n_i^{\uparrow}{}_{Q_{1-1}}}{ni_{Q_1}} \times a_i \qquad Q_1 = 4{,}7 + \frac{13-7}{15} \times 2{,}4 = 5{,}66cm$$

Commentaire : 25% des branches d'arbres (soit 13 branches) ont une longueur inférieur à 5,66 cm, et le reste soit (39 branches) ont une longueur supérieur à 5.66.

$$Q_3 = e_{\min} + \frac{N \cdot \left(\frac{3}{4}\right) - n_{i\,Q_{3-1}}^{\uparrow}}{ni_{Q3}} \times ai \qquad Q_3 = 9{,}5 + \frac{39-31}{12} \times 2{,}4 = 11{,}09cm$$

Commentaire : 75% des branches d'arbres (soit 39 branches) ont une longueur supérieur à 11,09 cm, et le reste soit (13 branches) ont une longueur supérieur à 11,09 cm.

$$D_9 = e_{\min} + \frac{N \cdot \left(\frac{9}{10}\right) . - n_{i\,D_{9-1}}^{\uparrow}}{ni_{D_9}} \times ai \qquad D_9 = 11{,}9 + \frac{46-43}{5} 2{,}4 = 13{,}34$$

Commentaire : 99% des branches d'arbres (soit 51 branches) ont une longueur inférieur à 13,34 cm, et le reste soit (une branche) ont une longueur supérieur à 13,34 cm.

Conclusion

La réduction des données d'une série statistique se fait par le calcule des paramètres de position (mode, médiane, moyenne, etc). Cette étape est importante, non seulement pour réduire les données, mais aussi pour comparer plusieurs séries et dégager des conclusions raisonnables.

Exercices sur le chapitre III

Exercice n°1 :

Soit le tableau ci-après qui représente le nombre de pièces dans 80 logements de type LPP.

xi	1	2	3	4	5	Total
Ni	26	30	12	8	4	80

1. Quelle est la population statistique étudiée, son caractère et son type ?
2. Déterminé le mode de la série et représenter-le graphiquement.
3. Calculer la médiane et représenter-la graphiquement.
4. Que signifié le résultat de la médiane.
5. Calculer la moyenne arithmétique.
6. Calculer les quantiles suivants : Q_1, Q_3 , D_5, et interpréter les résultats.
7. Déduire les quantiles suivants : C_{75}, Q_2, C_{25}.
8. Quel est le nombre de logements qui ont plus de 3 pièces ?
9. Quel est le nombre de logements qui ont moins de 4 pièces ?

Exercice n°2 :

Soit la distribution suivante qui représente la répartition de 82 exploitations en fonction du nombre d'hectares cultivés.

Nombre d'hectares	[0-22[	[22-44[	[44-66[	[66-88[	[88-110[	Total
Nombre d'exploitation	7	15	30	25	5	82

1. Quelle est la population statistique étudiée, son caractère et son type ?
2. Calculer la valeur du mode et de la médiane et interpréter les résultats.
3. Représenter le mode et la médiane graphiquement.
4. Calculer la moyenne arithmétique.
5. Calculer les quantiles suivants : Q_1. Q_3. D_5. et C_{75} et interpréter les résultats.
6. Quel est le pourcentage d'exploitation qui ont au plus 44 hectares.
7. Quel est le nombre d'exploitation qui ont au moins 66 hectares.

Exercice n°03 :

Soit la série suivante qui représente le nombre de chambres défectueuses dans un cartier de 50 logements après un séisme qui a frappé la région.

Nombre de pièces déf. xi	1	2	3	4	5	6	Total
Nombre de log. ni	6	8	15	7	8	6	50

1. Calculer les quatre moyennes pour cette série statistique.
2. Comparer les résultats.

Chapitre IV : Les paramètres de dispersion et de forme.

Introduction

1. Les paramètres de dispersion
   1.1.L'étendue.
   1.2.Les écarts interquartiles.
      1.2.1. L'intervalle interquartile.
      1.2.2. L'intervalle interdécile.
      1.2.3. L'intervalle intercentile.
   1.3.L'écart absolu moyen.
   1.4.La variance et l'écart type.
      1.4.1. La variance.
      1.4.2. L'écart type.
   1.5.Le coefficient de variation.
   1.6.Les moments simples et les moments centrés.
      1.6.1. Les moments centrés d'ordre K.
      1.6.2. Les moments simples
2. Les paramètres de forme.
   2.1.L'asymétrie.
      2.1.1. Le coefficient de l'asymétrie de Pearson.
      2.1.2. Le coefficient de l'asymétrie de Fisher.
      2.1.3. Le coefficient de l'asymétrie de Yule.
   2.2.L'aplatissement.
      2.2.1. Le coefficient d'aplatissement de Pearson.
      2.2.2. Le coefficient d'aplatissement de Fisher.
      2.2.3. Le coefficient d'aplatissement de Kelley.

Exercices sur le chapitre IV

Chapitre IV : Les paramètres de dispersion et les paramètres de forme

Introduction

La connaissance de la valeur centrale (le mode, la médiane et la moyenne) ne nous donne pas l'information complète sur une série statistique. Il convient donc de donner d'autres mesures pour rendre compte de l'importance que constitue la variabilité ou la dispersion dans une série de donnée. Pour cela, on utilise les paramètres de dispersion et de forme qui permettent d'expliquer la composition interne ainsi que la forme d'une série.

Dans ce présent chapitre, on expose les différentes mesures de dispersion et de forme.

1. Les paramètres de dispersion

Sont des nombres qui mesurent la dispersion des valeurs observées autour d'un paramètre de position (X,Me, Mo, ..). Ils s'expliquent dans la même unité que les observations et permettent de comparer des séries statistiques de même nature.

Nous étudierons ici les caractéristiques de dispersion couramment utilisées dans la pratique, en proposant le classement suivant :

1.1.L'étendue (notée *E*) : Est la différence entre la plus grande et la plus petite des observations faite sur une variable statistique, classées préalablement dans l'ordre croissant ou décroissant. (Fredon, 2007)

$$E = X_{max} - X_{min}$$

Exemple 1 (Cas d'une variable discrète) : Distribution de 500 familles selon le nombre d'enfants

Nombre d'enfants xi	0	1	2	3	4
Nombre de famille ni	50	120	130	110	90

Dans ce cas, l'étendue est égale à :

E= (4-0)=4 enfants par famille.

Exemple 2 (Cas d'une variable continue) : Reprenons l'exemple de la répartition de 52 branches d'arbre selon leur longueur

Classe	n_i
[2,3 - 4,7[	7
[4,7 - 7,1[	15
[7,1 - 9,5[	9
[9,5 - 11,9[	12
[11,9- 14,3[	5
[14,3- 16,7[	2
[16,7- 19,1[	2

Dans ce cas, l'étendue est égale à : $E = 19{,}1 - 2{,}3 = 16{,}8\,\text{cm}$

1.2. Les écarts interquartiles

1.2.1. L'intervalle interquartile $[Q_1 ; Q_3]$: L'intervalle interquartile est une mesure de la variation (dispersion) qui n'est pas influencée par les valeurs extrêmes, contrairement à l'étendue.

L'intervalle interquartile est l'intervalle qui contient 50% des observations, en laissant 25% des observations de part et d'autre de l'intervalle.

L'écart interquartile noté Q est égale à : $Q = Q_3 - Q_1$

1.2.2. L'intervalle interdécile $[D_1 ; D_9]$: Cet intervalle contient 80% des observations en laissant respectivement 10% des observations à droite et à gauche de l'intervalle. Cela signifie que l'on élimine 10% des valeurs se trouvant aux extrémités des distributions.

L'écart interdécile noté D est égale à : $D = D_9 - D_1$

1.2.3. L'intervalle intercentile $[C_1 ; C_{99}]$ **:** Cet intervalle contient 98% des observations en laissant respectivement 1% des observations à droite et à gauche de l'intervalle. (Grandjacquot, 1999)

L'écart interdécile noté C est égale à : $C = C_{99} - C_1$

1.3. Ecart absolu moyen (Noté $e_a(X)$ **) :** L'écart moyen absolu est la moyenne arithmétique des valeurs absolues des écarts entre la valeur prise par la variable statistique X et une valeur type choisie qui est généralement la moyenne $\overline{X}$.

Cas d'une variable discrète : Dans le cas d'une distribution discontinue de N observation, l'écart moyen des valeurs à la moyenne est :

$$e_a(X) = \frac{\sum |X_i - \overline{X}|}{N}$$

Il s'agit d'une moyenne des valeurs absolues des variables centrées.

Exemple1 : La série suivante représente les notes de 6 modules d'un étudiant en première année.

14 15 16 12 11 10

Calculons d'abord la moyenne :

$$\overline{X} = \frac{\sum x_i}{N} = \frac{78}{6} = 13$$

D'où :

$$e_a(X) = \frac{\sum |X_i - \overline{X}|}{N} = \frac{|14-13| + |15-13| + |16-13| + |12-13| + |11-13| + |10-13|}{6} = \frac{12}{6} = 2$$

Dans le cas d'une variable discrète pondérée par les effectifs, la formule est :

$$e_a(X) = \frac{\sum_{i=1}^{n} n_i |X_i - \overline{X}|}{\sum_{i=1}^{n} n_i}$$

Exemple 2 : la série suivante représente le nombre d'absences, d'un échantillon de 10 étudiants, à un cours de statistique.

Nombres d'absences xi	Nombres d'étudiants ni	$n_i \cdot X_i$	$ni \cdot \lvert X_i - \overline{X} \rvert$
1	2	2	8
2	1	2	3
4	1	4	1
6	1	6	1
7	4	28	8
9	1	9	4
Total	10	51	25

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i X_i}{\sum_{i=1}^{n} n_i} = \frac{51}{10} = 5,1 = 5 \qquad e_a(X) = \frac{\sum_{i=1}^{n} n_i \lvert X_i - \overline{X} \rvert}{\sum_{i=1}^{n} n_i} = \frac{25}{10} = 2,5 \approx 3\text{absences}$$

Cas d'une variable continue

- Dans le cas continu l'écart absolu moyen s'exprime par la formule suivante :

$$e_a(X) = \frac{\sum_{i=1}^{n} n_i \lvert c_i - \overline{X} \rvert}{\sum_{i=1}^{n} n_i}$$

Avec c_i : centre de la classe d'une variable continue

Et n_i : effectif de la classe i

Exemple : Soit la distribution suivante qui représente la répartition des salariés selon leur salaire brut mensuel.

e_i (10^3 DA)	n_i	c_i	$\lvert c_i - \overline{X}\rvert$	$n_i \cdot \lvert c_i - \overline{X}\rvert$
[4-5[	12	4,5	4,95	59,40
[5-6[	23	5,5	3,95	90,85
[6-7[	42	6,5	2,95	123,90
[7-9[	56	8	1,45	81,20
[9-11[	34	10	0,55	18.70
[11-15[	32	13	3,55	113,60
[15-23[	16	19	9,55	152,80
[23-31[	4	27	17,55	70,20
Total	219	-	-	710,65

$\overline{X} = 9{,}45 \times 10^3$ DA.

D'où : $e_a(X) = \dfrac{\sum_{i=1}^{n} n_i \lvert X_i - \overline{X}\rvert}{\sum_{i=1}^{n} n_i} = \dfrac{710{,}65}{219} = 3{,}24 \times 10^3$ DA

L'écart absolu moyen et donc 3240 DA.

1.4. La variance et l'écart-type

1.4.1. La variance (Notée V(x)) : On appelle la variance de la variable X, la moyenne des carrés des écarts des valeurs observées par rapport à leur moyenne arithmétique $\overline{X}$. Elle signifie la fluctuation (ou la variation) des valeurs observées autours de leur moyenne.

Ce carré permet de mesurer l'aptitude de la variable X à varier ou à fluctuer autour de sa moyenne. La variance est toujours positive.

On note : $$V(X)=\frac{\sum_{i=1}^{n} n_i(X_i-\overline{X})^2}{\sum_{i=1}^{n} n_i}=\sum f_i(X_i-\overline{X})^2$$ (Ayache, 2019)

Calculer aussi par la formule suivante :

$$V(X)=\frac{\sum_{i=1}^{n} n_i x_i^2}{\sum_{i}^{n} n_i}-\overline{X}^2$$

Remarque : Dans le cas continu on remplace xi par le centre de classe ci

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

1.4.2. L'écart type (Notée $\sigma(X)$) : L'écart type est la racine au carrée de la variance. (Chauvat, 2002)

$$\sigma(X)=\sqrt{v(x)}$$

Exemple (cas d'une variable discrète) : on reprend l'exemple précédent des données représentant le nombre d'absences d'un échantillon de 10 étudiants.

X_i	n_i	n_iX_i	$n_i(X_i-\overline{X})^2$	$n_iX_i^2$
1	2	2	32	2
2	1	2	9	4
4	1	4	1	16
6	1	6	1	36
7	4	28	16	196
9	1	9	16	81

Total	10	51	75	335

$$\overline{X} = \frac{\sum_{i=1}^{n} n_i x_i}{N} = \frac{51}{10} \approx 5$$

$$V(X) = \frac{\sum_{i=1}^{n} n_i X_i^2}{\sum_{i}^{n} n_i} - \overline{X}^2 = \frac{335}{10} - (5{,}1)^2 = 7{,}49 \quad ,$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{7{,}49} = 2{,}73 \approx 3\text{absences.}$$

Exemple 2 : Cas d'une variable continue

Exemple : On reprend le tableau de la répartition des salariés selon leur salaire mensuel brut.

Classe (10^3 DA)	n_i	c_i	$n_i \times c_i$	$n_i \cdot c_i^2$
4-5	12	4,5	54	243
5-6	23	5,5	126.5	695,75
6-7	42	6,5	273	1774,50
7-9	56	8	448	3584
9-11	34	10	340	3400
11-15	32	13	416	5408
15-23	16	19	304	5776
23-31	4	27	108	2916
Total	219	-	2069.5	23797,25

D'où :

$$V(X) = \frac{\sum_{i=1}^{n} n_i c_i^2}{\sum_{i}^{n} n_i} - \overline{X}^2 \frac{23797{,}25}{219} - (9{,}45)^2 = 19{,}3651$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{19{,}3651} = 4{,}40 \times 10^3\,\text{DA ou } 4400\text{DA}$$

1.5. Le coefficient de variation (Notée CV) : nommé aussi l'écart type relatif, il est défini comme le rapport entre l'écart type et la moyenne. Sa valeur est un nombre sans unité de mesure. Il est donné généralement en pourcentage. Le CV est également indépendant des unités choisies, il permet de comparer les séries statistiques exprimées en unités de mesure différentes. (Fredon, 2007)

$$CV = \frac{\sigma(X)}{\overline{X}} \times 100$$

Plus la valeur de CV est grande, plus la distribution est dispersée à la moyenne.

1.6. Les moments simples et les moments centrés

1.6.1. Les moments centrés d'ordre K : Si $X_0 = \overline{X}$, nous parlons de moment centré d'ordre k et nous le notons μ_k (k entier positif quelconque) (Grandjacquot, 1999). Son expression devient alors :

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

Ainsi peut-on établir :

- ✓ $\mu_1(X) = 0$ (Somme des variables centrées nulle)
- ✓ $\mu_2(X) = V(X)$ (Variance).

Le moment centré d'ordre 2 est évidemment le plus important parmi les moments centrés.

1.6.2. Les moments simples d'ordre K : Le principe de calcul est identique au précédent. Seulement, les variables ne sont pas centrées sur la moyenne, mais sur la valeur 0. Ainsi :

Si $X_0 = 0$, nous parlons de **moment simple d'ordre k** et nous le notons m_k

Nous reconnaîtrons les expressions :

$$m_1(X) = \frac{\sum_{i=1}^{n} n_i X_i}{\sum_{i=1}^{n} n_i} = \overline{X}$$: Moyenne arithmétique, moment du 1er ordre.

$$m_2(X)=\frac{\sum_{i=1}^{n} n_i X_i^2}{\sum_{i=1}^{n} n_i}$$: Carré de la moyenne quadratique, moment du 2ème ordre.

La formule générale des moments par rapport à l'origine est donc :

$$m_k(X)=\frac{\sum_{i=1}^{n} n_i X_i^k}{\sum_{i=1}^{n} n_i}$$

Remarque :

Les moments par rapport à l'origine (simples), faciles à calculer, sont utilisés pour le calcul des moments centrés :

$$\mu_2(X)=V(X)=\frac{\sum_{i=1}^{n} n_i X_i^2}{\sum_{i=1}^{n} n_i}-\overline{X}^2=m_2-m_1^2$$

$$\mu_3(X)=m_3-3m_1m_2+2m_1^3$$

$$\mu_4(X)=m_4-4m_1m_3+6m_1^2m_2-3m_1^4$$

$\mu_3(X)$ et $\mu_4(X)$ sont aussi importants que le moment centré d'ordre 2 (variance), puisqu'ils sont utilisés dans le calcul des caractéristiques de forme.

2. Les paramètres de forme

Les différents indicateurs d'asymétrie et d'aplatissement permettent de comparer entre les distributions statistiques. On trouve les paramètres d'asymétries et les paramètres d'aplatissement.

2.1. L'asymétrie : l'asymétrie d'une distribution peut être approchée par une comparaison entre le mode, la médiane et la moyenne arithmétique.

On dit qu'une distribution est symétrique si les trois valeurs de tendance centrale sont égales :

- ✓ Moyenne = Médiane = Mode.

Ainsi que, le premier quartile et le troisième quartile, le premier décile et le neuvième décile, le premier centile et le 99e centile, sont équidistance à la médiane :

- ✓ $Q_3 - Me = Me - Q_1$.
- ✓ $D9 - Me = Me - D_1$.
- ✓ $C_{99} - Me = Me - C_1$.

Les principaux coefficients d'asymétrie sont le coefficient de Fisher, le coefficient de Yule et le coefficient de Pearson. Ils ne sont valables que si les séries sont unimodales et si la variable statistique prend un nombre assez élevé de valeurs.

2.1.1. Le coefficient d'asymétrie de Pearson (Notée β_1) : Pearson propose deux coefficients pour mesurer l'asymétrie (monino, 2010) :

Le premier : $$\beta_1 = \frac{\overline{X} - M_0}{\sigma(X)}$$

Sens de l'asymétrie :

$\beta_1 = 0$, La courbe est symétrique

$\beta_1 \succ 0$, Étalement à droite (M_0 est à gauche de M_e)

$\beta_1 \prec 0$ Étalement à gauche (M_0 est à droite de M_e)

Le deuxième : $$\beta_2 = \frac{\mu_3^2(X)}{\mu_2^3(X)}$$

Il est calculé à partir des moments centrés de la série statistique.

Sens de l'asymétrie :

Si $\beta_2 = 0 \rightarrow$ courbe symétrique

Si $\beta_2 \neq 0 \rightarrow$ courbe asymétrique ; seulement, le sens de l'étalement est donné par le signe de $\mu_3(X)$.

En effet :

Si $\mu_3(X) \succ 0 \rightarrow$ Étalement à droite (M0 est à gauche de la médiane).

Si $\mu_3(X) \prec 0 \rightarrow$ Étalement à gauche (M0 est à droite de la médiane).

2.1.2. Le coefficient d'asymétrie de Fisher : c'est la racine carrée du coefficient de Pearson. (Fredon, 2007)

$$F_1 = \frac{\mu_3(X)}{\sigma^3(X)}$$

Sens de l'asymétrie

Si F1 < 0 , Donc la distribution est asymétrique à droite.

Si F 1 = 0 , Donc la distribution est symétrique.

Si F1 > 0 , Donc la distribution est asymétrique à gauche.

2.1.3. Coefficient d'asymétrie de Yule : Ce coefficient est se défini à partir des quartiles.

$$C_y = \frac{Q_3 - Q_2 - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

Sens d'asymétrie :

$C_y = 0 \rightarrow$ Courbe symétrique

$C_y \succ 0 \rightarrow$ Étalement à droite (M0 est à gauche de la médiane).

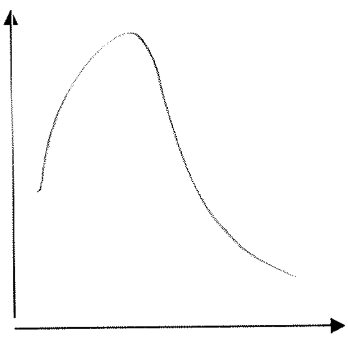
$C_y \prec 0 \rightarrow$ Étalement à gauche (M0 est à droite de la médiane).

Remarque : la dissymétrie (asymétrie) d'une série statistique est le contraire d'une série statistique.

Si nous avons Mode = Médiane = moyenne, la série est symétrique

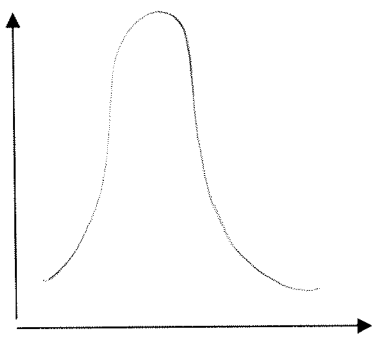
Si nous avons Mode < Médiane < Moyenne, la série est dissymétrique à droite.

Si nous avons le Mode > Médiane > Moyenne, la série est dissymétrique à droite.



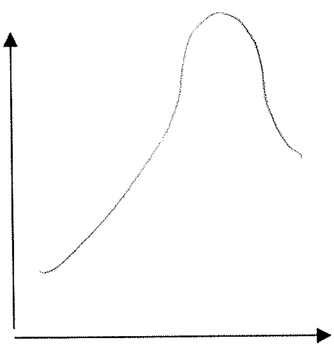
Mo < Me < X

Asymétries à gauche



Mo = Me = X

symétrie



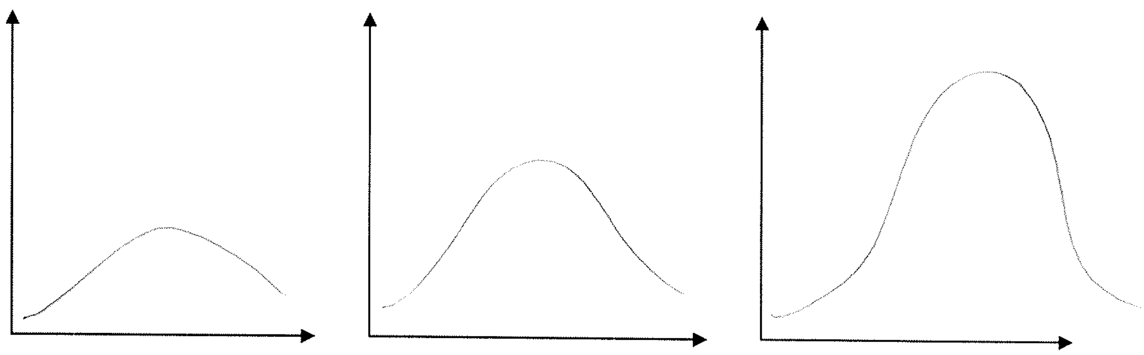
Mo >Me > X

Asymétrie à droite

2.2. L'aplatissement : Comme pour l'asymétrie, on prend comme référence, la courbe normale.

- Une courbe plus aplatie que la courbe normale est dite platicurtique.

- Une courbe moins aplatie que la courbe normale est dite leptocurtique.



Courbe platicurtique **courbe mésocurtique** **courbe leptocurtique**

Les principaux coefficients exprimant le degré d'aplatissement sont les suivants :

2.2.1. Le coefficient d'aplatissement de Pearson (Noté β$_3$)

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)} \qquad \text{(Chauvat, 2002)}$$

Le coefficient β_3 est toujours supérieur à 1, puisque $\mu_4 \succ \mu_2^2$

Les sens d'aplatissement :

Si $\beta_3 = 3 \rightarrow$ courbe normale

Si $\beta_3 \succ 3 \rightarrow$ courbe leptocurtique (moins aplatie que la courbe normale).

Si $\beta_3 \prec 3 \rightarrow$ courbe platicurtique (plus aplatie que la courbe normale).

Remarque : la valeur de coefficient de Pearson ne dépasse pas rarement 5 généralement.

2.2.2. Le coefficient d'aplatissement de Fisher (Noté F_2) : C'est un coefficient sans dimension, invariant par changement de variable et nul pour les distributions symétriques. (Chauvat, 2002) . Sa formule est comme suit

$$F_2 = \beta_3 - 3 = \frac{\mu_4(X)}{\sigma^4(X)} - 3$$

Le sens d'aplatissement

Si $F_2 = 0 \rightarrow$ courbe normale

Si $F_2 \succ 0 \rightarrow$ courbe leptocurtique (moins aplatie que la courbe normale).

Si $F_2 \prec 0 \rightarrow$ courbe platicurtique (plus aplatie que la courbe normale).

Le coefficient de Fisher F_2 est toujours supérieur à -2 du fait de l'intégralité signalée précédemment $\mu_4(X) \succ \sigma^4(X)$.

2.2.3. Le coefficient d'aplatissement de Kelley (Noté C_k) : il existe un autre coefficient d'aplatissement défini à l'aide des quartiles et les déciles. (Hamdani, 2001) :

$$C_k = \frac{1}{2} \times \frac{Q_3 - Q_1}{D_9 - D_1}$$

Exercice : Le tableau ci-après donne la distribution des ventes journalières d'une entreprise pour un certain produit.

Les ventes	0	1	2	3	4	5	**Total**
Effectif ni	98	232	119	85	50	16	**600**

1. Calculer le mode, la médiane et la moyenne de la série statistique.
2. La série est-elle symétrique ?
3. Calculer la variance et l'écart type.
4. Calculer le coefficient de variation.
5. Calculer le coefficient d'asymétrie de Pearson et de Fisher et commenter les résultats.
6. Calculer d'aplatissement de Pearson et de Fisher et commenter les résultats.

Solution :

Les ventes xi	**Effectif ni**	**fi**	**Fi**	**fi xi**	**fi xi²**
0	98	0,1633	0,1633	0	0
1	232	0,3867	0,5500	0,3867	0,3867
2	119	0,1983	0,7483	0,3966	0,3966
3	85	0,1417	0,8900	0,4251	1,2753
4	50	0,0833	0,9733	0,3332	1,3328
5	16	0,0267	1,0000	0,1335	0,6675
Total	**600**	**1**	-	**1,6751**	**4,0589**

1. **Le mode :** Mo = 1 (est la valeur de xi qui correspond à la valeur la plus grande de ni).

La médiane : Me = 1 (corresponds à la valeur supérieur ou égale à 0,5 des fréquences relatives croissantes, soit 0,55 dans cet exercice).

La moyenne : $\overline{X} = \frac{\sum_i n_i \times x_i}{N}$

On utilise les valeurs de fi pour calculer la moyenne, alors on utilise la formule suivante :

$$\overline{X} = \sum fixi = 1{,}675 \cong 2$$

2. **La série est-elle symétrique ?**

Mo = Me $\neq \overline{X}$, Donc la série n'est pas symétrique.

3. **Calcule de la variance et l'écart type :**

La variance : on utilise la formule suivante :

$V(X) = \sum fixi - \overline{X}^2 = 4{,}0589 - (1{,}675)^2 = 1{,}2529$

L'écart type :

$\sigma(X) = \sqrt{v(x)} = 1{,}193$

4. **Calculer le coefficient de variation.**

$$CV = \frac{\sigma(X)}{\overline{X}} \times 100 = \frac{1{,}193}{1{,}675} = 0{,}7122$$

5. **Calculer le coefficient d'asymétrie de Pearson et de Fisher et commenter les résultats.**

- Le coefficient d'asymétrie de Pearson : $\beta_2 = \dfrac{\mu_3^2(X)}{\mu_2^3(X)} = 0{,}49$

- Le coefficient d'asymétrie de Fisher : $F_1 = \dfrac{\mu_3(X)}{\sigma^3(X)} = \mathbf{0{,}69}$

Commentaire :

La valeur du coefficient d'asymétrie de Pearson égale à 0.49, alors la distribution est non symétrique. La valeur de coefficient d'asymétrie de Fisher égale à 0.69 > 0, donc la distribution est asymétrique à gauche.

6. **Calculer le coefficient d'aplatissement de Pearson et de Fisher et commenter les résultats.**

- Le coefficient d'aplatissement de Pearson : $\beta_3 = \dfrac{\mu_4(X)}{\mu_2^2(X)} = \dfrac{\mu_4(X)}{\sigma^4(X)} = 2{,}76.$

- Le coefficient d'aplatissement de Fischer : $F_2 = \beta_3 - 3 = \dfrac{\mu_4(X)}{\sigma^4(X)} - 3 = -0{,}24.$

Commentaire :

La valeur du coefficient d'aplatissement de Pearson est égale à 2,76 < 3 , donc la courbe de distribution est platicurtique, ce qui est confirmé aussi par la valeur de coefficient d'aplatissement de Fisher qui est négative (-0,24).

Exercices sur le chapitre IV

Exercice n°1 :

L'évolution de la production mensuelle de deux enterprises B1 et B2 est présentée dans le tableau suivant;

Mois	Jan	Fev	Mar	Avr	Mai	Jui	Juil	Aou	Sep	Oct	Nov	Dec
B1	250	360	380	400	450	300	150	140	300	350	400	450
B2	40	30	35	40	50	55	20	15	30	40	45	50

1. Calculer L'écart type pour les deux entreprises.
2. Calculer le coefficient de variation.
3. Comparer les résultats.

Exercice n° 02 :

Soit la série suivante qui représente la distribution de 100 colis, selon le poids en Kg.

Poids xi	[10-12[	[12-14[	[14-16[	[16-18[	[18-20[	[20-22[	[22-24[	[24-26[	Total
Effectif ni									100

1. Calculer le mode, la médiane et la moyenne de la série statistique.
2. La série est-elle symétrique ?
3. Calculer la variance et l'écart type de la série.
4. Calculer le coefficient de Pearson et de YULE.

Exercice n° 03 :

Soit la série suivante qui indique le nombre de ménage d'un certain quartier selon le nombre d'enfant.

Nombre d'enfant/ménage (xi)	1	2	3	4	5	6	Total
Effectif (ni)	3	4	7	6	3	2	25

1. Calculer la moyenne d'enfants par ménage.
2. Calculer la variance et l'écart type.
3. Calculer le coefficient d'asymétrie de Pearson et de Fisher.
4. Calculer le coefficient d'aplatissement de Pearson et de Fisher.
5. Commenter les résultats de la question 3 et 4

Conclusion générale

Ce support de cours du module statistique descriptive est un manuel destiné aux étudiants de première année SEGC-LMD. Pour comprendre cette matière et de faciliter ses notions de base, pour les étudiants, nous avons adopté des méthodes simples et faciles selon le programme officiel. Nous avons développé chaque élément avec des exemples et des exercices d'illustrations, cela pour appuyer l'aspect théorique.

Nous avons proposé pour chaque chapitre une série d'exercice, également, nous avons donné un exemple d'un sujet d'examen avec un corrigé-type, qui permet à l'étudiant d'auto contrôler ses acquisitions.

Université Abderrahmane Mira Béjaia
Département SEGC-LMD
Première année SEGC-LMD

EMD Statistique descriptive
(Durée 1h30min)

Exercice n°1: (04 points)

Le tableau ci-dessous résume les résultats des étudiants de première année après la délibération du premier semestre.

Les résultats de semestre I	Admis	Admis avec dettes	Ajournés	Abandon
Effectifs	1100	260	600	40

1. Déterminer la population statistique étudiée et l'unité statistique.
2. Déterminer le caractère étudié, sa nature et ses modalités.
3. Représenter la série graphiquement.
4. Déterminer le mode et interpréter le résultat.

Exercice n°2 : (06 Points)

Le tableau suivant représente la distribution du nombre d'appels reçus durant une journée des 60 salariés d'une entreprise agroalimentaire.

Nombre d'appels	0	1	2	3	4	5	6
Effectif	30	14	8	2	1	2	3

1. Déterminer la population statistique étudiée, le caractère et sa nature.
2. Tracer le diagramme cumulatif des effectifs croissants et décroissants.
3. Calculer la médiane et déterminer sa valeur graphiquement.
4. Calculer le nombre d'appels moyen.
5. Combien de salariés ayant reçu au plus 2 appels ?
6. Combien de salariés ayant reçu au moins 3 appels ?

Exercice n°03 : (10 points)

Une étude réalisée par une étudiante en master sur la répartition de 1000 entreprises industrielles selon le chiffre d'affaire réalisé en milliers d'unités, fait sortir les résultats suivants.

Chiffre d'affaire (en milliers d'unités)	[0 – 40 [	[40 – 60[	[60 - 80[	[80- 100[	[100-120[	[120-150[
Effectifs	354	231	177	81	56	101

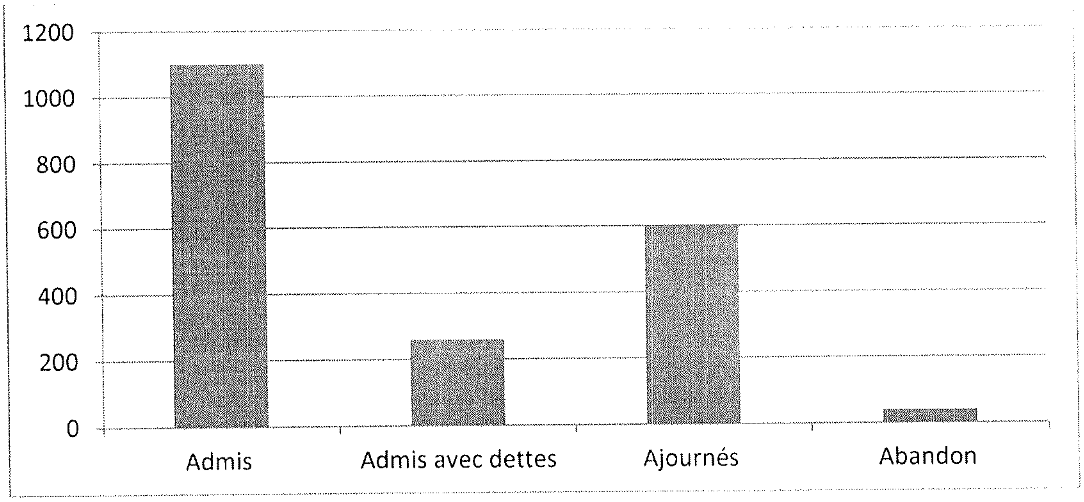
1. Déterminer la population statistique étudiée, le caractère et son type.
2. Calculer le mode de cette série et représenté-le graphiquement.
3. Calculer la médiane et interpréter le résultat.
4. Calculer le chiffre d'affaire moyen et la variance.
5. Quelle est la proportion des entreprises ayant réalisées un chiffre d'affaire supérieur ou égal à 40 et inférieur à 100 milliers d'unités.

Bon courage

Corrigé-type de l'EMD

Corrigé de l'exercice n°1 (04 points)

1. **La population statistique étudiée** : Les étudiants de première année.
   - Unité statistique : Un étudiant.
2. **Caractère étudié** : Les résultats du premier semestre.
   - Sa nature : Qualitative ordinale.
   - Ses modalités : Admis, admis avec dettes, ajournée et abandon.
3. **Représentation graphique de la série statistique :**

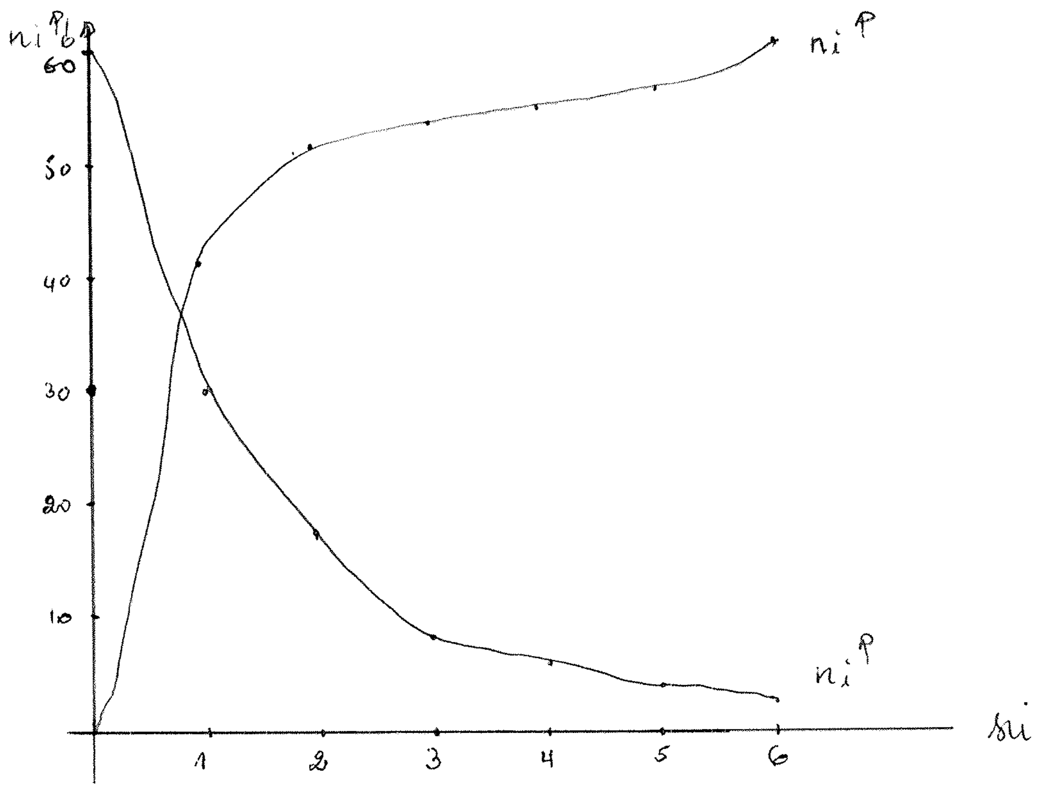


4. **Le mode** : M_0 = Admis. (il correspond à la modalité dont l'effectif est le plus élevé).
   - Interprétation du mode : La valeur du mode montre que la majorité des étudiants de première année sont admis en deuxième année (soit 1100 étudiants).

Corrigé de l'exercice n°2 (06 points)

Nombre d'appels	Effectif ni	$n_i \uparrow$	$n_i \downarrow$	$n_i \times x_i$
0	30	30	60	0
1	14	44	30	14
2	8	52	16	16
3	2	54	8	6
4	1	55	6	4
5	2	57	5	10
6	3	60	3	18
Total	**60**	-	-	**68**

1. **La population statistique étudiée, le caractère et sa nature**
   - La population étudiée : 60 salariés.
   - Le caractère étudié : Nombre d'appels reçu.
   - Nature du caractère : Quantitative discrète.
2. **Le diagramme cumulatif des effectifs croissants et décroissants**



3. **Calcule de la médiane (Me)** : nous avons N= 60 (Pair), donc la médiane est :

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

$$M_e = \frac{x_{30} + x_{31}}{2} = \frac{0+1}{2} = 0{,}5$$

Me = 01 appel.

Présentation graphique de la médiane : Présentée sur le graphe précédent.

4. **Calcule du nombre d'appels moyen**

$$\bar{X} = \frac{\sum ni \times xi}{N} = \frac{68}{60} = 1{,}13 \approx 1\ appel$$

5. **Les salariés ayant reçu au plus 2 appels** : X inférieur ou égal à 2 = (30+14+8) = 52 salariés

6. **Les salariés ayant reçu au moins 3 appels** : X supérieur ou égal à 3 = (2+1+2+3) = 8 salariés.

Corrigé l'exercice n°3 (10 points)

Classes	ni	a_i	n_{ic}	di	$n_i \uparrow$	c_i	$n_i \times c_i$	$n_i \cdot c_i^2$
[0 - 40[	354	40	177	8,85	354	20	7080	141600
[40 - 60[	231	20	231	11,55	585	50	11550	577500
[60 - 80[	177	20	177	8,85	762	70	12390	867300
[80 - 100[	81	20	81	4,05	843	90	7290	656100
[100-120[	56	20	56	2,8	899	110	6160	677600
[120-150[	101	30	67,33	3,37	1000	135	13635	1840725
Total	1000	-	-	-		-	58105	4760825

1) **La population statistique** : Les 1000 entreprises industrielles

Le caractère est : le chiffre d'affaire réalisé

Son type est : quantitatif continu

2) **Le calcule du mode :**

L'effectif le plus grand des n_i corrigés est égal à 231.

D'où la classe modale est : [40 - 60[

$$Mo = X_{\min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i \Rightarrow Mo = 40 + \frac{(231-177)}{(231-177)+(231-177)} \times 20 = 50MU$$

3) **Le calcul de la médiane**

$M_e \in [40 - 60[$

$M_e = X_{min} + \frac{N/2 - n\uparrow_{Me-1}}{ni_{Me}} \times ai$

$M_e = 40 + \frac{500-354}{231} \times 20 = 52{,}64\ MU$

Interprétation de la médiane : Au moins 50% des entreprises ont réalisées un chiffre d'affaire plus que 52,64 MU.

4) **Le calcul du chiffre d'affaire moyen**

$$\bar{X} = \frac{\sum ni \times ci}{N} = \frac{58105}{1000} = 58{,}105\ MU$$

- **Le calcul de la variance**

$$V(X) = \frac{\sum ni \times c_i^2}{N} - \bar{X}^2$$

$$V(X) = \frac{4760825}{1000} - (58{,}105)^2 = 1384{,}633$$

5) **La proportion des entreprises ayant réalisées un chiffre d'affaire supérieur ou égal à 40 million et inférieur à 100 millions** est :

$$P = \left(\frac{231 + 177 + 81}{1000}\right) 100 = \left(\frac{489}{1000}\right) 100 = 48{,}9\%$$

Bibliographie

Ayache, A. (2019). *Statistique descriptive et calcul de probabilités* . Lille : Université de Lille

Bahouayila, B. (2016). *cours de statistiques descriptives* . Gongo: DEUG.

Barthe, R. (1989). *La Statistique Descriptive en 10 Leçons.* Paris : Economica.

Chakroun, a. (2018). *Statistique descriptive et excercices* . Tlemcen: Université Abou Bekr Belkaid .

Chauvat, G. (2002). *Statistiques descriptives.* Paris: Armand Colin.

Fredon, D. (2007). *Statistique et probabilités en 30 fiches* . Paris: Dunod.

Grandjacquot, M.-P. (1999). *Outils statistiques* . Paris: ESKA.

Hamdani, H. (2001). *Statistique Descriptive avec initiation aux méthodes d'analyse de l'information économique* . Alger : Office des Publications Universitaires .

Meunier, J.-M. (2008). *Statistiques descriptives: Résumés et exercices* . France : Saint-Denis .

monino, J.-L. (2010). *Statistiques descriptive.* Paris: Dunod.

Salmon, J. (2014). *Statistique: statistique descriptive* . Paris: TelecomParisTech.