



Université d'Oran 2

Faculté des Sciences Economiques, Commerciales et des Sciences de Gestion

POLYCOPIÉ

Statistique 1

Domaine S.E.G.C (Semestre 1)

Socle commun

Présenté par :

Mme DJENNANE Hayat

Maître de Conférences " A "

Avril 2026

Avant-propos

Ce polycopié s'adresse aux étudiants de première année universitaire (L1, système LMD) dans les disciplines liées à l'économie, la gestion, les finances et les Sciences commerciales. Ils trouveront les éléments de base de la Statistique I (Statistique descriptive) Nous nous sommes attachés à réaliser un polycopié conforme au programme officiel de L1 adopté par le CPND-SEGC (Comité Pédagogique National du Domaine de Formation Sciences Economiques, Gestion et de Sciences Commerciales) arrêté n° 1017 du 11 septembre 2025 modifiant et complétant l'arrêté n° 802 du 07 juillet 2022 fixant le programme d'enseignement pour de la licence dans le domaine S.E.G.C.

Dans ce manuel vous trouverez :

- Le cours précis et complet les résultats à connaître sont mis en valeur.
- Des exemples et des exercices avec solutions détaillées et commentées.
- D'autres exercices sont laissés sans solution. Leurs solutions sont manifestement reportées sur d'autres manuels.
- Des méthodes, regroupées sous la forme d'exemples les plus classiques possibles.

Outre un cours détaillé se rapportant aux méthodes élémentaires de la statistique descriptive (Statistique 1) : classement des données après enquête ; représentation graphiques ; résumés numériques des séries statistiques à une dimension, étude des séries statistiques à deux dimensions : Régression et corrélation, étude des indices statistiques ; ce manuel contient un certains nombre d'exemples et d'exercices avec corrections et commentaires détaillés.

Enfin, je teins à remercier tous ceux qui mon aidé et soutenu lors de la réalisation de ce manuel : ma mère, mon époux, mes enfants, ma fille, son époux et mes petit fils Zayd et Mohmed Racim, mes frères et surtout à ma sœur Nawel et ses enfants et à la mémoire de père Ahmed Djennane et de mon beau-frère Ahmed Khantache qui nous ont quittés très tôt.

Dr DJENNANE Hayat
Maître de Conférences A
FSECG - Université d'Oran 2
Mohamed Ben Ahmed
Avril 2026

SOMMAIRE

Avant-propos	I
Introduction générale	1
Chapitre I : La présentation des statistiques	2
I : Définitions :	2
II: La statistique utile aux économistes	3
III: Les limites de la méthode statistique	4
IV: La terminologie statistique	4
V: Quelques symboles mathématiques utilisés	9
Chapitre II : Les représentations graphiques	11
I : Cas d'une variable qualitative :	11
II : Cas d'une variable quantitative :	13
A- Séries statistiques discrètes	13
B - Séries statistiques continues	17
Chapitre III : Les caractéristiques de tendance centrale ou paramètres de position	23
I : Le Mode :	23
II : La Médiane (Me)	28
III : La moyenne arithmétique ($\bar{X}$)	32
IV : Autres moyennes	38
Chapitre II : Les caractéristiques de dispersion	46
I : L'étendu ou l'intervalle de variation :	47
II : Les écarts inter quartiles	47
III : Mesure de la dispersion autour de la moyenne	54
IV : Variance et écart-type	56
V : Coefficient de variation	59
IV : Coefficient interquartile (relatif)	60
Chapitre V : Les caractéristiques de forme	61
I : Les moments d'une série statistique	61
II : Les coefficients d'asymétrie	62
III : Les coefficients d'aplatissement	65

Chapitre VI : Les caractéristiques de concentration	68
I : La détermination algébrique de la concentration	69
II : La détermination graphique de la concentration	71
Chapitre VII : Corrélation et ajustement linéaire	73
I : Ajustement linéaire d'un nuage de points	73
II : La covariance entre x et y	78
III : Le coefficient de corrélation linéaire entre x et y	79
IV : Le coefficient de détermination entre x et y	84
Chapitre VIII : Les indices statistiques	85
I : Les indices élémentaires	85
II : Les indices synthétiques	87
1- Les indices de Laspeyres	88
2- Les indices de Paasche	89
3- Les indices de Fisher	90
Exercices corrigés	91
Bibliographie	106
Table des matières	110

Introduction générale

La statistique peut être définie comme la méthode d'analyse numérique qui vise à décrire quantitativement des ensembles comportant un grand nombre d'éléments. La statistique est aussi l'instrument qui permet la connaissance d'un phénomène collectif qui échappe, en raison de son étendu ; de sa complexité et de sa diversité, à la méthode uniquement expérimentale.

La statistique est composée de deux grandes parties : statistique descriptive et statistique mathématique.

La statistique mathématique se propose d'étudier les phénomènes observés par des modèles probabilistes qui permettent de calculer la probabilité d'un événement particulier et de procéder à problèmes d'échantillonnage ; d'estimation et de tests.

La statistique descriptive (objet de notre polycopié) se propose de classer les données ; les organiser et de les présenter d'une façon claire afin de tirer des informations chiffrées et utilisables par le calcul de certains paramètres qui expriment les principales caractéristiques de la distribution.

Notre objectif dans ce polycopié est double. Dans les deux premiers chapitres nous exprimons les méthodes qui permettent d'ordonner les observations statistiques en distribution ; de les représenter graphiquement. Dans les chapitres III, IV, V et VI nous présentons les méthodes qui permettent de les résumer. Le septième chapitre aborde les distributions à deux dimensions et également l'ajustement linéaire entre deux variables statistiques (deux phénomènes économiques). Et enfin dans le dernier chapitre nous étudions les indices statistiques.

Ce polycopié comprend donc huit chapitres :

- Le premier est consacré au vocabulaire statistique et à quelques symboles mathématiques utilisés.
- Le second chapitre est consacré aux représentations graphiques.
- Les chapitres III, IV, V et VI sont consacrées à la définition de principaux résumés numériques d'une distribution à une variable (caractéristiques de tendance centrale, de dispersion ; de forme et de concentration).
- Le chapitre VII est consacré à l'étude de la corrélation et de l'ajustement linéaire (régression).
- Le chapitre VIII est dernier chapitre est consacré à l'étude des nombres indices.

CHAPITRE I : LA PRESENTATION DES STATISTIQUES

I. Définitions :

1. Les Statistiques :

Le mot "Statistiques", au pluriel, désigne l'ensemble des données chiffrées qui regroupent toutes les observations faites sur des faits relatifs à un même phénomène qui concerne un groupe d'individus ou d'objets. Des données sont essentiellement tirées des recensements de la population, des déclarations du registre d'état civil ou d'enquêtes appropriées et sont groupées sous forme de tableaux, de graphiques et d'indicateurs statistiques. On trouve des statistiques qui concernent la démographie, l'emploi, la sante, l'industrie, les transports, le commerce intérieur, le commerce extérieur, les indices de prix, la finance...etc. Ainsi, les statistiques ce sont des données chiffrées relatives à un phénomène étudié.

2. La Statistique :

Le mot "statistique", au singulier, désigne la discipline scientifique constituée par l'ensemble des procédés, des techniques ou des méthodes d'analyse visant, d'une part, à fournir, au moyen d'un nombre limité de caractéristiques, une description simple et la plus complète possible d'une population envisagée sous l'angle d'un caractère donné. D'autre part, la statistique permet d'interpréter les caractéristiques aussi déterminées afin de tirer des conclusions concernant la population étudiée et de prendre des décisions.

La Statistique, en tant que méthode d'analyse des données quantitatives et qualitatives, comporte deux niveaux :

- **La statistique descriptive**, qui consiste en la collecte et la présentation de données, ainsi que leur première analyse. Le but est de représenter d'une manière compréhensible et utilisable l'information fournie par les données. La statistique descriptive est la partie de la statistique qui sert à décrire un phénomène, c'est-à-dire de mesurer, classer les mesures, présenter ces mesures par quelques indicateurs de manière à donner une idée simple et rapide d'un phénomène étudié.
- **La statistique inférentielle** ou **statistique mathématique**, qui cherche à trouver les caractéristiques de la population mère a partir des observations faites sur un échantillon. Elle prend la suite de la statistique descriptive et fait appel au calcul des probabilités.

Une opération statistique se déroule en général en quatre étapes :

- **La collecte des données** : cette étape peut se réaliser de deux manières :
 - Par recensement, auquel cas l'information porte sur la totalité des individus qui forment la population en question. (exemple : le recensement de la population et de l'habitat effectué par l'Office National de la Statistique, une fois tous les dix ans).
 - Par sondage, auquel cas on se limite à une partie de la population qu'on appelle échantillon. Un échantillon est un sous ensemble de la population totale. Il doit être représentatif, c'est-à-dire doit être choisi de telle sorte qu'il ait la même structure et les mêmes propriétés fondamentales de l'ensemble dont il est issu (population mère).
- **La représentation et l'organisation statistique des données** : cette étape peut se réaliser, soit à l'aide de graphiques soit à l'aide d'indicateurs statistiques.
- **La modélisation** : On distingue deux types de modèles, les modèles explicatifs et les modèles prévisionnels.
- **L'interprétation des résultats.**

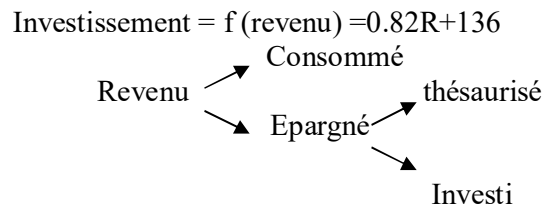
On appelle statistique la méthode scientifique qui vise à observer, collecter, analyser des données quantitatives.

II. La statistique utile aux économistes :

La statistique est un outil indispensable aux économistes : tant aux théoriciens qu'aux praticiens de l'économie :

1. La statistique est utile aux théoriciens de l'économie :

- Elle permet de mettre en évidence (révéler) l'existence d'interdépendance entre différents phénomènes économiques.
- Elle permet de tester la validité d'une hypothèse théorique.



2. La statistique est utile aux praticiens de l'économie :

- La statistique permet aux entrepreneurs de mieux contrôler la gestion de leurs entreprises.
- Elle permet également au pouvoir public de mieux définir les politiques économiques, fiscales, monétaires et d'emploi.

III. Les limites de la méthode statistique :

Pour éviter des erreurs d'interprétation due à une mauvaise utilisation statistique, il faut savoir :

- La statistique s'intéresse au grand nombre, elle ignore les cas particuliers.
- La résultante d'un grand nombre d'informations peut être différente de la sommation de ces différentes informations : comportement collectif n'est pas égal à la sommation des comportements individuels

Quand on étudie un phénomène on n'est jamais certain que l'on dispose de toutes les informations le concernant.

Il ne faut pas oublier que la statistique n'est qu'un outil au service de l'économiste, ce qui nous oblige de ne jamais oublier après une interprétation statistique de faire une analyse économique des résultats :

- Les mêmes causes ne produisent pas les mêmes effets.
- Les corrélations mêmes très parfaites ne signifient pas toujours qu'il y a interdépendance entre les phénomènes étudiés.

IV. La terminologie statistique (Vocabulaire de base) :

1. Population :

C'est l'ensemble de référence, c'est-à-dire l'ensemble des unités observées, qui constitue l'objet de l'étude de la statistique que l'on cherche à connaître. (La population des étudiants de l'université d'Oran 2 Mohamed Ben Ahmed, la population des salariés d'une entreprise industrielle...). Autrement dit, c'est l'ensemble sur lequel porte l'étude. Une population ne signifie pas exclusivement un ensemble de personnes physiques, mais peut concerner des personnes morales ou des objets (entreprises, exploitations agricoles, université, ampoules, voitures). Une population doit être bien définie. Sa définition est importante car elle conditionne l'homogénéité des unités observées et aussi la fiabilité des résultats.

Exemple : Âge des étudiants inscrits en 1^{ère} année Sciences économiques : l'ensemble étudié c'est l'âge.

2. Unité Statistique ou Individu :

Une Unité Statistique ou Individu est tout élément de la population ou de l'échantillon. Une population se compose d'éléments chaque élément est appelé unité statistique. La totalité des individus correspond à la population.

Exemple : La population d'étudiants : l'Unité statistique est un étudiant.

3. Caractère et modalité :

Pour chaque individu extrait d'une population ou d'un échantillon, on relève de la valeur d'une ou plusieurs de ses caractéristiques. Le caractère ou variable statistique est un aspect particulier de l'individu que l'on désire étudier.

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

3.1. Caractère qualitatif ou variable qualitative :

Un caractère est dit **qualitatif** lorsqu'il est lié à une observation qui n'est pas mesurable. **Exemple :** lors de l'étude de la population algérienne on peut s'intéresser à quelques-unes de ses caractéristiques telles que :

- Le milieu de résidence (urbain, rural) ;
- Le sexe (masculin, féminin) ;
- La région de résidence (nord, sud, centre, est, ouest) ;
- L'état matrimonial (marié, veuf, divorcé, célibataire).

Les **modalités** d'un caractère sont simplement les différentes rubriques d'une nomenclature définie a priori et associées a un caractère qualitatif. Une modalité est donc une des réponses possibles à un caractère. Autrement dit, on appelle une modalité les différentes situations possibles que peut prendre un caractère.

Par exemple :

- le caractère milieu de résidence comporte deux modalités : milieu rural et milieu urbain.
- le caractère « sexe » : 02 modalités possibles : Masculin/Féminin
- le caractère « état matrimonial » : 4 modalités possibles : célibataire/marié/divorcé/veuf.

Ces modalités doivent former une partition, c'est-à-dire doivent être exhaustives et disjointes : pour chaque individu on doit pouvoir lui associer une modalité et une seule.

La **Nomenclature** désigne l'ensemble des modalités d'un caractère précédées d'un numéro.

Présentation du tableau statistique associée à un caractère qualitatif

Modalités (numérotées)	Effectifs	Fréquences
M_i	n_i	f_i
M_1 (1)	n_1	f_1
M_2 (2)	n_2	f_2
...	...	...
...	...	...
...	...	...
M_r (r)	n_r	f_r
Ensemble	N	1

L'effectif total **N** est le nombre total d'individus observés

$$N = n_1 + n_2 + \dots + n_i = \sum_{i=1}^r n_i$$

L'effectif **n_i** d'une modalité, appelé aussi **fréquence absolue**, est le nombre de fois ou la modalité numéro i a été observée.

La fréquence relative **f_i** d'une modalité est le rapport de l'effectif **n_i** à l'effectif total **N** :

$$f_i = \frac{n_i}{N} = \frac{n_i}{\sum_{i=1}^r n_i}$$

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

Remarque :

- Les fréquences relatives peuvent être exprimées en pourcentage.
- Lorsque les modalités ne permettent pas l'exhaustivité, c'est-à-dire lorsqu'il y a des individus qu'on ne peut classer dans le tableau, on peut rajouter une modalité, en bas du tableau, qu'on appelle "autres" ou "non réponses".

3.2. Caractère quantitatif ou variable quantitative :

Lorsque les observations relatives a un caractère sont mesurables, le caractère est dit **quantitatif** (taille, âge, poids, production, superficie....). A chaque modalité correspond un nombre différent. **Exemple :** lors de l'étude de la population algérienne, on peut s'intéresser à quelques-unes de ses caractéristique telles que :

- Le nombre d'enfants par ménage, le nombre d'années d'études, nombre de voitures par ménage...).
- L'âge, le poids, la taille, le revenu, la facture de l'électricité et du gaz, l'impôt sur le revenu, les dépenses en loyer...).

On distingue deux types de caractères quantitatifs :

- Les **caractères quantitatifs discrets**, si les valeurs possibles de la variables sont des nombres isolés (en général des nombres entiers comme par exemple le nombre d'enfants d'un ménage, le nombre de voitures par ménages...).
- Les **caractères quantitatifs continus**, si les valeurs possibles de la variable sont a priori en nombre infini dans un intervalle de valeurs (comme par exemple la taille, le revenu, l'âge...).

Remarque : Certains caractères discrets sont de préférence traités en tant que caractères continus. Exemple nombre d'ouvriers dans chaque entreprise, nombre de places de cinémas associées a chaque salle...).

Présentation du tableau statistique associée à un caractère quantitatif discret

Valeurs observées x_i	Effectifs n_i	Fréquences f_i	Fréquences cumulées $F_i \uparrow$
x_1	n_1	f_1	$F_1 = 0$
x_2	n_2	f_2	$F_2 = f_1$
...	...	...	$F_3 = f_1 + f_2$
...	...	...	...
...	...	...	...
x_p	n_p	f_p	F_p
Ensemble	N	1	

Fréquences cumulées croissantes F_i : Le cumul des fréquences associées aux valeurs du caractère inférieures strictement a la valeur x_i .

$$F_i = \sum_{j=1}^{i-1} f_j \text{ pour } i = 2, 3, \dots, p. \text{ et } F_1 = 0$$

Présentation du tableau statistique associe à un caractère quantitatif continu

Classes $[b_{i-1} - b_i[$	Centres de classes C_i	Effectifs n_i	Fréquences f_i	Fréquence cumulée $F_i \uparrow$
$[b_0 - b_1[$	C_1	n_1	f_1	$F_1 = f_1$
$[b_1 - b_2[$	C_2	n_2	f_2	$F_2 = f_1 + f_2$
$[b_2 - b_3[$	C_3	n_3		$F_3 = f_1 + f_2 + f_3$
...	...	...	...	...
...	...	...	...	...
$[b_{p-1} - b_p[$	C_p	n_p	f_p	F_p
Ensemble		N	1	

Remarque :

- Par convention, les classes sont fermées à gauche et ouvertes à droite. Une classe est dite bornée si : $b_{i-1} \neq -\infty$, $b_i \neq +\infty$
- Le centre d'une classe bornée est : $C_i = \frac{b_{i-1} + b_i}{2}$, $C_i \approx x_i$
- L'amplitude d'une classe bornée est : $a_i = b_i - b_{i-1}$

4. Effectifs (fréquences absolues) :

C'est le nombre d'unités statistiques relatif à une modalité donnée :

Age	Effectifs
[18-20[	20
[20-22[	35
[22-24[	5
total	60

5. Fréquences relatives :

C'est le rapport de la fréquence absolue à l'effectif total :

Exemple : $20/60 = 33/100 = 0.333$ est la fréquence relative de première modalité.

6. Série statistique :

Distribution de fréquences, distribution de statistiques ou tableau statistique, c'est un tableau qui nous donne ou résume l'ensemble des valeurs mesurant le caractère.

Exemples :

Séries simples :

Sexe	Effectifs n_i
Masculin	180
Féminin	240
Total	60

Nombre d'enfants x_i	Nombre de ménages n_i
2	18
3	28
4	10
5	4
Total	60

Séries avec classes :

Salaires 10^3 DA	Nombre de ménages n_i
[40 – 60 [	10
[60 – 70 [	25
[70 – 80 [	05
Total	40

7. Classe :

On appelle classe un groupement de valeurs du caractère selon des intervalles qui peuvent être égaux ou inégaux.

Pour chaque classe on peut définir :

- Une limite inférieure
- Une limite supérieure
- Intervalle de classe (amplitude) = limite (sup)- limite (inf)
- Centre de classe = [limite (sup) + limite (inf)]/2

NB : « [40-60[» signifie qu'on comptabilise les salariés qui gagnent entre 40000 et 60000 DA, en incluant ceux qui gagnent 40000 DA et excluant ceux qui gagnent 60000 DA.

V. Quelques symboles mathématiques utilisés :

1. L'opérateur Somme de : Sigma Σ

L'opérateur Σ (lettre grecque sigma majuscule) permet d'écrire de manière contractée la somme d'une variable indicée entre deux bornes.

On peut par exemple écrire :

$$x_1 + x_2 + x_3 + x_4 + x_5 = \sum_{i=1}^5 x_i$$

D'une manière générale :

$$x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i$$

Cette formule se lit de bas en haut : **somme de i égal 1 à i égal n de x indice i.**

Remarque : On peut établir les résultats suivants :

$$\begin{aligned} \sum_{i=1}^n x_i + \sum_{i=1}^n y_i &= \sum_{i=1}^n (x_i + y_i) \\ \sum_{i=1}^n a &= na \\ \sum_{i=1}^n (x_i + a) &= \sum_{i=1}^n x_i + na \\ \sum_{i=1}^n ax_i &= a \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i &\neq \sum_{i=1}^n x_i \sum_{i=1}^n y_i \\ \sum_{i=1}^n \left(\frac{x_i}{y_i} \right) &\neq \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} \\ \left(\sum_{i=1}^n x_i \right)^2 &\neq \sum_{i=1}^n (x_i)^2 \\ \sum_{i=1}^n (x_i + y_i)^2 &= \sum_{i=1}^n (x_i)^2 + \sum_{i=1}^n (y_i)^2 + 2 \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n \sum_{j=m}^{j=1} x_i y_j &= \sum_{j=1}^{j=m} \sum_{i=1}^n x_i y_j = \sum_{i=1}^n x_i \sum_{j=1}^{j=m} y_j = \sum_{j=1}^{j=m} y_j \sum_{i=1}^n x_i \end{aligned}$$

2. L'opération de produit : ($\prod$)

– Notation : Le produit de x variables s'écrit :

$$x_1 \cdot x_2 \cdot x_3 \dots x_n = \prod_{i=1}^n x_i$$

➤ Propriétés

$$\prod_{i=1}^n a = a^n \qquad \prod_{i=1}^n ax_i = a^n \prod_{i=1}^n x_i$$

CHAPITRE II : LES REPRESENTATIONS GRAPHIQUES

Les tableaux statistiques donnent un premier résumé statistique des résultats d'une enquête. Cependant, dans le cas où la variable présente plusieurs modalités ou dans le cas où nous avons à comparer deux ou plusieurs distributions, il est préférable de représenter les résultats à l'aide de graphiques.

L'intérêt d'un graphique c'est de synthétiser des informations statistiques d'une manière imagée, c'est à dire globale.

I. Cas d'une variable qualitative :

Il y a plusieurs manières de représenter graphiquement une variable qualitative. Le choix du type de la représentation graphique dépend des différentes modalités du caractère. On distingue essentiellement le **diagramme circulaire**, appelé aussi **diagramme à secteurs** et le **diagramme en tuyaux d'orgues**, appelé aussi **diagramme à bandes** ou encore **diagramme en barres**.

Exemple : Le tableau suivant représente la répartition de la population active occupée dans l'agriculture par statut dans la profession est la suivante :

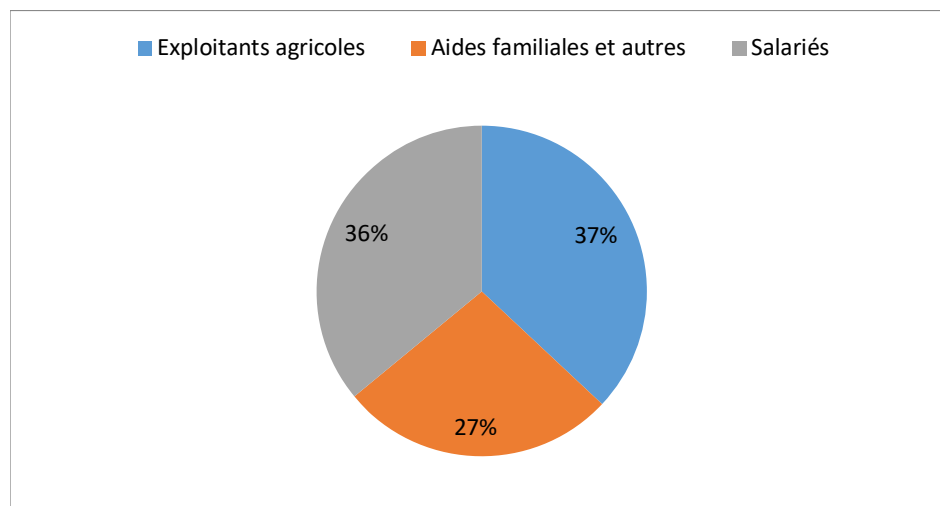
Statut dans la profession M_i	Effectifs n_i	Fréquences f_i	Angles α_i
Exploitants agricoles	190.5	0.374	135°
Salariés	182.2	0.357	129°
Aides familiales et autres	137	0.269	97°
Ensemble	509.7	1	360°

On peut remarquer que ceux qui travaillent dans l'agriculture en tant que salariés, représentent un peu plus du tiers de la population active agricole et que l'agriculture dans cet exemple est caractérisée par le travail familial.

1. Diagramme en secteurs (circulaire) :

Le principe de la représentation du **diagramme en secteurs** est le suivant : effectif total représenté par un disque, modalité représentée par un secteur circulaire dont la surface est proportionnelle à la fréquence, angle de chaque secteur est égale à :

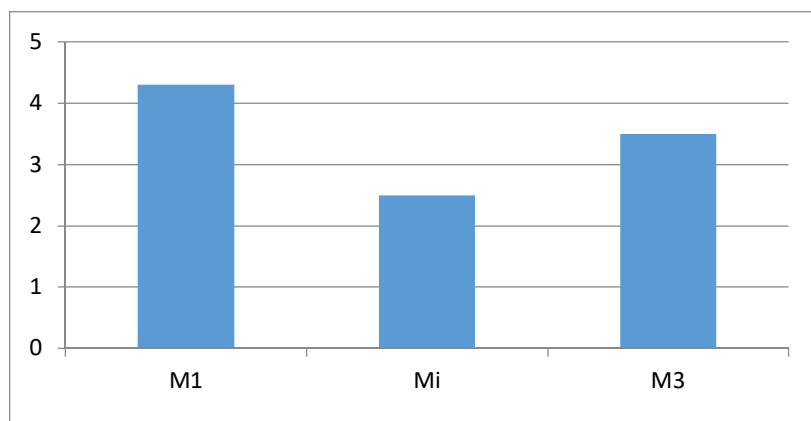
$$\alpha_i^\circ = 360^\circ \cdot f_i$$



Graphique 2-1

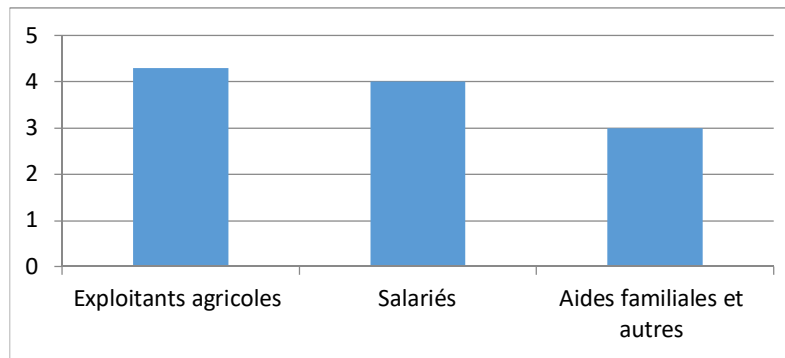
2. Diagramme en tuyaux d'orgue :

Le principe de la représentation du diagramme à bandes ou en tuyaux d'orgues est le suivant : Association à chacune des modalités M_i du caractère, qui sont placées sur un axe horizontal, une bande verticale ayant une hauteur proportionnelle à la fréquence f_i (ou à l'effectif n_i). Les bases des bandes doivent être égales et équidistantes.



Graphique 2-2

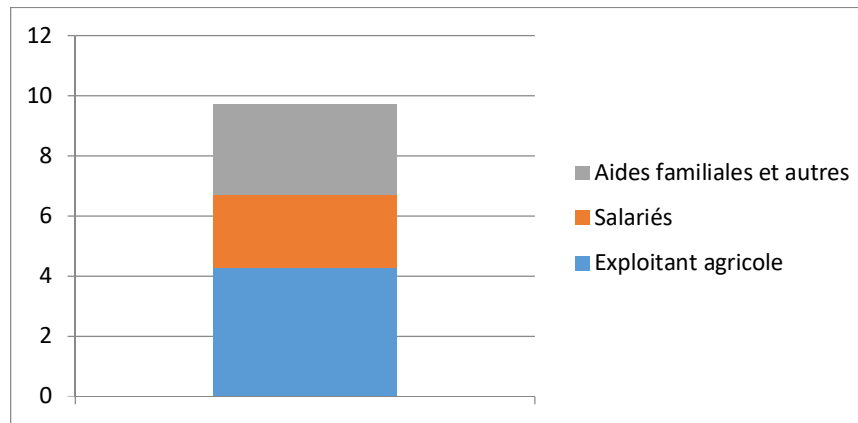
Concernant l'exemple de la distribution de la population active agricole : la représentation par **un diagramme en tuyaux d'orgue** est la suivante :



Graphique 2-3

3. Diagramme rectiligne :

On peut aussi utiliser une seule bande dont la hauteur est égale à 100% et la subdiviser selon le nombre de modalités. Chaque sous-bande est proportionnelle à la fréquence f_i .



Graphique 2-4

II. Cas d'une variable quantitative :

On peut envisager plusieurs types de représentations graphiques mais qui sont spécifiques à chaque type de caractère statistique étudié.

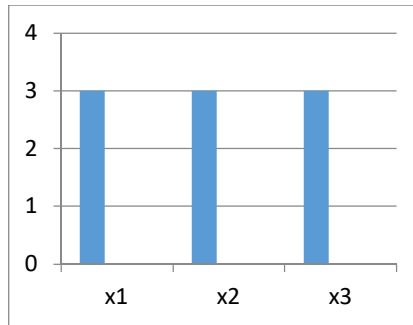
A. Séries statistiques discrètes :

1. Diagramme en bâtons :

Ce diagramme s'applique uniquement aux caractères quantitatifs discrets.

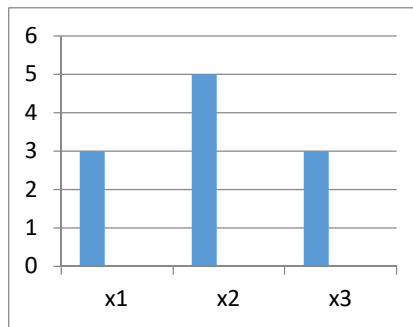
Il s'agit de la figure obtenue sur un repère cartésien en associant à chaque point de coordonnées $(x_i, 0)$ un segment vertical dont la longueur est proportionnelle à la fréquence f_i (ou à l'effectif n_i).

Uniforme



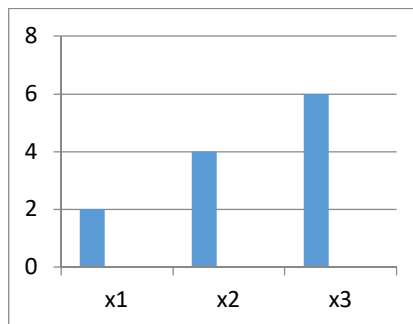
Graphique 2-5

Symétrique



Graphique 2-6

Asymétrique



Graphique 2-7

L'intérêt de cette représentation est double. D'une part, elle permet de donner une idée générale sur la forme de la distribution. D'autre part, elle permet de repérer les valeurs aberrantes.

Exemple : La distribution d'un échantillon d'individus selon le nombre de personnes par ménage est résumée dans le tableau suivant :

Nombre d'enfants x_i	Nombre de ménages n_i
1	25
2	42
3	38
4	15
5	6
6	-
Total	126

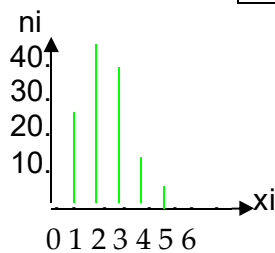


Diagramme en bâtons

Graphique 2-8

2. Fréquences cumulées croissantes, fonction de répartition et diagramme cumulatif :

On appelle fonction de répartition d'une variable statistique quantitative toute application définie par :

$$F: \mathfrak{R} \rightarrow [0, 1]$$

$$x_i \mapsto F(x_i) = \text{prop} (X < x_i)$$

$F(x_i)$ est égale à la proportion des individus ayant une valeur du caractère strictement inférieure à x_i .

Exemple : On donne la répartition de 69 ménages selon le nombre d'enfants par ménage :

Valeurs x_i	Effectifs n_i	Fréquence f_i	Fréquences cumulées croissantes $F_i \nearrow$
1	22	0.32	0
2	11	0.16	0.32
3	16	0.23	0.48
4+	20	0.29	0.71
Ensemble	69	1	1

$$F(1) = \text{prop}(x < 1) = 0$$

$$F(1,5) = \text{prop}(x < 1,5) = \text{prop}(x = 1) = 0,32$$

$$F(2) = \text{prop}(x < 2) = \text{prop}(x = 1) = 0,32$$

$$F(2,5) = \text{prop}(x < 2,5) = \text{prop}(x = 1) + \text{prop}(x = 2) = 0,48$$

$$F(3) = \text{prop}(x < 3) = \text{prop}(x = 1) + \text{prop}(x = 2) = 0,48$$

$$F(4+) = \text{prop}(x < 4+) = \text{prop}(x = 1) + \text{prop}(x = 2) + \text{prop}(x = 3) = 0,71$$

Ainsi, la formulation de la fonction de répartition de cette distribution statistique est comme suit :

$$F = \begin{cases} 0 & \text{si } x \leq 1 \\ 0.32 & \text{si } 1 < x \leq 2 \\ 0.48 & \text{si } 2 < x \leq 3 \\ 0.71 & \text{si } 3 < x \leq 4+ \\ 1 & \text{si } x > 4+ \end{cases}$$

La représentation graphique de la fonction de la répartition, appelée diagramme cumulatif ou diagramme intégral.

Ce diagramme permet de visualiser l'évolution des fréquences cumulées liées aux valeurs de la variable. Le caractère étant discret, la courbe des fréquences cumulées croissante est la représentation graphique d'une fonction en escalier.

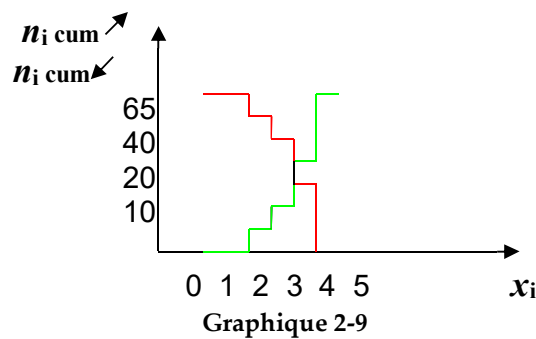
D'une manière générale, la fonction de répartition est constante par intervalle. Sa formulation est la suivante :

$$F = \begin{cases} 0 & x \leq x_1 \\ f_1 & x_1 < x \leq x_2 \\ f_1 + f_2 & x_2 < x \leq x_3 \\ f_1 + f_2 + \dots + f_{p-1} & x_{p-1} < x \leq x_p \\ 1 & \text{si } x > x_p \end{cases}$$

L'intérêt de la représentation graphique est qu'elle permet de retrouver pour toute valeur de x_i donnée, la proportion des individus ayant une valeur de la variable strictement inférieure à x_i .

Exemple : On donne la répartition de 65 ménages selon le nombre d'enfants par ménage :

Nombre d'enfants x_i	Nombre de ménages n_i	n_i cum ↗	n_i cum ↘
1	5	5	65
2	10	15	60
3	30	45	50
4	20	65	20
Ensemble	65	$\leq x_i$	$\geq x_i$



B. Séries statistiques continue :

1. Histogramme :

Graphiquement, on représente une série statistique continue par un **histogramme**. Il s'agit d'une figure obtenue sur un repère cartésien en représentant pour chaque classe $[b_{i-1}, b_i]$ un rectangle de surface S_i proportionnelle à l'effectif n_i ou à la fréquence f_i . Les rectangles de l'histogramme sont contigus.

Principe de construction de l'histogramme :

$$S_i = \text{base} \times \text{hauteur} = a_i \times h_i = n_i \times a^* \text{ d'où : } h_i = \frac{n_i}{a_i} \times a^* = d_i \times a^*$$

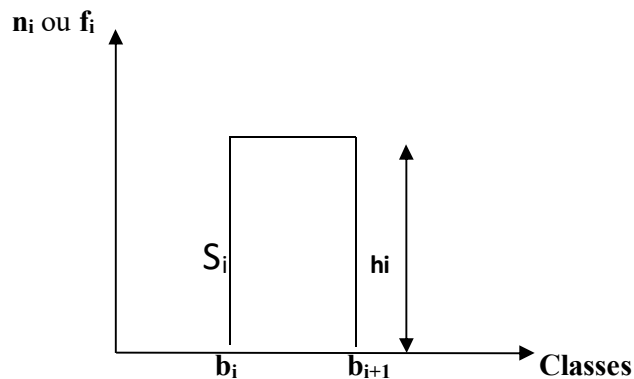
Le a^* est appelée amplitude de référence. Elle est choisie arbitrairement de manière à faciliter la représentation graphique (valeurs sur l'axe des ordonnées).

La hauteur h_i est dans ce cas appelée effectif corrigé qu'on note par n_i^c .

La densité d_i d'une classe est : $d_i = \frac{n_i}{a_i}$. Il s'agit du nombre d'individus par unité d'amplitude.

Remarque :

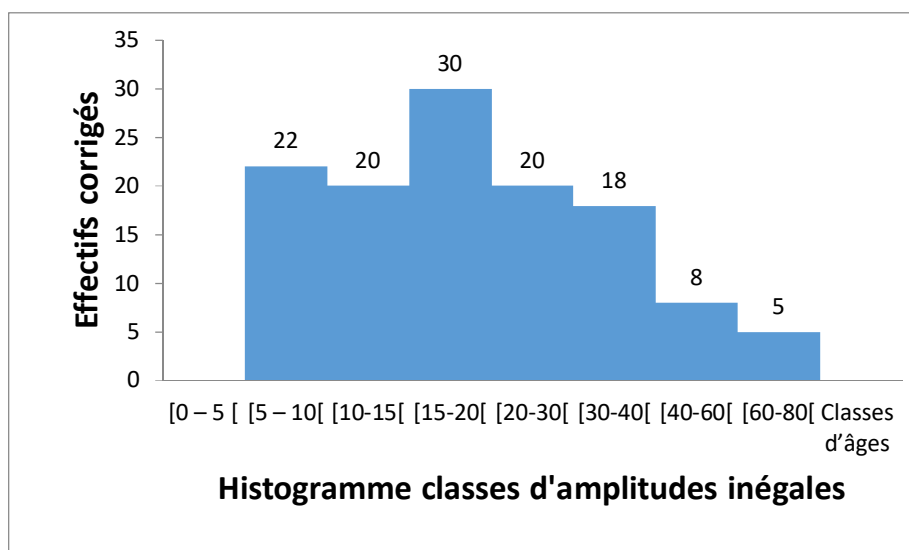
- On peut utiliser les fréquences corrigées à la place des effectifs corrigés.
- Dans le cas de classes d'**amplitudes égales**, il n'est pas nécessaire de calculer les fréquences corrigées ou les effectifs corrigés. On peut utiliser directement les effectifs ou les fréquences comme hauteurs des rectangles. En revanche, dans le cas de classes d'**amplitudes inégales**, les hauteurs des rectangles doivent être proportionnelles à la **densité**, afin d'avoir une **surface** proportionnelle à l'effectif.



Graphique 2-10

Exemple : La répartition de 100 individus par classes d'âges est donnée par le tableau suivant :

Classes d'âges	Effectifs n_i	Amplitudes a_i	Densité d_i	Effectifs corrigés n_i^c	Fréquences f_i	Fréquences corrigées f_i^c
[5 – 10 [	11	5	2.2	22	0.11	0.22
[10-15[	10	5	2	20	0.10	0.20
[15-20[	15	5	3	30	0.15	0.30
[20-30[	20	10	2	20	0.20	0.20
[30-40[	18	10	1.8	18	0.18	0.18
[40-60[	16	20	0.8	8	0.16	0.8
[60-80[	10	20	0.5	5	0.10	0.5
Total	100				1	

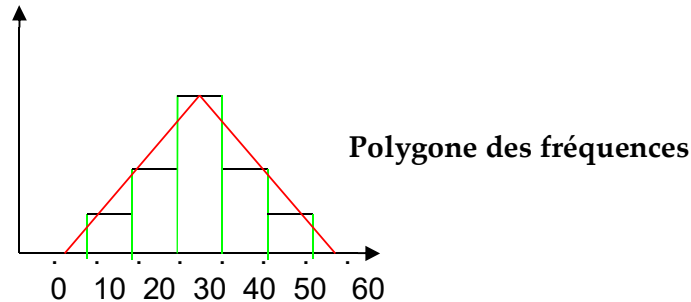


Graphique 2-11

2. Polygone des fréquences :

Il s'agit d'une ligne brisée reliant les milieux des sommets des rectangles de l'histogramme. La fermeture se fait par deux points sur l'axe des abscisses situés respectivement à un demi-intervalle de la borne inférieure de la première classe et de la borne supérieure de la dernière classe.

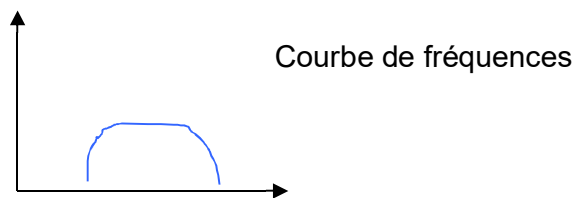
Il permet de donner une image plus lisse du phénomène que l'histogramme. On l'obtient en joignant les milieux des sommets des rectangles de l'histogramme.



Graphique 2-12

Remarque :

- La surface sous le polygone = la surface de l'histogramme.
- Lorsqu'il y a un très grand nombre de classe, l'intervalle de classe devient de plus en plus petit et le polygone de fréquences se transforme en courbe de fréquence.



Graphique 2-13

3. Courbes des effectifs ou des Fréquences cumulées croissantes et décroissantes :

La courbe cumulative permet de connaître le nombre d'observations supérieures ou inférieures à une valeur donnée.

Les deux types de courbes cumulatives sont :

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

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

Considérons le tableau suivant donnant la répartition de 100 individus par classe d'âges :

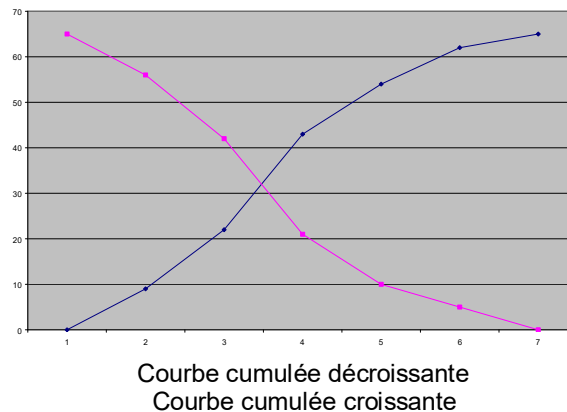
Classes d'âges	Effectifs n_i	Effectifs cumulés croissants $n_i \nearrow$	Fréquences relatives f_i	Fréquences cumulées croissantes $F_i \nearrow$ $F(b_i)$
[5 – 10 [	11	11	0.11	0.11
[10 – 15 [	10	21	0.10	0.21
[15 – 20 [	15	36	0.15	0.36
[20 – 30 [	20	56	0.20	0.56
[30 – 40 [	18	74	0.18	0.74
[40 – 60 [	16	90	0.16	0.9
[60 – 80 [	10	100	0.10	1
Total	100		1	

La lecture des fréquences cumulées croissantes se fait par rapport à la borne supérieure de chaque classe.

La lecture des fréquences cumulées décroissantes se fait par rapport à la borne inférieure de chaque classe.

La représentation graphique de la fonction de répartition appelée courbe cumulative est la suivante :

D'une manière générale, la courbe cumulative, dans le cas d'une variable continue, est une ligne brisée obtenue en joignant différents points de coordonnées (b_i, F_i) où b_i désigne la borne supérieure (ou inférieure) de la classe i , F_i et la fréquence cumulée croissante (ou décroissante) correspondante.



Graphique 2-14

Remarque :

- On peut aussi représenter graphiquement la courbe des fréquences cumulées décroissantes, lesquelles sont définies par la proportion des individus ayant une valeur du caractère supérieure ou égale à la borne inférieure de la classe i .
- On obtiendrait le même graphique si on remplace les fréquences absolues par les fréquences relatives (les pourcentages)

CHAPITRE III : LES CARACTERISTIQUES DE TENDANCE CENTRALE OU LES PARAMETRES DE POSITION

L'observation d'une série de données n'est pas suffisante pour caractériser une distribution. On a souvent besoin de synthétiser ces données à l'aide d'indicateurs ou de paramètres appropriés.

On distingue à cet effet quatre types de caractéristiques :

- Les caractéristiques de tendance centrale
- Les caractéristiques de dispersion
- Les caractéristiques de forme
- Les caractéristiques de concentration

La représentation graphique d'une série statistique nous donne une idée assez générale sur la distribution. Pour confirmer certaines impressions sur la série et pour en donner plus de précision, nous serons amenés à trouver une ou plusieurs valeurs centrales de la variable, capables de résumer la série en caractérisant l'ordre de grandeur des observations. De telles valeurs centrales sont appelées paramètres de tendance centrale ou caractéristiques de position. Un indicateur de position doit être défini de manière rigoureuse et objective, doit tenir compte de l'ensemble des observations de la série et doit être exprimé dans la **même unité** que la variable.

Dans ce chapitre nous allons voir comment synthétiser une série par quelques chiffres. Ces nombres sont appelés caractéristiques de tendance centrale ou paramètres de position.

I. Le Mode :

1. Définition :

On appelle mode ou valeur dominante d'une série statistique la valeur observée de la variable ayant le plus grand effectif (ou la fréquence la plus élevée). Autrement dit, C'est la valeur du caractère le plus fréquent. On note généralement le mode Mo .

Remarque :

- Le mode est exprimé dans la même unité que la variable.
- Si toutes les modalités ont la même fréquence alors la distribution statistique ne possède pas de mode. On parle alors de distribution uniforme.

- Lorsqu'une série possède un seul mode, on dit que la distribution est unimodale. En revanche, lorsqu'elle en possède deux ou plusieurs elle est respectivement qualifiée de bimodale et multimodale.

2. Calcul de M_o :

Le calcul du mode dépend de la nature de la variable, discrète ou continue.

2.1 Cas d'une variable discrète :

Exemple :

On considère les notes obtenues en statistique par un groupe de 20 étudiants :

7, 13, 5, 15, 12, 9, 7, 8, 14, 16, 13, 6, 13, 6, 13, 10, 13, 12, 10, 7, 12, 13

Le mode de cette série correspond à la note la plus fréquente, soit $M_o=13$, valeur qui apparaît cinq fois. L'interprétation en est que la note la plus fréquente est 13.

Exemple :

On considère une distribution statistique d'une population de 69 ménages selon le nombre d'enfants :

Valeurs x_i	Effectifs n_i	Fréquences f_i
1	22	0.32
2	11	0.16
3	16	0.23
4 et plus	20	0.29
Total	69	1.00

Le mode de cette série est : $M_o = 1$. Il signifie que la plupart des ménages ont **un seul** enfant.

Graphiquement, le mode correspond à l'abscisse du bâton le plus élevé.

2.2 Cas d'une variable continue :

Dans le cas d'une variable continue groupée en classes, on parle plutôt de classe modale. La classe modale est la base du rectangle ayant la hauteur la plus élevée.

Cependant, on distingue deux cas selon que les amplitudes des classes sont égales ou inégales.

a) Cas d'amplitudes identiques (égales) :

Dans ce cas, la classe modale est la classe d'effectif n_i la plus élevée, soit $[b_{i-1} - b_i[$. L'effectif de la classe qui précède la classe modale est n_{i-1} et celui de la classe qui suit la classe modale est n_{i+1} .

La détermination du mode à partir de la classe modale se fait de la façon générale suivante :

$$M_o = b_{i-1} + a_i \left(\frac{m_1}{m_1 + m_2} \right) = \frac{b_{i-1} \times m_2 + b_i \times m_1}{m_1 + m_2}$$

Avec :

b_{i-1} : borne inférieure de la classe modale

b_i : borne supérieure de la classe modale

a_i : amplitude de la classe modale

$$m_1 = n_i - n_{i-1}$$

$$m_2 = n_i - n_{(i+1)}$$

Exemple : Soit la distribution de la population de 20 ménages selon le revenu des deux parents :

Classe de Revenu en DA 10 ²	Amplitudes	Effectifs n_i	Fréquences f_i
[200-300[	100	40	0.20
[300-400[	100	60	0.30
[400-500[	100	30	0.15
[500-600[	100	50	0.25
[600-700[	100	20	0.10
Total		200	1

La classe modale est la classe ayant la fréquence la plus élevée. C'est la classe [300-400[dans notre exemple. Dans ce cas, le mode est calculé par :

$$M_o = 300 + 100 \left(\frac{60-40}{(60-40)+(60-30)} \right) = 340 \text{ DA} \times 10^2 = 34000 \text{ DA}$$

Interprétation : Le salaire le plus fréquent est de 34000 DA.

Remarque : On peut aussi utiliser les fréquences relatives au lieu des effectifs.

Dans ce cas, on aura :

$$M_o = 300 + 100 \left(\frac{0.30 - 0.20}{(0.30 - 0.20) + (0.30 - 0.15)} \right) = 34000 \text{ DA}$$

b) Cas d'amplitudes inégales :

Dans le cas où les amplitudes sont différentes, la classe modale est la classe de densité (ou fréquence corrigée) la plus élevée, ou encore d'effectif corrigé le plus élevé.

Le mode est donné par : $M_o = b_{i-1} + a_i \left(\frac{m_i}{m_i + m_{i+1}} \right) = \frac{b_{i-1} \times m_2 + b_i \times m_1}{m_1 + m_2}$

b_{i-1} : borne inférieure de la classe modale

b_i : borne supérieure de la classe modale

a_i : amplitude de la classe modale

$m_1 = h_i - h_{i-1}$, ($m_1 = n_i^c - n_{i-1}^c$);

$m_2 = h_i - h_{i+1}$, ($m_2 = n_i^c - n_{i+1}^c$)

ou h_i , h_{i-1} et h_{i+1} sont les effectifs corrigés

Classes numérotée $[b_{i-1} - b_i[$	Amplitudes a_i	Effectifs n_i^c	Fréquences corrigées f_i^c ou h_i
$[b_0 - b_1[$	a_1	n_1	f_1^c ou h_1
$[b_1 - b_2[$	a_2	n_2	f_2^c ou h_2
		n_{i-1}^c	f_{i-1}^c ou h_{i-1}
$[b_{i-1} - b_i[$	a_i	n_i^c	f_i^c ou h_i
		n_{i+1}^c	f_{i+1}^c ou h_{i+1}
$[b_{p-1} - b_i[$			
Total		N	1

Exemple : Soit la distribution de 100 personnes selon leur âge :

Classes d'âges	Effectifs n_i	Amplitudes a_i	Densités d_i	Effectifs corrigés n_i^c ou h_i
[5-10[	11	5	2,2	22
[10-15[	10	5	2	20
[15-20[	15	5	3	30
[20-30[	20	10	2	20
[30-40[	18	10	1,8	18
[40-60[	16	20	0,8	8
[60-80[	10	20	0,5	5
Total	100			

La plus grande hauteur appartient à la classe [15 – 20[. Donc : $M_o \in [15 - 20[$

$$\text{et } M_o = 15 + 5 \left(\frac{30 - 20}{(30 - 20) + (30 - 20)} \right) = 17.5 = \frac{15 \times 10 + 20 \times 10}{10 + 10}$$

Interprétation : L'âge observé le plus fréquemment est d'environ 17 ans et 6 mois.

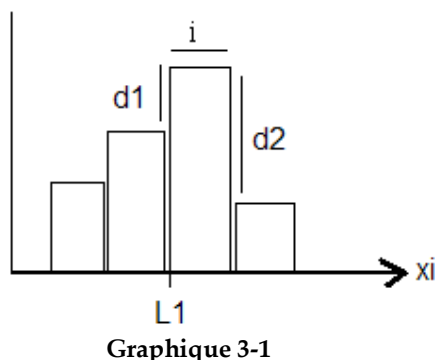
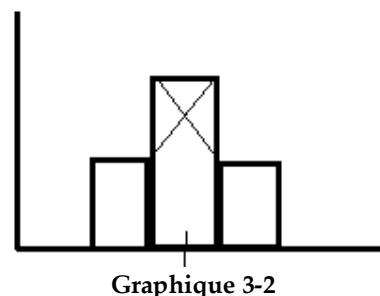
Remarque : On peut aussi utiliser les fréquences corrigées à la place des effectifs corrigés.

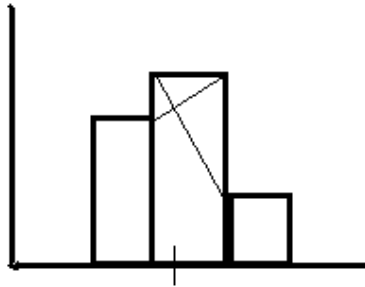
Dans ce cas on aura :

$$M_o = 15 + 5 \left(\frac{0.3 - 0.2}{(0.3 - 0.2) + (0.3 - 0.2)} \right) = 17.5 = \frac{15 \times 0.10 + 20 \times 0.10}{0.1 + 0.1}$$

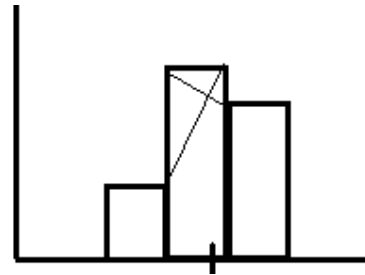
3. Détermination du Mode par la méthode graphique :

Elle consiste d'abord à construire l'histogramme :





Graphique 3-3



Graphique 3-4

N.B : Ne pas oublier, lorsqu'on construit l'histogramme de corriger les effectifs.

II. La Médiane (Me) :

1. Définition :

On appelle médiane d'une série classée par ordre croissant ou décroissant, la valeur du caractère qui partage en deux parties égales les effectifs.

C'est la valeur du caractère telle que la moitié des effectifs lui est supérieure et l'autre moitié lui est inférieure (50% de l'effectif lui soit inférieur et 50% lui soit supérieur). Autrement dit, la médiane d'un ensemble de nombres rangés par ordre de grandeur croissante est la valeur du milieu ou la moyenne arithmétique des deux valeurs centrales.

2. Calcul de Me :

2.1 Cas de données non groupées (DNG) :

Exemple 1 : Nombre d'observation impair : 3-4-4-5-6-8-8-8-10 a pour médiane 6.

Exemple 2 : Nombre d'observation pair : 5-5-7-9-11-12-15-18 a pour médiane.

Intervalle Médian [9 -11]

On prend le centre de l'intervalle comme médiane :

$$\mathbf{Me} = \frac{1}{2} \cdot (9+11) = 10$$

2.2 Cas de données groupées (DG) :

a. Variable statistique discrète (discontinue) :

La médiane est la valeur de la variable statistique qui occupe le $n/2^{em}$ rang (détermination direct).

Exemple :

On considère la répartition d'une population de 80 ménages selon le nombre d'appareils électroménagers existant au sein de leur domicile :

Nombre d'appareils électroménagers (xi)	Nombre de ménages (ni)	Effectifs cumulés croissants n_i ↗
1	8	8
2	12	20
3	16	36
4	32	68
5	12	80
Total	80	

Le rang de la médiane **RMe** = $N/2 = 80/2 = 40$.

D'après le tableau la valeur de la variable qui occupe le 40^{ème} rang est égal à 4 appareils électroménagers donc **Me** = 4.

Interprétation : Cela signifie que 50% des ménages ont moins que 4 appareils électroménagers et 50% des ménages ont plus que 4 appareils électroménagers.

Remarque : La médiane peut aussi déterminée graphiquement et ce en lisant sur le diagramme en escaliers, l'abscisse correspondant à la fréquence cumulée croissante 0.5.

b. Variable statistique continue :

Il n'a aucune différence de calcul pour la médiane selon que les classes sont d'amplitudes constantes ou variables.

Le calcul de la médiane dans le cas de variable continue passe, d'abord, par la détermination de la classe médiane. Ensuite, par interpolation linéaire, on peut calculer la valeur précise de la médiane à l'intérieur de la classe médiane.

Soit $[b_{i-1} - b_i[$ la classe médiane, a_i l'amplitude de la classe médiane, N_i l'effectif cumulé croissant de la classe médiane, N_{i-1} l'effectif cumulé croissant de la classe avant la classe médiane et N l'effectif total.

L'expression de la médiane est donnée par interpolation linéaire :

$$M_e = b_{i-1} + a_i \left(\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}} \right)$$

La même démarche pourrait être utilisée en remplaçant les fréquences absolues par les fréquences relatives :

$$M_e = b_{i-1} + a_i \left(\frac{0.5 - F_{i-1}}{F_i - F_{i-1}} \right)$$

ou F_i désigne la fréquence cumulée croissante de la classe médiane, F_{i-1} la fréquence cumulée croissante de la classe qui précède la classe médiane.

Exemple : En reprenant notre exemple sur la répartition des 100 individus selon leur âge :

Classes d'âges	Effectifs n_i	Effectifs cumulés croissants $n_i \nearrow$	Fréquences relatives f_i	Fréquences cumulées croissantes $F_i \nearrow$ $F(b_i)$
[5 – 10 [	11	11	0.11	0.11
[10 – 15 [	10	21	0.10	0.21
[15 – 20 [	15	36	0.15	0.36
[20 – 30 [	20	56	0.20	0.56
[30 – 40 [	18	74	0.18	0.74
[40 – 60 [	16	90	0.16	0.9
[60 – 80 [	10	100	0.10	1
Total	100		1	

Le calcul, par interpolation linéaire, de la médiane donne :

$$\frac{M_e - 20}{30 - 20} = \left(\frac{50 - 36}{56 - 36} \right) = \left(\frac{0.50 - 0.36}{0.56 - 0.36} \right)$$

Dans notre exemple : $\frac{N}{2} = 50$. La classe médiane est la classe à laquelle appartient la valeur médiane, c'est-à-dire la classe [20-30[, d'où :

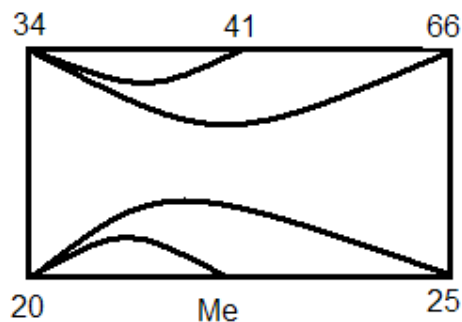
$$M_e = 20 + 5 \left(\frac{50 - 36}{66 - 36} \right) = 27 \text{ ans}$$

Interprétation : 50 % des individus sont âgés de moins de 27 ans.

Méthode rapide d'interpolation :

Exemple :

Salaires	Effectifs n_i	Effectifs cumulés croissants $n_i \nearrow$
[10 – 15 [	9	9
[15 – 20 [	25	34
[20 – 25 [	32	66
[25 – 30 [	16	82
Total	82	

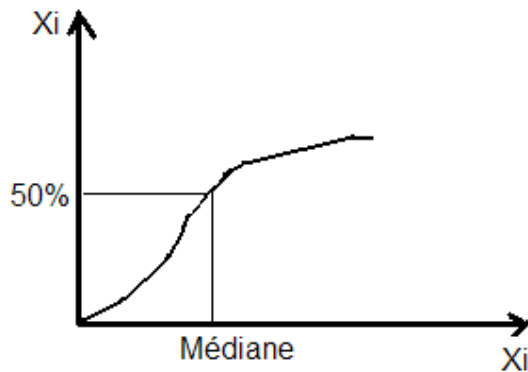


Graphique 3-5

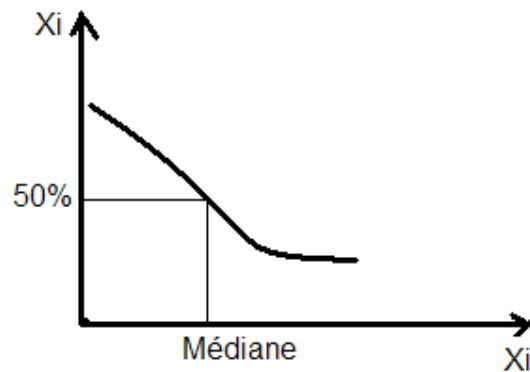
$$\frac{Me - 20}{25 - 20} = \frac{41 - 34}{66 - 34} \Leftrightarrow Me = \frac{7 \times 5}{32} + 20 \approx 21$$

Interprétation : Le 41^{ème} individu normalement la médiane devrait se situer entre le 41^{ème} et le 42^{ème}, mais on convient lorsque les effectifs sont nombreux de prendre ($N / 2$).

3. Détermination graphique de la médiane :



Graphique 3-6



Graphique 3-7

Courbes cumulatives

III. La moyenne arithmétique :

1. Définition :

La moyenne arithmétique, dite simplement moyenne est notée $\bar{x}$, est la caractéristique de tendance centrale la plus usuelle.

2. Calcul de la moyenne arithmétique :

2.1. Cas de données non groupées (DNG) :

Soit une série statistique de N observations : $x_1, x_2, x_3, \dots, x_n$. La moyenne arithmétique simple de ces observations, utilisées pour des valeurs individuelles (pour les valeurs isolées). Elle est donnée par :

$$\bar{X} = \frac{\text{Somme de toutes les observations}}{\text{Le nombre d'observations}}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N} = \frac{\sum_{i=1}^N x_i}{N}$$

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

: Une moyenne arithmétique simple

Exemple :

On observe les notes en statistique d'un groupe d'étudiants :

14 , 16, 12, 9, 11, 16, 7, 9, 7, 9

La moyenne arithmétique simple de ces notes est :

$$\bar{x} = \frac{14 + 16 + 12 + 9 + 11 + 16 + 7 + 9 + 7 + 9}{10} = 11$$

2.2. Cas de données groupées (DG) :

2.2.1 Cas d'une variable statistique discrète :

Dans le cas d'une variable statistique discrète, la moyenne arithmétique s'écrit :

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + n_3x_3 + \dots + n_px_p}{N} = \frac{\sum_{i=1}^{i=p} n_ix_i}{N}$$

$$\bar{X} = \frac{x_1 + x_1 + \dots + x_2 + x_2 + \dots}{n_1 + n_2 + \dots + n_n}$$

où x_i , et n_i , $i=1,2,\dots,p$ représentent respectivement la valeur du caractère et l'effectif correspondant, et p est le nombre de valeurs prises par la variable.

Dans le cas d'une variable continue, ou les données sont groupées en classes, on applique la même formule, en remplaçant les x_i par les centres de classes x_i dans ce cas on a :

$$\bar{x} = \frac{\sum_{i=1}^{i=p} n_ix_i}{N}$$

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

: Une moyenne arithmétique pondérée

Exercice :

Soit la série de notes suivante :

2-6-12-10-12-10-10-6

$$\bar{X} = \frac{2+6+12+10+12+10+10+6}{8} = \frac{68}{8}$$

$$\bar{X} = 8,5$$

Exercice :

Soit la série des notes de l'exercice qui peut être présentée de la manière suivante :

Notes x_i	Effectifs n_i	$n_i x_i$
2	1	2
6	2	12
10	3	30
12	2	24
Total	8	68

$$\bar{X} = \frac{68}{8} = 8,5$$

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

2.2.2 Cas d'une variable statistique continue :

Exercice 3 :

Soit la série des notes de l'exercice qui peut être présentée de la manière suivante :

Classe d'âges	Effectifs n_i	Centre de classes x_i	$n_i x_i$
[20-25[	8	22,5	180
[25-30[	10	27,5	275
[30-35[	20	32,5	650
[35-40[	25	37,5	937.5
[40-45[	15	42,5	637.5
[45-50[	10	47,5	475
Total	88		3155

L'âge moyen est donnée par :

$$\bar{X} = \frac{3155}{88} = 35,85 \text{ Années : Moyenne de l'âge ou l'âge moyen}$$

2.2.3 Méthode de changement de variable :

Lorsque les calculs sont compliqués, on peut les simplifier en précédant à un changement de variable

Par changement d'échelle :

La variable x_i peut s'écrire : $x_i = a \cdot z_i$

a = nouvelle échelle (amplitude de la classe)

z_i = nouvelle variable

Exemple :

$$\begin{array}{c|c|c|c|c} X_i & & a & * & Z_i \\ \hline 24 & = & 1 & * & 24 \\ \hline 36 & = & 1 & * & 36 \end{array}$$

$$\begin{array}{l} a=1 \\ x_i = z_i \end{array}$$

$$\begin{array}{c|c|c|c|c} X_i & & a & * & Z_i \\ \hline 24 & = & 6 & * & 4 \\ \hline 36 & = & 6 & * & 6 \end{array}$$

$$\begin{array}{ll} a = 6 & Z_i = 4 \\ a = 6 & Z_i = 6 \end{array}$$

Par changement d'origine et d'échelle : tout variable x_i peut s'écrire :

$$x_i = x_0 + a z_i$$

x_0 = nouvelle origine

a : nouvelle échelle

x'_i : nouvelle variable

Exemple :

$$\begin{array}{c|c|c|c|c|c|c} X_i & & X_0 & + & a & * & X'_i \\ \hline 14 & = & 4 & + & 2 & * & 5 \\ \hline 22 & = & 4 & + & 2 & * & 9 \end{array}$$

Si on pose $x_i = x_0 + a x'_i \Rightarrow z_i = \frac{x_i - x_0}{a}$

La moyenne arithmétique :

$$\begin{aligned}\bar{X} &= \frac{\sum n_i x_i}{\sum n_i} \\ &= \frac{\sum n_i (x_0 + a z_i)}{\sum n_i} \\ &= \frac{x_0 \sum n_i + a \sum n_i z_i}{\sum n_i} \\ \bar{X} &= x_0 + a \frac{\sum n_i z_i}{\sum n_i} \\ \bar{X} &= x_0 + a \bar{z} \text{ avec } \bar{z} = \frac{\sum n_i z_i}{\sum n_i}\end{aligned}$$

x_0 = nouvelle origine

a = nouvelle échelle (amplitude de classe)

$\bar{Z}$ = Moyenne provisoire

La moyenne arithmétique par la méthode de changement de variables :

$$\bar{X} = x_0 + a \bar{z}$$

Remarque :

- On utilise cette relation pour simplifier les calculs de la manière suivante.
- On prend pour x_0 la valeur de caractère la plus fréquente.
- On prend « a » l'intervalle des classes lorsque les classes sont égaux.
- Lorsque l'on a des classes d'amplitude inégales on prend le plus grand diviseur commun (PGDC) des amplitudes comme amplitude de classe

Exercice :

On prend les mêmes données du tableau précédant et on calcule la moyenne arithmétique par la méthode de changement de variable :

Classe d'âges	Effectifs n_i	Centre de classes x_i	$z_i = (x_i - x_0)/a$	$n_i z_i$
[20-25[	8	22,5	-3	-24
[25-30[	10	27,5	-2	-20
[30-35[	20	32,5	-1	-20
[35-40[	25	37,5	0	0
[40-45[	15	42,5	1	15
[45-50[	10	47,5	2	20
Total	88			-29

La moyenne arithmétique par la méthode de changement de variables :

$$\bar{X} = x_0 + a\bar{z}$$

$x_0 = 37,5$ c'est le centre de classe modale

$a = 5$

$z_i = (x_i - x_0)/5$

$$\bar{z} = \frac{\sum n_i z_i}{\sum n_i} = \frac{-29}{88} = -5.27 \quad \text{Moyenne provisoire}$$

$$\bar{X} = 37,5 + 5(-29/88) = 35,8 \text{ ans} \quad \text{Moyenne arithmétique}$$

2.2.4 Calcul de la moyenne arithmétique à l'aide des fréquences relatives :

$$\bar{X} = \frac{\sum n_i x_i}{\sum n_i} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_i x_i + n_n x_n}{\sum n_i}$$

$$\begin{aligned} \bar{X} &= \frac{n_1 x_1}{\sum n_i} + \frac{n_2 x_2}{\sum n_i} + \dots + \frac{n_n x_n}{\sum n_i} \\ &= f_1 x_1 + f_2 x_2 + \dots + f_n x_n \end{aligned}$$

Fréquence relative : $f_i = \frac{n_i}{\sum n_i}$

d'où : $\bar{X} = f_1 x_1 + f_2 x_2 + \dots + f_n x_n$

Exemple :

x_i	Effectifs n_i	f_i	$f_i x_i$
10	5	0.125	1.25
11	8	0.20	2.2
12	10	0.25	3
13	12	0.3	3.9
14	5	0.125	1.75
Total	40	1	12.1

$$\bar{X} = \sum f_i x_i$$

$$\bar{X} = 12,1$$

3. Propriétés de la moyenne arithmétique :

- La somme des écarts à la moyenne arithmétique est toujours nulle :

- Pour des données non groupées : $\sum_{i=1}^N (x_i - \bar{x}) = 0$

En effet :

$$\sum_{i=1}^N (x_i - \bar{x}) = \sum_{i=1}^N x_i - \sum_{i=1}^N \bar{x} = \sum_{i=1}^N x_i - N\bar{x} = N\bar{x} - N.\bar{x} = 0$$

- Pour des données groupées en classes, On a : $\sum_{i=1}^N n_i(x_i - \bar{x}) = 0$

En effet :

$$\sum_{i=1}^N n_i x_i - N.\bar{x} = N.\bar{x} - N.\bar{x} = 0$$

- La moyenne arithmétique $\bar{x}$ d'une population d'effectif N composée de K sous-populations d'effectifs N_k et de moyenne $\bar{x}_k$ (**Moyenne arithmétique combinée**) est égale à :

$$\bar{x} = \frac{N_1 \bar{x}_1 + \dots + N_p \bar{x}_p}{N} = \frac{\sum_{i=1}^p N_i \bar{x}_i}{N}$$

- La moyenne arithmétique est le critère le plus fréquemment utilisé pour définir une valeur moyenne d'observations d'une variable additive comme par exemple : la taille, le poids, l'âge...etc. Il y a d'autres variables dont le calcul de la moyenne se traite autrement, comme, par exemple, le taux de chômage ou d'inflation, le taux de change, la vitesse sur différents parcours...etc.

IV. Autres moyennes :

1. La moyenne géométrique (G) :

La moyenne géométrique d'une variable, notée généralement G, est égale à la racine N^{ième} du produit des N valeurs observées de cette variable. Elle est utilisée souvent dans le calcul des taux de croissance moyens et de certains indices statistiques synthétiques.

Cas de données non groupées (DNG) :

La moyenne géométrique simple est donnée par :

Étant donnée n observations connues individuellement $(X_1, X_2, X_3, \dots, X_n)$.
On appelle moyenne géométrique simple de ces n observations la grandeur G :

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n} = (X_1 \cdot X_2 \cdot \dots \cdot X_n)^{1/n}$$

$$G = \left(\prod_{i=1}^{i=n} x_i \right)^{1/n}$$

Remarque :

Calculer G est plus facile en passant par le logarithme, en effet :

$$G = \sqrt[n]{X_1 \cdot X_2 \cdot \dots \cdot X_n} = (X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n)^{1/n}$$

$$\log G = 1/n \log (X_1 \cdot X_2 \cdot \dots \cdot X_n)$$

$$= 1/n [\log X_1 + \log X_2 \cdot \dots + \log X_n]$$

$\log G = \frac{\sum \log X_i}{N_i}$

La moyenne géométrique pondérée :

Cas de données groupées (DG) :

Lorsque les observations sont groupées ; chaque variable X_i sera pondérée par l'effectif correspondant, la moyenne géométrique pondérée s'écrit :

$$G = \sqrt[n]{X_1 \cdot X_1 \cdot X_1 \cdot \dots \cdot X_2 \cdot X_2 \cdot X_2 \cdot \dots \cdot X_3 \cdot \dots \cdot X_3} \Leftrightarrow G = \sqrt[n]{X_1^{n_1} \cdot X_2^{n_2} \cdot X_3^{n_3} \cdot \dots \cdot X_n^{n_n}}$$

$$N = n_1 + n_2 + \dots + n_n$$

$$G = \sqrt[n]{x_1^{n_1} x_2^{n_2} \dots x_n^{n_n}}$$

$$G = (x_1^{n_1} x_2^{n_2} \dots x_n^{n_n})^{1/n}$$

$$\begin{aligned} \log G &= \frac{1}{n} \log(x_1^{n_1} x_2^{n_2} \dots x_n^{n_n}) = \frac{\log(x_1^{n_1} x_2^{n_2} \dots x_n^{n_n})}{n} \\ &= \frac{n_1 \log x_1 + n_2 \log x_2 + \dots + n_n \log x_n}{n} \end{aligned}$$

$$\log G = \frac{\sum n_i \log x_i}{\sum n_i}$$

La moyenne géométrique peut être exprimée en fonction des fréquences relatives de la manière suivante :

$$G = x_1^{\left(\frac{n_1}{N}\right)} \cdot x_2^{\left(\frac{n_2}{N}\right)} \cdot x_3^{\left(\frac{n_3}{N}\right)} \dots x_p^{\left(\frac{n_p}{N}\right)} = x_1^{(f_1)} \cdot x_2^{(f_2)} \dots x_p^{(f_p)}$$

Résumé :

- Le logarithme de la moyenne géométrique est égal à la moyenne arithmétique des logarithmes des x_i .
- En pratique, le calcul de la moyenne géométrique passe par le logarithme. Ainsi, dans le cas des données non groupées, on a :

$$\ln(G) = \frac{1}{N} \sum_{i=1}^N \ln(x_i)$$

- Et dans le cas des données groupées par classes, on a :

$$\ln(G) = \frac{1}{N} \sum_{i=1}^p n_i \ln(x_i)$$

Remarque :

- La moyenne géométrique est utilisée quand les valeurs de la variable sont liées de façon multiplicative les unes aux autres.

- La moyenne géométrique est notamment utilisée pour calculer des taux de croissance moyens.
- La moyenne géométrique d'un produit de deux variables est égale au produit de leur moyenne géométrique.
- La moyenne géométrique d'un rapport de deux variables (n'égal pas 0) est égale au rapport de leurs moyennes géométriques.

Exercice : On va calculer la moyenne géométrique du tableau suivant :

x_i	Effectifs n_i	$\log x_i$	$n_i \cdot \log x_i$
2	1	0.3010	0.3010
6	2	0.7781	1.5562
10	3	1	3
12	2	1.0791	2.158
Total	8		7.0152

Solution :

$$\log G = \frac{\sum \log X_i}{N_i}$$

$$\Rightarrow \log G = \frac{7,0152}{8} = 0,8769$$

$$\Rightarrow G = 10^{0,8769} = 7.531$$

2. La moyenne harmonique (H) :

La moyenne harmonique, notée H, est égale à l'inverse de la moyenne arithmétique des inverses des valeurs :

Cas de données non groupées (DNG) :

Étant donnée n observations connues individuellement $X_1, X_2, X_3, \dots, X_n$.

On appelle moyenne harmonique le nombre H tel que :

Moyenne harmonique simple

$$\frac{1}{H} = \frac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n} = \frac{\sum \frac{1}{x_i}}{n}$$

Cas de données groupées (DG) :

Si les observations sont groupées la moyenne harmonique H s'écrit :

$$\frac{1}{M} = \frac{x_1 \cdot \frac{1}{x_1} + x_2 \cdot \frac{1}{x_2} + \dots + x_n \cdot \frac{1}{x_n}}{n_1 + n_2 + \dots + n_n} = \frac{\sum n_i \frac{1}{x_i}}{\sum n_i}$$

$$H = \frac{\sum n_i}{\sum n_i \frac{1}{x_i}}$$

Moyenne harmonique pondérée

Remarques :

- L'inverse de la moyenne harmonique est égale à la moyenne arithmétique des inverses des x_i .
- La moyenne harmonique est généralement employée lorsque la variable observée est égale au rapport de deux variables exprimées dans deux unités différentes, par exemple le prix d'un bien exprime en unités monétaires par unité de bien, la vitesse exprimée en unités de distance par unité de temps.

Exemple : On va calculer la moyenne harmonique du tableau suivant :

x_i	Effectifs n_i	$\frac{1}{x_i}$	$n_i \cdot \frac{1}{x_i}$
2	1	0.5	0.5
6	2	0.166	0.332
10	3	0.1	0.3
12	2	0.083	0.166
Total	8		1.298

Solution :

$$\frac{1}{H} = \frac{\sum \frac{1}{x_i}}{n}$$

$$\Rightarrow \frac{1}{H} = \frac{\sum x_i \frac{1}{x_i}}{n_i} = \frac{1,298}{8}$$

$$\Leftrightarrow H = \frac{8}{1,298} = 6,16$$

Remarque :

$$\frac{1}{H} = \frac{\sum n_i \frac{1}{x_i}}{\sum n_i} = \frac{\sum n_i \cdot X_i}{\sum n_i} \text{ avec } x_i = \frac{1}{X_i}$$

L'inverse de la moyenne = moyenne des inverses

3. La moyenne quadratique (Q) :

La moyenne quadratique d'une variable statistique, notée Q, est égale à la racine carrée de la moyenne arithmétique des carrés des valeurs de la variable.

Cas de données non groupées (DNG) :

Etant donné n observations connues individuellement $X_1 ; X_2 ; \dots, X_n$

$$Q^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} \Leftrightarrow Q^2 = \frac{\sum x_i^2}{n}$$

Moyenne quadratique simple :

$$\Leftrightarrow Q = \sqrt{\frac{\sum x_i^2}{n}}$$

Cas de données groupées (DG) :

Si les observations sont groupées, la moyenne quadratique s'écrit :

$$Q^2 = \frac{n_1 \cdot x_1^2 + n_2 \cdot x_2^2 + \dots + n_n \cdot x_n^2}{n_1 + n_2 + \dots + n_n}$$

$$Q^2 = \frac{\sum n_i \cdot x_i^2}{\sum n_i} \Leftrightarrow Q = \sqrt{\frac{\sum n_i \cdot x_i^2}{\sum n_i}} \quad \text{Moyenne quadratique pondérée}$$

Exemple : On va calculer la moyenne quadratique du tableau suivant :

x_i	Effectifs n_i	x_i^2	$n_i \cdot x_i^2$
2	1	4	4
6	2	36	72
10	3	100	300
12	2	144	288
Total	8	284	664

Solution :

$$Q^2 = \frac{\sum x_i^2}{n} \Leftrightarrow Q = \sqrt{\frac{\sum x_i^2}{n}}$$

$$\Rightarrow Q^2 = \frac{\sum n_i \cdot x_i^2}{\sum n_i} = \frac{664}{8} = 83$$

$$\Leftrightarrow Q = \sqrt{83} = 9.11$$

Remarque :

$$Q^2 = \frac{\sum n_i \cdot x_i^2}{\sum n_i} = \frac{\sum n_i \cdot X_i}{\sum n_i} \text{ avec } X_i = x_i^2$$

Carré de la moyenne = la moyenne des carrés

- La moyenne quadratique est souvent utilisée dans le calcul de la variance (voir le chapitre suivant).
- Les moyennes quadratique et arithmétique tiennent compte davantage des valeurs les plus élevées de la série statistique. En revanche, les moyennes géométrique et harmonique réduisent l'influence des observations les plus élevées.
- Les relations existantes entre les différentes moyennes.

4. Généralisation de la notion de moyenne : Moyenne d'ordre r

On appelle moyenne d'ordre r la quantité M_r tel que :

$$M_r = \left(\frac{x_1^r + x_2^r + \dots + x_n^r}{n} \right)^{1/n}$$

$$M_r^r = \frac{x_1^r + x_2^r + \dots + x_n^r}{n}$$

$$\text{Si } r = 1 \quad M_1^1 = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \Rightarrow M_1^1 = \bar{X}$$

$$\text{si } r = 2 \quad M_2^2 = \frac{x_1^2 + x_2^2 + \dots + x_n^2}{n} \Rightarrow M_2^2 = Q^2 \Rightarrow M_2 = Q$$

$$\text{si } r = -1 \quad M_{-1}^{-1} = \frac{x_1^{-1} + x_2^{-1} + \dots + x_n^{-1}}{n} \Rightarrow \frac{1}{M_{-1}} = H^{-1} = \frac{1}{H} \Rightarrow M_{-1} = H$$

si $r = \varepsilon \rightarrow 0$.

5. Classement des moyennes : Les inégalités entre les moyennes

On démontre que les moyennes s'ordonnent selon la valeur de r, c'est-à-dire que si :

$$r_1 < r_2 \Rightarrow M_{r_1} < M_{r_2}$$

Ce qui nous donne : $M_{-1} < M_0 < M_1 < M_2$

$$X_{\text{Min}} < H < G < \bar{X} < Q < X_{\text{Max}}$$

Dans notre exemple, on trouve : $2 < 6.16 < 7.531 < 8.5 < 9.11 < 12$

Le choix d'une moyenne :

L'utilisation d'une moyenne parmi les quatre moyennes dépend du problème posé.

CHAPITRE IV : LES CARACTERISTIQUES DE DISPERSION

Très souvent les indicateurs de tendance centrale (mode, médiane et moyenne) s'avèrent insuffisants pour permettre de résumer à eux seuls et de comparer deux ou plusieurs séries statistiques, car deux séries statistiques, peuvent avoir le même mode, la même médiane, la même moyenne et, cependant, être distribuées différemment.

Prenons, à titre d'exemple, deux séries A et B représentant les nombres mensuels de pièces jugées défectueuses par un contrôleur de pièces mécaniques pendant 7 mois :

Série A	78	79	80	80	80	81	82
Série B	30	35	80	80	80	125	130

On constate que ces deux séries ont un même mode ($Mo = 80$), une même médiane ($Me = 80$) et une même moyenne ($\bar{X} = 80$). Cependant, elles ne sont pas comparables : leur distribution se fait d'une manière nettement différente. Autrement dit, elles sont différentes par la dispersion de leurs valeurs.

Dans la série, les valeurs sont resserrées ou concentrées autour de la valeur centrale : 80 ; tandis que dans la série B les valeurs sont plus étalées, plus dispersées par rapport à cette valeur centrale typique. D'où la nécessité de calculer d'autres indicateurs capables de rendre compte des écarts entre les différentes valeurs observées et la valeur centrale. Ces indicateurs, qui nous informent sur la variabilité des valeurs observées, sont appelés indicateurs de dispersion.

Ainsi pour donner une représentation plus suffisante du phénomène étudié, il est important de mesurer la dispersion entre les différentes valeurs prise par une variable statistique et la valeur centrale.

Il faut donc lorsqu'on résume une série, indiquer par un nombre si les valeurs sont proches ou éloignées de la valeur centrale.

Ce nombre est appelé caractéristiques ou indicateurs de dispersion.

I. L'étendu ou l'intervalle de variation :

On appelle étendue ou intervalle de variation d'une série statistique, la différence entre la plus élevée et la plus faible des valeurs observées, soit :

$$E = X_{\text{MAX}} - X_{\text{MIN}}$$

$$\text{Série 1: } \mathbf{E_1} = 19 - 1 = 18$$

$$\text{Série 2 : } \mathbf{E_2} = 12 - 8 = 4$$

L'étendu ou l'intervalle de variation n'est pas un indicateur toujours fiable, car il dépend des valeurs extrêmes qui peuvent être fausses ou « aberrantes ».

II. Les écarts inter quantiles :

Il s'agit des écarts entre les premiers et les derniers principaux quantiles.

1. Les quantiles :

Comme pour la médiane où l'on s'est intéressé à la valeur de la variable qui partage la population en deux parties d'égal effectif, on s'intéresse ici aux valeurs qui partagent la population en quatre, en dix ou en cent parties de même effectif. Ces valeurs sont appelés respectivement quartiles, déciles et centiles.

D'une manière générale, on appelle quantile d'ordre α (alpha), la valeur de la variable X_α telle que $\alpha\%$ des valeurs lui sont inférieures.

On peut alors écrire : $F(X_\alpha) = \alpha\%$, où F désigne la fonction de répartition de la variable. La détermination des différents quantiles se fait de la même manière que la médiane (par interpolation linéaire).

Les principaux quantiles sont les quartiles, les déciles et les percentiles (ou les centiles).

2. Les quartiles :

Les quartiles en nombre de trois notés Q_1 , Q_2 et Q_3 , sont les valeurs d'une variable, rangées par ordre croissant ou décroissant, qui partagent la population étudiée en quatre parties de même effectif. L'expression des trois quartiles peut-être dérivée de la même manière que la médiane.

Soit $[b_{i-1} - b_i[$ la classe d'amplitude a_i à laquelle appartient Q_i , N_i l'effectif cumulé croissant de cette classe, N_{i-1} l'effectif cumulé croissant de la classe précédant la classe $[b_{i-1} - b_i[$ et N l'effectif total.

L'expression du premier quartile est donnée par :

$$Q_1 = b_{i-1} + a_i \left(\frac{\frac{N}{4} - N_{i-1}}{N_i - N_{i-1}} \right)$$

La même démarche pourrait être utilisée en remplaçant les fréquences absolues par les fréquences relatives :

$$Q_1 = b_{i-1} + a_i \left(\frac{0.25 - F_{i-1}}{F_i - F_{i-1}} \right)$$

Où F_i désigne la fréquence cumulée croissante de la classe du premier quartile, F_{i-1} la fréquence cumulée croissante de la classe qui précède la classe du premier quartile.

Si Q_2 appartient à $[b_{i-1} - b_i[$ alors :

L'expression du second quartile est donnée par :

$$Q_2 = b_{i-1} + a_i \left(\frac{\frac{2N}{4} - N_{i-1}}{N_i - N_{i-1}} \right)$$

La même démarche pourrait être utilisée en remplaçant les fréquences absolues par les fréquences relatives :

$$Q_2 = b_{i-1} + a_i \left(\frac{0.5 - F_{i-1}}{F_i - F_{i-1}} \right) = Me$$

Q_2 (deuxième quartile) : valeur de la variable telle que 50% des observations lui soient inférieures.

Si Q_3 appartient à $[b_{i-1} - b_i[$ alors :

L'expression du second quartile est donnée par :

$$Q_3 = b_{i-1} + a_i \left(\frac{\frac{3N}{4} - N_{i-1}}{N_i - N_{i-1}} \right)$$

La même démarche pourrait être utilisée en remplaçant les fréquences absolues par les fréquences relatives :

$$Q_3 = b_{i-1} + a_i \left(\frac{0.75 - F_{i-1}}{F_i - F_{i-1}} \right)$$

Q₂ (deuxième quartile) : valeur de la variable telle que 75% des observations lui soient inférieures.

Résumé :

- Q₁ : 1^{er} quartile Q₁ la valeur du caractère tel que : 25% des observations lui sont inférieurs et 75% lui sont supérieurs. 25% < ; 75% >
- Q₂ : 2^{ème} quartile Q₂ = Me 50% < ; 50% >
- Q₃ : 3^{ème} quartile Q₃ = 75% < ; 25% >

3. Les déciles :

Les déciles en nombre de neuf notés D₁, D₂,...et D₉, sont les valeurs d'une variable, rangées par ordre croissant ou décroissant, qui partagent la population étudiée en dix sous-populations de même effectif. L'expression des neufs déciles peut-être dérivée de la même manière que la médiane et les quartiles.

De la même manière, on peut définir les expressions des déciles :

$$D_1 = b_{i-1} + a_i \left(\frac{\frac{1}{10}N - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left(\frac{0.10 - F_{i-1}}{F_i - F_{i-1}} \right)$$

$$D_2 = b_{i-1} + a_i \left(\frac{\frac{2}{10}N - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left(\frac{0.20 - F_{i-1}}{F_i - F_{i-1}} \right)$$

.....

$$D_9 = b_{i-1} + a_i \left(\frac{\frac{9}{10}N - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left(\frac{0.90 - F_{i-1}}{F_i - F_{i-1}} \right)$$

D₁ (premier décile) : valeur de la variable telle que 10% des observations lui soient inférieures.

D₂ (deuxième décile) : valeur de la variable telle que 20% des observations lui soient inférieures.

.....

D₉ (neuvième décile) : valeur de la variable telle que 90% des observations lui soient inférieures.

Remarque :

$$Q2 = Me = D5$$

4. Les percentiles (ou les centiles) :

Les percentiles en nombre de 99 notés $C_1, C_2, \dots$ et C_{99} et appelés aussi centiles, sont les valeurs d'une variable, rangées par ordre croissant ou décroissant, qui partagent la population étudiée en cent sous-populations d'égal effectifs. L'expression des neufs déciles peut-être dérivée de la même manière que la médiane et les quartiles.

De la même manière, on peut définir les expressions des percentiles :

$$P_1 = b_{i-1} + a_i \left(\frac{\frac{1}{100}N - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left(\frac{0.01 - F_{i-1}}{F_i - F_{i-1}} \right)$$

$$P_2 = b_{i-1} + a_i \left(\frac{\frac{2}{100}N - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left(\frac{0.02 - F_{i-1}}{F_i - F_{i-1}} \right)$$

.....

.....

.....

$$P_{99} = b_{i-1} + a_i \left(\frac{\frac{99}{100}N - N_{i-1}}{N_i - N_{i-1}} \right) = b_{i-1} + a_i \left(\frac{0.99 - F_{i-1}}{F_i - F_{i-1}} \right)$$

P_1 (premier percentile) : valeur de la variable telle que 1% des observations lui soient inférieures.

P_2 (deuxième percentile) : valeur de la variable telle que 2% des observations lui soient inférieures.

.....

P_{99} (99^{ième} percentile) : valeur de la variable telle que 99% des observations lui soient inférieures.

Remarque :

- $Q2 = Me = D5 = P50$
- Les quartiles, les déciles et les percentiles (ou centiles) permettent de calculer les différents intervalles interquartiles. La longueur de ces intervalles correspond aux écarts inter quantiles qui sont des indicateurs qui sont des

indicateurs de dispersion. Plus la longueur de l'intervalle est grande, plus la dispersion est forte.

5. Les intervalles inter quantiles :

On distingue :

- L'intervalle interquartile, qui contient 50% des observations, est : $\{Q1, Q3\}$.
L'écart interquartile est égal à : $EQ = Q3 - Q1$.
- L'intervalle inter décile, qui contient 80% des observations, est : $\{D1, D9\}$.
L'écart inter décile est égal à : $ED = D9 - D1$.
- L'intervalle inter percentile, qui contient 50% des observations, est : $\{Q1, Q3\}$.
L'écart interquartile est égal à : $EP = P99 - Q1$.

Exercice :

Salaires 10 ³ DA	Effectifs
10-15	9
15-20	25
20-25	32
25-30	16
Total	82

Travail à faire : Calculer l'écart inter quartile.

Solution :

$$Q3 - Q1$$

Calcul de Q1 :

$$N = 82$$

$$\text{Rang} : 82/4 = 20,5$$

Classe du Q1: [15-20]

Interpolation : $15 + \Delta$

Salaires 10 ³ DA	Effectifs
10-15	9
15-20	25
20-25	32
25-30	16
Total	82

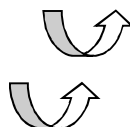
Ecart Inter quartile

$$\begin{aligned} Q3 - Q1 \\ &= 24,3 - 17,3 \\ &= 7 \text{ DA} \end{aligned}$$

Ni Cum	↗
9	
34	
66	
82	

Interprétation : Si 25 individus

Si 01 Individu

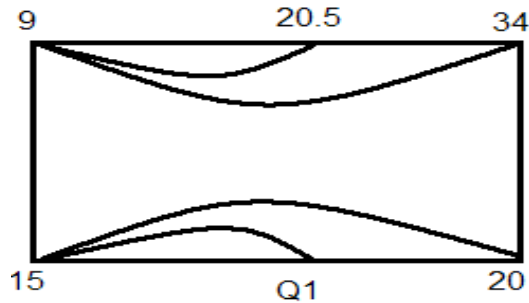


Augmentation de $5 \cdot 10^3$ DA

Augmentation $5/25 \cdot 10^3$ DA

$$(20,5 - 9) = 11,5 \quad \curvearrowright \quad 5/25 * 11,5$$

Donc $Q_1 = 15 + 5/25 * 11,5 = 17,3 \cdot 10^3 \text{ DA} = 17300 \text{ DA}$



$$\begin{aligned} Q_1 - 15 &= 20.5 - 9 \\ 20 - 15 &= 34 - 9 \end{aligned}$$

Graphique 4-1

2ème Méthode :

$$Q_1 = b_{i-1} + a_i \left(\frac{\frac{N}{4} - N_{i-1}}{N_i - N_{i-1}} \right)$$

$$Q_1 = 15 + (20 - 15) \left(\frac{\frac{82}{4} - 9}{34 - 9} \right) = 17,3 \cdot 10^3 \text{ DA} = 17300 \text{ DA}$$

Interprétation : Ce qui signifie que 25% des salariés perçoivent un salaire inférieur à 17300 DA et 75% des salariés perçoivent un salaire supérieur à 17300 DA.

Calcul de Q3 :

Rang : $82 * 3/4 = 61,5$

Classe du Q3 = [20-25]

Interpolation : Si 32 individus $\curvearrowright$

Augmentation de 5 DA

01 Individu $\curvearrowright$

Augmentation de 5/32

$(61,5 - 34) = 27,5$ individus $\curvearrowright$

Augmentation $5/32 * 27,5$

Donc $Q_3 = 20 + [(5/32) * 27,5]$

2ème Méthode :

$$Q_3 = b_{i-1} + a_i \left(\frac{\frac{3N}{4} - N_{i-1}}{N_i - N_{i-1}} \right)$$

$$Q_3 = 20 + (25 - 20) \left(\frac{\frac{3.82}{4} - 34}{66 - 34} \right) = 24,3 \cdot 10^3 \text{ DA} = 24300 \text{ DA}$$

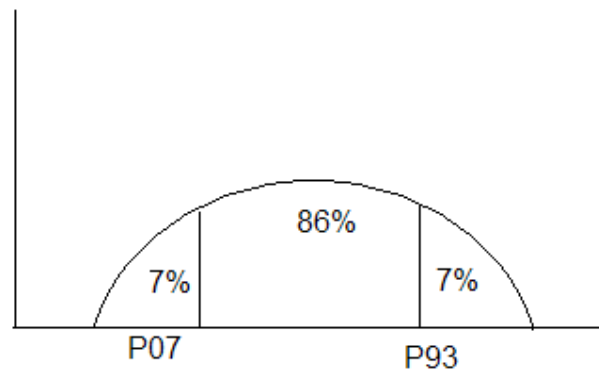
Interprétation : Ce qui signifie que 75% des salariés perçoivent un salaire inférieur à 24300 DA et 25% des salariés perçoivent un salaire supérieur à 24300 DA.

Intervalle Inter quartile : $Q_3 - Q_1 = 24,3 - 17,3 = 7 \cdot 10^3 \text{ DA} = 7000 \text{ DA}$

Interprétation : Pour 50% des effectifs l'écart Maximum de salaire est de 7000 DA.

Remarque :

- Les déciles : valeur du caractère tel que 10 % des observations ont une valeur qui est inférieure à D1 et 90% des observations ont une valeur qui est supérieure à D1.
On appelle 9^{ème} décile D9 la valeur du caractère tel que 90% des observations lui sont inférieures, et 10% des observations lui sont supérieures. L'intervalle inter décile D9 - D1 contient 80% des observations
- Les percentiles : On appelle percentiles P1 la valeur du caractère telle que (1%) des observations ont une valeur inférieure à P1 et 99% ont une valeur supérieure à P1.
- Pour le statisticien KELLY pour supprimer les valeurs aberrantes, il suffit de calculer l'intervalle inter percentile $P_{93} - P_{07}$ qui contient 86% des observations.



Graphique 4-2

III. Mesure de la dispersion autour de la moyenne :

Pour définir une mesure de la dispersion qui prend en compte toutes les données, une solution consiste à calculer les écarts ($x_i - \bar{x}$) entre chaque valeur observée x_i et une valeur centrale (moyenne, médiane ou mode).

Comme il est impossible de prendre connaissance en même temps de tous ces écarts, on peut en faire la synthèse en calculant leur moyenne arithmétique.

Exercice n°1 :

Considérons les notes suivantes en statistique d'un groupe de 6 étudiants :

2 -17 -7 - 18 - 3 - 13

La moyenne des notes est : $\bar{X}=10$. Faut-il conclure alors que ce groupe est homogène ? En d'autres termes, faut-il conclure que les étudiants ont le même niveau ? La réponse est non, car 50% des étudiants seulement ont la moyenne.

Pour mesurer cette dispersion autour de la moyenne on peut calculer les différentes distances (écarts) entre la moyenne et les notes observées. On obtient :

x_i	$x_i - \bar{X}$
2	-8
3	-7
7	-3
13	+3
17	+7
18	+8
Σ	0

$$\bar{x} = \frac{1}{6} \sum_{i=1}^6 x_i = \frac{60}{6} = 10$$

$$\frac{1}{n} \sum (x_i - \bar{x}) = \frac{1}{6} \sum (x_i - \bar{x}) = 0$$

Remarque : On a toujours : $\frac{1}{n} \sum (x_i - \bar{x}) = 0$

Ceci traduit le fait que certains étudiants ont des notes supérieures à la moyenne et d'autres ont des notes qui lui sont inférieures (certaines différences sont positives et d'autres sont négatives).

Une première solution consiste à prendre les valeurs absolues de ces écarts et de calculer leur moyenne.

IV. L'écart absolu moyen par rapport à la moyenne :

L'écart absolu moyen par rapport à la moyenne, noté $\bar{e}_{\bar{x}}$ d'une série statistique est égal à la moyenne arithmétique de la valeur absolue des écarts entre les valeurs observées et leur moyenne.

Cas de données non groupées (DNG) :

$$\bar{e}_{\bar{x}} = \frac{1}{n} \sum_{i=1}^k |x_i - \bar{x}|$$

Cas de données groupées (DG) :

$$\bar{e}_{\bar{x}} = \frac{1}{n} \sum_{i=1}^k n_i |x_i - \bar{x}|$$

Exercice n°2 : Soit le tableau suivant :

Poids en Kg	n_i	x_i	$n_i * x_i$	$ x_i - \bar{x} $	$n_i x_i - \bar{x} $
55-60	12	57,5	690	10,25	123
60-65	17	62,5	1062,50	5,25	89,25
65-70	36	67,5	2430	0,25	9
70-75	24	72,5	1740	4,75	114
75-80	11	77,5	852,50	9,75	107,25
Σ	100		6775		442,5

$$\bar{e}_{\bar{x}} = 442.5 / 100 = 4.42 \text{ Kg}$$

$$\bar{X} = 67.75 \text{ Kg}$$

Interprétation : $\bar{e}_{\bar{x}} = 4.42 \text{ Kg}$ signifie qu'en moyenne, chaque individu s'éloigne de la moyenne (67.75 Kg) de 4.42 Kg.

Remarque : Pour dire si une dispersion est grande ou non, pour comparer deux séries entre elles, on se sert de l'indice de dispersion relatif $= \frac{\bar{e}_{\bar{x}}}{\bar{X}} * 100$

Exemple :

Poids de filles

$$\bar{X}=52 \text{ Kg}$$

$$\bar{e}_x= 2 \text{ Kg}$$

$$2/52 * 100 = 3.8\%$$

Dispersion Faible

Poids des garçons

$$\bar{X}=68 \text{ Kg}$$

$$\bar{e}_x= 17 \text{ Kg}$$

$$17/68 * 100 = 25\%$$

Dispersion plus importante

Remarque : Cet indicateur de dispersion tient compte de tous les écarts entre chaque valeur observée et la moyenne. Ces écarts sont exprimés dans la même unité que la variable. Le calcul de l'écart absolu moyen ne se prête pas aisément au calcul algébrique (n'est pas commode) à cause de la valeur absolue.

IV. Variance et écart-type :

Une solution alternative consiste à considérer la moyenne des écarts des différences (dans ce cas toutes les valeurs négatives deviennent positives).

1. Variance :

On appelle **variance** d'une variable la moyenne arithmétique des carrés des écarts de valeurs de cette variable à sa moyenne :

➤ Cas de données non groupées (DNG) :

Formule de définition (1) et formule développée (ou formule de Kooning) (2) :

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

➤ Cas de données groupées (DG) :

Formule de définition (1) et formule développée (ou formule de Kooning) (2) :

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

Remarque :

- La formule développée de la variance est plus aisée à retenir et plus rapide à calculer.
- La variance est exprimée dans le carré de l'unité de la variable. Par exemple la variance de la variable salaire est exprimée en « dinars au carré (DA²) ». C'est la raison pour laquelle on ne doit pas interpréter la variance, mais plutôt sa racine carrée. On obtient, l'écart-type.

2. Ecart-type :

On appelle **écart-type** (ou écart quadratique moyen) que l'on note par δ_x d'une variable la moyenne arithmétique des carrés des écarts de valeurs de cette variable à sa moyenne :

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

➤ **Cas de données non groupées (DNG) :**

Formule de définition (1) et formule développée (ou formule de Kooning) (2) :

$$\delta(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{N}} \quad (1) \quad \delta(x) = \sqrt{\frac{\sum_{i=1}^n (x_i)^2}{N} - (\bar{X})^2} \quad (2)$$

On demande maintenant de calculer l'écart-type de l'exercice n°1 :

$$V(x) = \frac{1}{4} \sum_{i=1}^4 (x_i - \bar{x})^2 = 40,66 \Rightarrow \delta_x = \sqrt{V(x)} = \sqrt{40,66} = 6,37$$

Interprétation : Certains étudiants (les bons) auront approximativement la note moyenne (10) plus 6,37, les autres (les mauvais) auront la note moyenne (10) moins 6,37.

➤ **Cas de données groupées (DG) :**

$$V(x) = \frac{1}{4} \sum_{i=1}^4 (x_i - \bar{x})^2 = \frac{40}{4} = 10 \Rightarrow \delta_x = \sqrt{V(x)} = 3,16$$

Formule de définition (1) et formule développée (ou formule de Kooning) (2) :

$$\delta(x) = \sqrt{\frac{\sum_{i=1}^n n_i (x_i - \bar{X})^2}{N}} \quad (1) \quad \delta(x) = \sqrt{\frac{\sum_{i=1}^n n_i x_i^2}{N} - (\bar{X})^2} \quad (2)$$

Remarque :

L'écart-type est exprimé dans la même unité de mesure que la variable. Plus l'écart-type est grand, plus la dispersion des observations autour de la moyenne de la variable est forte

Exemple :

Prendre le même tableau de l'exercice n° 2 concernant le poids en Kg de 100 individus :

$(x_i - \bar{X})^2$	$ni \cdot (x_i - \bar{X})^2$
105,0625	1260,75
27,5625	468,5625
0,0625	2,25
22,5625	541,50
95,0625	1045,6875
	3318,75

$$\sigma = \sqrt{\sum ni(x_i - \bar{x})^2 / \sum ni} = \sqrt{3318.75/100} = 5.76$$

Interprétation : En moyenne chaque individu s'écarte du poids moyen (67.5 kg) de 5.76 kg.

Exercice n°3 :

Considérons les notes suivantes en statistique d'un groupe de 4 étudiants :

8 -12 -9 - 11

TAF : Calculer l'écart-type des notes et comparer le résultat obtenu avec le résultat de l'exercice n° 1.

$$\bar{x} = \frac{1}{4} \sum_{i=1}^4 x_i = \frac{40}{4} = 10$$

$$V(x) = \frac{1}{4} \sum_{i=1}^4 (x_i - \bar{x})^2 = \frac{40}{4} = 10 \Rightarrow \sigma_x = \sqrt{V(x)} = \sqrt{10} = 3,16$$

Comparaison : La dispersion des notes dans l'exercice n°1 est deux fois plus importante que celle de l'exercice n°3. Le second groupe d'étudiant est un groupe plus homogène que le 1^{er} groupe.

V. Coefficient de variation :

Pour une variable statistique réelle X , on appelle **coefficient de variation** le rapport :

$$CV = \frac{\delta_x}{\bar{x}} \times 100$$

Remarque : Le coefficient de variation est indépendant des unités de mesures. Il s'exprime en pourcentage. C'est pourquoi il est commode de l'utiliser quand on veut comparer des distributions où les unités sont différentes. Par contre il cesse d'être commode quand $\bar{X}$ est presque nulle.

Exemple :

On prend les données de l'exercice n°2 :

$$\bar{X} = 67.75 \text{ Kg} \quad CV = \frac{\delta_x}{\bar{x}} \times 100 = (5.76/67.75) * 100 = 8.5\%$$

Exemple :

Soient 2 modèles d'ampoules électriques **A** et **B** dont on a relevé les durées de vie en heures (h) :

A : Durée de vie moyenne 1400 h
Ecart-type = 100 h

B : Durée de vie moyenne 1800 h
Ecart-type = 250 h

Modèle A
 $\delta/\bar{X} * 100 = 100/1400 * 100 = 7\%$

Modèle B
 $250/1800 * 100 = 14\%$

Le modèle A est plus faible que le modèle B

VI. Coefficient interquartile (relatif) :

C'est un autre paramètre permet aussi d'écarter dans son calcul l'unité de mesure.

Il s'agit du coefficient interquartile noté C_Q dont la formule est la suivante : $Q_2 = M_e$

Remarque :
$$C_Q = \frac{Q_3 - Q_1}{Q_2} * 100$$

Pour l'étude de séries statistiques simples : Il est nécessaire de séparer clairement deux types de calculs :

- Moyenne, écart-type à réaliser à partir des centres de classes et des effectifs de classes.
- Médiane, quartiles, intervalles interquartile...à réalisés à partir des extrémités de classes et des effectifs cumulés.

CHAPITRE V : LES CARACTERISTIQUES DE FORME

Les polygones des fréquences nous livrent une représentation approximative de la distribution réelle des fréquences. Pour avoir une idée satisfaisante et plus précise sur la forme de la distribution, il est recommandé de calculer des indicateurs de forme.

On distingue les indicateurs d'asymétrie et les indicateurs d'aplatissement. Ces indicateurs sont sans unité de mesure. Ils sont indépendants d'un changement d'échelle et/ou de mesure.

Pour étudier la forme de la courbe d'une distribution, il est indispensable d'utiliser les moments d'une série statistique. Pour cela un rappel s'avère nécessaire.

I. Le Moments d'une série statistique :

1. Moments non centrés :

1.2 Cas de données non groupées (DNG) :

Le moment non centré d'ordre q , qu'on note m_q , d'une série statistique est :

$$m_q = \frac{1}{N} \sum x_i^q$$

1.2 Cas de données groupées (DG) :

$$m_q = \frac{1}{\sum n_i} \sum n_i x_i^q = \sum f_i x_i^q$$

Remarque :

- Le moment non centré d'ordre 1 est : $m_1 = \bar{x}$
- Le moment non centré d'ordre 2 est : $m_2 = \overline{x^2}$

2. Moments non centrés :

2.1 Cas de données non groupées (DNG) :

Le moment non centré d'ordre q , qu'on note μ_q , d'une série statistique est :

$$\mu_q = \frac{1}{N} \sum (x_i - \bar{x})^q$$

2.2 Cas de données groupées (DG) :

$$\mu_q = \frac{1}{\sum n_i} \sum n_i (x_i - \bar{x})^q = \sum f_i (x_i - \bar{x})^q$$

Remarque :

- Le moment centré d'ordre 1 est : $\mu_1 = 0$
- Le moment non centré d'ordre 2 est : $\mu_2 = V(x)$
- A partir de la formule développée de la variance, on a :

$$\mu_2 = V(x) = m_2 - (m_1)^2$$

- En général, les moments centrés d'ordre pair donnent une indication sur la dispersion des observations autour de la moyenne. Les moments centrés d'ordre impair donnent une indication sur le degré de symétrie de la distribution.

Cette étude se fait à l'aide de deux types de coefficients :

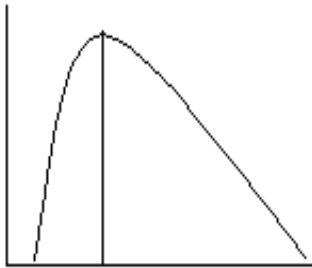
- Les coefficients d'asymétrie.
- Les coefficients d'aplatissement.

II. Les coefficients d'asymétrie :

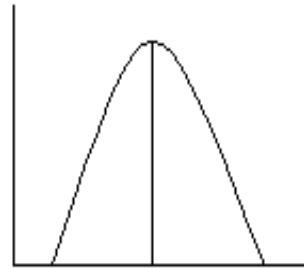
Pour une distribution parfaitement symétrique on a l'égalité entre les trois valeurs de tendance centrale à savoir le mode, la médiane et la moyenne arithmétique :

$$Me = Mo = \bar{x}$$

Autrement dit, une distribution est dite symétrique si les observations se répartissent dans la même proportion de part et d'autre des trois valeurs centrales (mode, médiane et moyenne) : Graphique 5-2.



**Distribution
Asymétrique
Graphique 5-1**



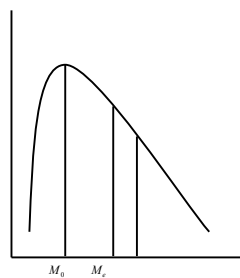
**Distribution
Symétrique
Graphique 5-2**

La symétrie est par rapport au mode qui correspond au plus grand effectif.

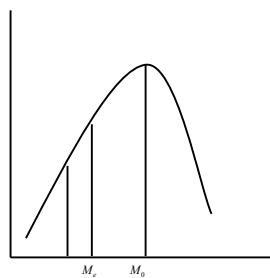
$$Me = Mo = \bar{x}$$

Cependant, dès que cette relation n'est pas vérifiée on aura une distribution asymétrique.

Les caractéristiques d'asymétrie permettent de quantifier le degré de déviation de la forme de distribution par rapport à une distribution symétrique.



**Asymétrie à gauche
Graphique 5-3**



**Asymétrie à droite
Graphique 5-4**

Il existe dans la littérature plusieurs coefficients pour mesurer l'asymétrie d'une distribution. Dans la suite on se limitera à la présentation de deux coefficients :

- a) Le coefficient empirique de Pearson :** basé sur la moyenne, le mode et l'écart-type. C'est un nombre sans unité défini par :

$$C_p = \frac{\bar{x} - M_o}{\delta}$$

- Si $C_p = 0$ alors la distribution est symétrique : **Graphique 5-2.**
- Si $C_p > 0$ alors la distribution est dite étalée à droite (Asymétrie à droite) : **Graphique 5-4.**
- Si $C_p < 0$ alors la distribution est étalée à gauche (Asymétrie à gauche) : **Graphique 5-3.**

- b) Le coefficient d'asymétrie de Yule :** basé sur les quartiles qu'on note :

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

- La distribution est dite symétrique dans le cas où $C_y = 0$.
- La distribution est dite étalée à droite (Asymétrie à droite) dans le cas où $C_y > 0$.
- La distribution est dite étalée à gauche (Asymétrie à gauche) dans le cas où $C_y < 0$.

- c) Le coefficient d'asymétrie de Pearson :**

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Pour étudier l'asymétrie on doit connaître le sens de l'asymétrie. Pour cela on doit calculer le coefficient de Fisher :

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}}$$

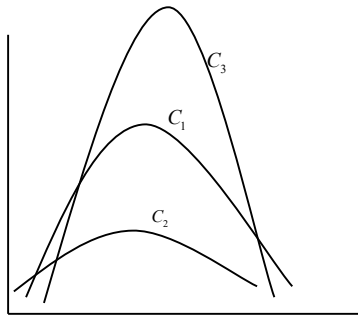
- La distribution est dite symétrique dans le cas où $\gamma = 0$. Dans ce cas on a : $\beta_1 = \gamma_1 = 0$.
- La distribution est dite étalée à droite (Asymétrie à droite) dans le cas où $\gamma < 0$.
- La distribution est dite étalée à gauche (Asymétrie à gauche) dans le cas où $\gamma > 0$.

III. Les coefficients d'aplatissement :

Le coefficient d'aplatissement permet de comparer la distribution étudiée par rapport à celle de la loi normale qui est une distribution symétrique ayant une allure en cloche.

On distingue trois formes de distributions :

- Distribution normale ou mésokurtique : **courbe C1**.
- Distribution aigue ou leptocurtique (moins aplatie que la distribution normale) : **courbe C3**.
- Distribution plate ou platycurtique (plus aplatie que la distribution normale) : **courbe C2**.



Graphique 5-5

Il existe dans la littérature plusieurs coefficients pour mesurer l'aplatissement d'une distribution. Dans la suite on se limitera à la présentation de deux coefficients :

a) Le coefficient d'aplatissement de Pearson :

Le coefficient d'aplatissement de Pearson noté β est un nombre sans unité défini par :

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\text{Moment centré d'ordre 4}}{(\text{Ecart-type})^2}$$

- La distribution est dite normale dans le cas où $\beta = 3$.
- La distribution est dite leptokurtique ou aigue (moins aplatie que la normale) dans le cas où $\beta > 3$.

- La distribution est dite platykurtique ou plate (plus aplatie que la normale) dans le cas où $\beta < 3$.

Remarque : Ce coefficient est égal à 3 pour une distribution normale, et varie de 1 jusqu'à ∞ .

b) Le coefficient d'aplatissement de Fisher : qu'on note par :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \beta - 3$$

- La distribution est dite normale dans le cas où $\gamma_2 = 0$.
- La distribution est dite leptokurtique ou aigue (moins aplatie que la normale) dans le cas où $\gamma_2 > 0$.
- La distribution est dite platykurtique ou plate (plus aplatie que la normale) dans le cas où $\gamma_2 < 0$.

Exercice :

Soit le tableau suivant relatif à la distribution des salaires mensuels en 10^3 DA des ouvriers dans une certaine entreprise :

Salaires 10^3 DA	Effectifs n_i
[150 – 160 [	11
[160 – 170 [	26
[170 – 185 [	63
[185 – 190 [	81
[190 – 200 [	35
[200 – 205 [	21
[205 – 210 [	13
Total	250

Travail à faire :

Calculer :

1. Le coefficient d'asymétrie de Fisher. Interpréter.
2. Le coefficient d'aplatissement de Fisher. Interpréter.

Solution :

Après calculs on obtient :

$$\bar{x} = 184.46 \quad ; \quad \sigma = 12.367$$

$$n_i(x_i - \bar{x})^3 = -180791.592 \quad ; \quad n_i(x_i - \bar{x})^4 = 16598215.89$$

1. Le coefficient d'asymétrie de Fisher est défini par :

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} = \frac{-180791.592}{\frac{250}{(12.367)^3}} = -0.382 < 0$$

Interprétation :

$\gamma_1 < 0$ négatif $\Rightarrow$ La distribution est étalée à gauche (Asymétrie à gauche).

2. Le coefficient d'aplatissement de Fisher est défini par :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \beta - 3 = \frac{16598215.89}{\frac{250}{(12.367)^4}} - 3 = -0.162$$

Interprétation :

$\gamma_2 < 0$ négatif $\Rightarrow$ Il s'agit d'une distribution **platykurtique** ou plate (plus aplatie que la normale).

CHAPITRE VI : LES CARACTERISTIQUES DE CONCENTRATION

L'étude de la concentration a pour objet de mesurer et de mettre en exergue d'éventuelles inégalités de répartition d'une valeur globale totale. Cette étude n'est pas centrée sur l'individu, elle est plutôt globale. L'analyse porte davantage sur la répartition de la masse totale.

Elle permet de compléter l'analyse de la dispersion relative d'une distribution et de caractériser la structure d'un grand nombre d'entités économiques dont les modalités du caractère sont susceptibles d'être additionnées. Les domaines d'application sont nombreux : concentration des salaires, des revenus, chiffres d'affaires, des superficies agricoles cultivables...etc.

On parle ainsi de concentration à chaque fois qu'on peut remarquer une disparité ou une inégalité entre les unités de la population considérée. Ainsi, la concentration ne s'applique qu'à des séries statistiques ou la concentration de la variable à un sens.

Remarque :

On ne peut pas parler de concentration d'âge.

Le concept de concentration a été élaboré dans les années 1910-14 par le statisticien italien Corrado Gini (1884-1965).

L'étude de la concentration porte sur toute série positive. La notion de concentration ne s'applique qu'à des variables quantitatives continues à valeurs positives cumulables, celle où le cumul a un sens.

La question fondamentale à laquelle on doit répondre est par exemple : La masse salariale totale est-elle répartie d'une manière égalitaire ? Dans le cas où elle s'est faite d'une manière inégalitaire, on observe un faible nombre d'individus détenir une grande partie de cette masse, la partie restante étant détenue par un grand nombre d'individus.

On peut déterminer la concentration soit algébriquement soit graphiquement.

I. La détermination algébrique de la concentration :

Cette détermination nécessite la connaissance de la « médiale ».

1. Notion de médiale (Mle) :

Si dans une série on désigne par x_i la valeur du caractère, par n_i les effectifs, la médiale est la valeur du caractère qui partage en deux parties égales le produit cumulé de $n_i x_i$.

Si x_i désigne un salaire et n_i désigne le nombre de salariés.

Le produit cumulé des $n_i x_i$ représente la totalité des salaires versés $\sum n_i x_i$, c'est-à-dire la masse salariale.

La médiale, c'est le salaire tel que la moitié de la masse salariale a servi à payer une partie qui touche moins de cette Médiale et l'autre moitié de la masse s a servi à payer les gens qui touchent plus de cette Médiale.

2. Mesure de la concentration :

ΔM sert à mesurer la différence entre Mle et Me :

$$\Delta M = Mle - Me$$

* Si $\Delta M = 0$ cela veut dire que $Mle = Me$

C'est-à-dire l'individu qui est au milieu l'effectif est en même temps celui qui est placé tel que la moitié de la masse salariale a été versée à des gens qui touchent moins que lui, et l'autre moitié à des gens qui reçoivent plus que lui, on a donc une distribution égalitaire concentration est nulle.

* Si $\Delta m \neq 0$ cela indique qu'il y a une concentration

* Si Δm est faible par rapport à l'intervalle de variation la concentration est faible

* Si $\frac{\Delta m}{\text{Inter variation}}$ est important, la concentration est forte

Application :

Salaires 10 ³ DA	Effectifs n_i	Centre de classes x_i	$n_i x_i$	$(n_i x_i) \nearrow$
[10-15]	9	12.5	112.5	112.5
[15-20]	25	17.5	437.5	550
[20-25]	32	22.5	720	1270
[25-30]	16	27.5	440	1710
Total	82		1710	

$$\Delta M = Mle - Me$$

Calcul de la Mle :

Rang de la médiale = $RMle = 1710/2 = 855 \Rightarrow$ Classe médiale : [20.25]

Interpolation linéaire :

$$720 \rightarrow 5 \text{ DA}$$

$$1 \text{ DA} \rightarrow 5/720 \text{ DA}$$

$$(855 - 550) = 305 \rightarrow 5/720 * 305 \text{ DA}$$

$$\text{Donc } Mle = 20 + 5/720 * 350$$

$$Mle = 22.12 \text{ DA}$$

}

$\longleftrightarrow$

$$\Delta M = ML - ME$$

$$= 22,12 - 21,09 \approx 1 \text{ DA}$$

$$\Delta M / \text{inter variation} = 1/20 = 5\% \Rightarrow \text{concentration faible}$$

L'intervalle de variation «Étant égale à : $(30 - 10) = 20$

Signification $Mle = 22.12 \text{ DA} = 22.12 * 10^3 = 22120 \text{ DA}$:

C'est le salaire tel que la moitié de la masse salariale a servi à payer des gens qui gagnent moins que 22120 DA et l'autre moitié de la masse salariale a servi à payer les gens qui gagnent plus que 22120 DA.

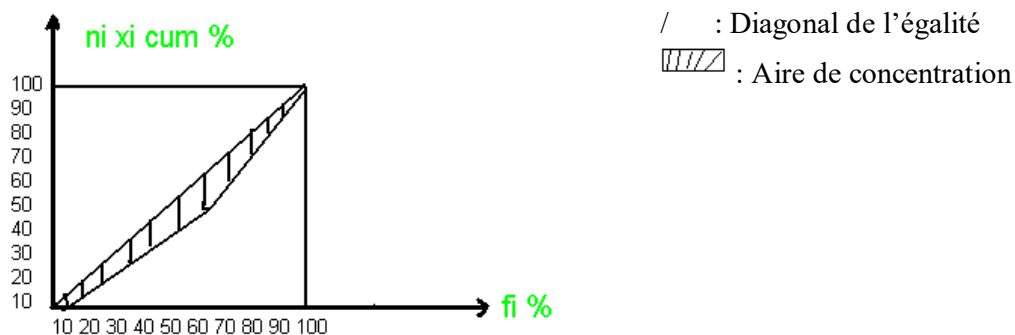
II. La détermination graphique de la concentration :

La courbe de Lorenz et le coefficient de Gini

1. La courbe de concentration (ou courbe de Lorenz) :

Lorenz de mesurer la concentration en mettant en abscisses les fréquences cumulées en%, et en ordonnées $n_i x_i$ cumulés en % :

Salaires 10 ³ DA	Effectifs n_i	$f_i\%$	$f_i\% \nearrow$	$n_i x_i$	$(n_i x_i)$	$n_i x_i\%$	$n_i x_i\% Cum$
[10-15[	9	11	11	112.5	112.5	6.5	6.5
[15-20[	25	30.5	41.5	437.5	550	25.5	32
[20-25[	32	39	80.5	720	1270	42	74
[25-30[	16	19.5	100	440	1710	26	100
Total	82	100		1710		100	



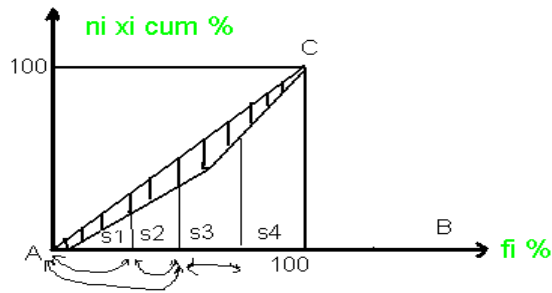
Graphique 6-1

Remarque :

- Si 10% de la population touchent 10% du revenu, 20% de la population touchent 20% du revenu. Dans le cas d'une répartition égalitaire du salaire, l'aire de concentration serait confondue avec diagonale.
- Dans le cas d'une répartition inégalitaire parfaite des salaires, (comme dans le cas théorique où 0.1% de la population toucherait 99.99% de la masse salariale : la courbe.

2. L'indice de Gini :

Gini propose de calculer la concentration à l'aide de coefficient suivant :



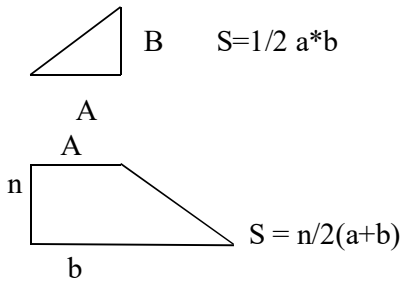
Graphique 6-2

$$C = \frac{\text{Aire de concentration}}{\text{Aire du triangle ABC}}$$

$$C = \frac{\text{Aire de G}}{5000(100 \cdot 100 / 2)}$$

On peut estimer l'aire de concentration de la manière suivante :

$$\text{Aire de concentration} = 5000 - (S1 + S2 + S3 + S4)$$



$$\begin{aligned} S1 &= \frac{1}{2}(116.6) \\ S2 &= (41.5 - 11) / 2(6.6 + 32.2) \\ S3 &= (80.5 - 41.5) / 2(32.2 + 74.3) \\ S4 &= (100 - 80.5) / 2(74.3 + 100) \\ \sum Si &= 4404 \end{aligned}$$

Graphique 6-3

Remarque :

- $0 < C < 1$
- $C = 0$ Concentration élevée $\Rightarrow$ Egalité parfaite
- $C = 1$ Concentration faible $\Rightarrow$ L'inégalité n'est pas très grande.

Donc : $C = 5000 - 4404 / 5000 \approx 0.12$

Interprétation :

L'inégalité n'est pas très grande : Elle est faible.

CHAPITRE VII : CORRELATION ET AJUSTEMENT LINEAIRE

Dans le cadre de ce chapitre, on s'intéresse à l'étude d'une éventuelle relation entre deux variables statistiques. Autrement dit, nous allons d'abord voir comment déterminer le sens de la liaison entre ces deux variables, ensuite, comment mesurer l'intensité ou le degré de la liaison entre elles, et enfin fournir une expression mathématique de la liaison entre ces deux variables.

I. Ajustement linéaire d'un nuage de points :

On considère deux variables statistiques quantitatives x et y et on s'intéresse à une relation éventuelle entre elles.

La représentation du nuage de points peut nous renseigner sur l'allure de la distribution à deux caractères. La forme de la relation entre les deux variables peut être mise en évidence graphiquement par les courbes de régression.

Généralement, on exprime y en fonction de x , on parle alors de la droite de régression de y sur x (ou de y en x). Dans ce cas, on cherche à expliquer la variable y par la variable x . De ce fait, y est dite variable explicative ou variable exogène.

1. La droite de régression de y sur x :

On considère N observations sur les deux variables x et y :

$$\begin{pmatrix} x_1, x_2, x_3, \dots, x_n \\ y_1, y_2, y_3, \dots, y_n \end{pmatrix}$$

Ces observations peuvent être représentées par un nuage de points. D'une manière générale, l'ajustement d'un nuage de point par une fonction mathématique, revient à estimer les valeurs des coefficients de cette fonction, revient à estimer les valeurs des coefficients de cette fonction de telle sorte que sa courbe représentative se rapproche au mieux du nuage de points.

Lorsqu'il s'agit d'une liaison linéaire entre les deux variables, on parle alors d'ajustement linéaire. L'ajustement linéaire consiste à estimer les coefficients de la droite de régression du type :

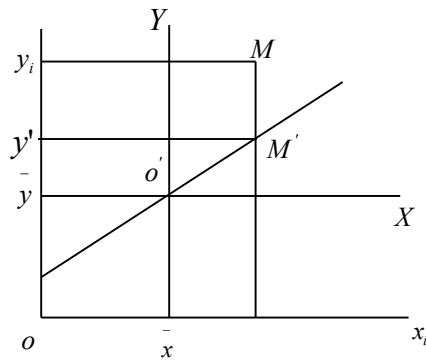
$$D : y/x \quad y_i = ax_i + b$$

C'est-à-dire à trouver la valeur de a et celle de b . Cette droite est supposée refléter l'évolution moyenne de la variable y (variable expliquée) en fonction de la variable explicative x .

La méthode d'ajustement que nous allons exposer est appelée « Méthode des **Moindres Carrés Ordinaires** » ou simplement « **MCO** ».

1.1 La méthode des moindres carrés :

Partons d'un nuage statistique théorique :



Graphique 7-1

$$D : y/x \quad y_i = ax_i + b$$

Il s'agit de résumer ce nuage par une droite.

Soit $y' = ax + b$ l'équation de la droite recherchée.

Pour toute valeur de x (x_i) nous avons une valeur réellement observée y' .

Pour toute valeur x_i , nous avons une valeur calculée sur la droite y' .

Pour toute une valeur x_i , nous avons une erreur d'estimation égale à $|y_i - y'_i|$.

La droite de régression idéale doit être de telle manière que la somme des erreurs d'estimation doit être la plus faible possible, $\sum |y_i - y'_i|$ doit être minimum.

Pour éviter les valeurs absolues, on convient de calculer les carrés des erreurs. La droite de régression doit être telle que :

$$\sum (y_i - y'_i)^2 \text{ minimum, et on appelle cela la condition des moindres carrés.}$$

1.2 Calcul des paramètres de la droite de régression :

Il s'agit de trouver $y' = ax + b$ sachant que : $\sum (y_i - y'_i)^2 \text{ min.}$

Remplaçons y'_i par sa valeur $\rightarrow \sum (y_i - (ax_i + b))^2 \text{ min.}$

Posons $\sum (y_i - ax_i - b)^2 = Z(a, b)$.

Pour que Z soit minimum, il suffit d'annuler (rendre nul) les dérivées de ce polynôme par rapport à 'a' et par rapport à 'b'.

Calcul de b :

Supposons 'a' est connu, et dérivons par rapport a 'b' et 'a'.

$$\delta Z / \delta b = 2 [\sum (y_i - ax ; -b)] (-1) = 0$$

$$\begin{aligned} Z &= U^2 \\ Z' &= 2UU' \end{aligned}$$

$$\sum [y_i - ax ; -b) = 0$$

$$\sum y_i - a \sum x_i - nb = 0$$

Divisons par n, on obtient $(\sum y_i / n - a \sum x_i / n - b = 0$

$$U' = (y_i - ax ; -b)$$

$$\tilde{y} - a \bar{x} = b$$

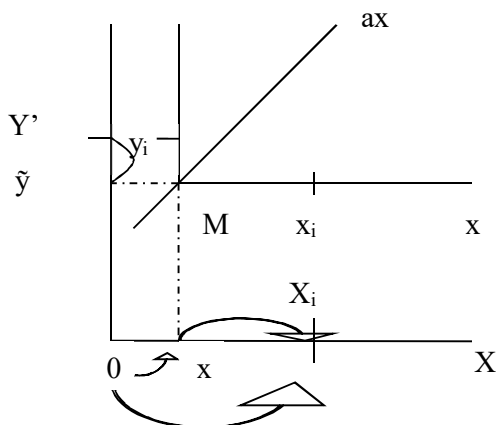
Donc :

$$b = \tilde{y} - a \bar{x}$$

La droite de régression passe donc par le point moyen $(\bar{x}, \tilde{y})$.

Calcul de a :

Y



Graphique 7-2

Le paramètre a Que nous cherchons correspond à la pente de la droite de régression qui passe par le point moyen M $(\bar{x} ; \tilde{y})$.

Procédons un changement d'origine, et prenons comme nouvelle origine le point moyen M $(\bar{x} ; \tilde{y})$, les nouvelles cordonnées deviennent :

$$\begin{cases} X_i = x_i - \bar{x} \\ Y_i = y_i - \tilde{y} \end{cases}$$

La droite de régression a pour équation $y' = ax$

La condition des moindres carrée s'écrit :

$$\sum (y_i - \tilde{y}_i)^2 \min$$

$$\sum (y_i - y'_i)^2 = \sum (y_i - ax_i)^2 \min$$

Dérivons par rapport à 'a' : $2 [\sum (y_i - ax_i)] (-X_i) = 0$

$$[\sum (y_i - ax_i)] X_i = 0 \Rightarrow \sum (y_i - a_i) X_i = 0 \Rightarrow \sum x_i y_i - a \sum x_i^2 = 0$$

$$\text{Donc } a = \sum x_i y_i / \sum x_i^2 = \sum (x_i - \bar{x}) (y_i - \bar{y}) / \sum (x_i - \bar{x})^2$$

L'équation de la droite de régression :

$$Dy(x) =$$

$$Y = ax + b$$

$$\begin{cases} a = \sum (x_i - \bar{x}) (y_i - \bar{y}) / \sum (x_i - \bar{x})^2 \\ b = \bar{y} - a \bar{x} \end{cases}$$

Formule de définition (1) :

$$a = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} (1)$$

Formule développée ou relation de Koning :

En utilisant la formule de définition :

$$a = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} (1)$$

On développe le numérateur :

$$\begin{aligned} \sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + n \bar{x} \bar{y} \\ &= \sum x_i y_i - n \bar{y} \bar{x} - n \bar{x} \bar{y} + n \bar{x} \bar{y} \\ &= \sum x_i y_i - n \bar{x} \bar{y} \end{aligned}$$

On développe le dénominateur :

$$\begin{aligned}
 \sum (x_i - \bar{x})^2 &= \sum (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\
 &= \sum x_i^2 - 2\bar{x} \sum x_i + n\bar{x}^2 \\
 &= \sum x_i^2 - 2n\bar{x} + n\bar{x}^2 \\
 &= \sum x_i^2 - n\bar{x}^2
 \end{aligned}$$

On obtient ainsi : La formule développée (relation de Koenig) :

$$a = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} \quad (2)$$

En divisant le numérateur et le dénominateur sur n on obtient :

$$\begin{aligned}
 a &= \frac{\frac{\sum x_i y_i}{n} - \bar{x} \bar{y}}{\frac{\sum x_i^2}{n} - \bar{x}^2} \\
 a &= \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} \\
 a &= \frac{\text{Cov}(x, y)}{V(x)} = \frac{\text{Cov}(x, y)}{\delta_x^2} \quad (3)
 \end{aligned}$$

$\text{Cov}(x, y)$ = Covariance entre x et y (Variance commune entre x et y).

= Variance de la variable x. $V(x) = \delta_x^2$

2. La droite de régression de x sur y :

On peut exprimer x en fonction de y. Dans ce cas, on appelle x une variable endogène ou expliquée et y une variable exogène ou explicative, et on parle de la droite de régression de x sur y :

$$D : x/y \quad x_i = a' y_i + b'$$

En utilisant la méthode des moindres carrés ordinaires, on retrouve la valeur de a' et de b' exprimées par :

Formule de définition (1) :

$$a' = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} \quad (1)$$

On obtient ainsi : La formule développée (relation de Konig) :

$$a' = \frac{\sum x_i y_i - N \bar{x} \bar{y}}{\sum y_i^2 - N \bar{y}^2} \quad (2)$$

$$a' = \frac{\text{Cov}(x, y)}{V(y)} = \frac{\text{Cov}(x, y)}{\sigma_y^2} \quad (3)$$

$\text{Cov}(x, y)$ = Covariance entre x et y (Variance commune entre x et y).

= Variance de la variable y. $V(Y) = \sigma_y^2$

Calcul de b' da la même manière :

On obtient :

$$b' = \bar{x} - a' \bar{y}$$

Remarque :

L'étude de la droite de régression de x sur y permet de prévoir x en fonction de y :

$$D : x/y \quad x_i = a' y_i + b'$$

II. La covariance entre x et y :

La covariance est égale à la moyenne des écarts des couples (x_i, y_i) de X et Y par rapport au point $(\bar{x}, \bar{y})$.

$$\text{Cov}(x, y) = \frac{\sum x_i y_i}{N} - (\bar{x} \cdot \bar{y})$$

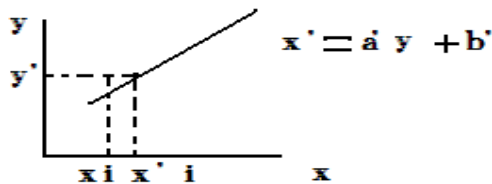
La covariance indique le sens de la relation entre les variables X et Y. Ainsi, on peut distinguer les cas suivants :

- Si $\text{Cov}(x,y) > 0$: alors on peut dire que la relation entre les deux variables est positive. Dans ce cas, ces deux variables varient dans le même sens.
- Si $\text{Cov}(x,y) < 0$: alors on peut dire que la relation entre les deux variables est négative. Dans ce cas, ces deux variables varient en sens inverse.
- Si $\text{Cov}(x,y) = 0$: alors on peut dire qu'il n'y a pas de relation entre les deux variables est positive. Dans ce cas, les variations de l'une n'entraînent pas la variation de l'autre.

III. Le coefficient de corrélation linéaire entre x et y :

Dans le paragraphe précédent, nous avons estimé y en fonction de x, et nous avons obtenu la droite de régression $Dy(x)$.

On peut pour le même nuage statistique estimer x en fonction de y, et trouver la droite de régression $Dx(y)$ lui aura pour équation.



Graphique 7-3

Pour toute y_i , nous avons une valeur observée x_i .

Pour toute y_i , nous avons une valeur estimée sur la droite x'_i

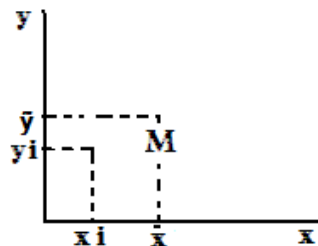
Pour toute y_i , nous avons une erreur d'estimation égale à $|x_i - x'_i|$

$Dx(y)$ idéale est tel que : $\sum |x_i - x'_i|$ minimum ou encore $\sum (x_i - x'_i)^2$ minimum.

En procédant de la même manière que dans le paragraphe précédent, on trouve l'équation de $Dx(y)$.

$$X = a'y + b'$$

$$\begin{cases} a' = \frac{\sum x_i y_i}{\sum y_i^2} \\ b' = \bar{x} - a' \bar{y} \end{cases}$$



Graphique 7-4

Dans le référentiel XMY nous obtenons 2 droites :

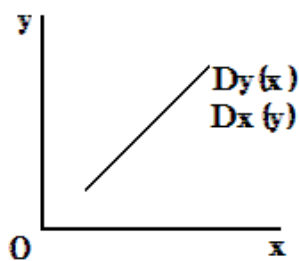
Soit $y = ax$ pour $Dy(x)$

Soit $x = a'y$ pour $Dx(y)$

Ou encore $y = 1/a' x$

Quatre cas peuvent se produire :

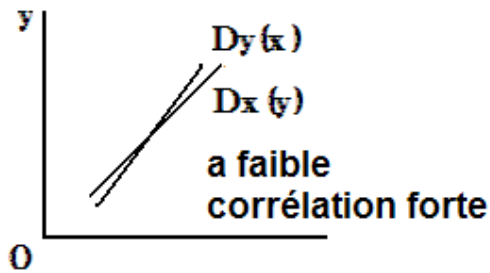
1^{er} cas : Les deux droites sont confondues :



Graphique 7-5

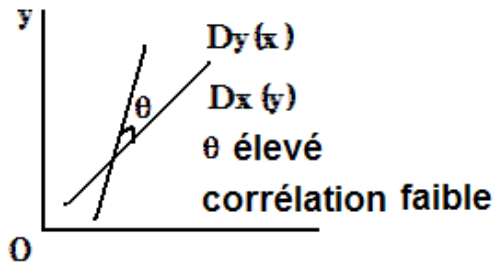
$$\begin{cases} Y = ax \\ X = a'y \\ Y = 1/y' \end{cases} \Leftrightarrow a = 1/a' \Leftrightarrow aa' = 1$$

2ème cas : Les deux droites font entre elles un angle très faible :



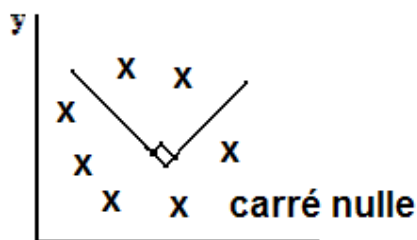
Graphique 7-6

3ème cas : Les deux droites font entre elles un angle très élevé :



Graphique 7-7

4ème cas : Les deux variables sont indépendantes l'une de l'autre :



Graphique 7-8

Le coefficient de corrélation linéaire est un nombre sans dimension qui permet de mesurer le degré ou l'intensité de la liaison linéaire entre deux variables statistiques. Ainsi, la formule du coefficient de corrélation linéaire entre X et Y est :

$$r_{x,y} = \frac{cov(x,y)}{\delta_x \delta_y} = \frac{cov(x,y)}{\sqrt{V(x)} \cdot \sqrt{V(y)}}$$

La covariance indique le sens de la relation entre les variables X et Y. Ainsi, on peut distinguer les cas suivants :

- Si $\mathbf{r_{x,y}} > 0$: les deux variables varient dans le même sens.
- Si $\mathbf{r_{x,y}} < 0$: les deux variables varient en sens inverse.
- Si $\mathbf{r_{x,y}} = 0$: les deux variables sont linéairement indépendants.

$$-1 < r_{x,y} < +1$$

Le coefficient de corrélation linéaire $r_{x,y}$ est toujours compris entre -1 et +1, son signe indique le sens de la liaison, tandis que sa valeur absolue en indique l'intensité.

Interprétation de la valeur de $r_{x,y}$:

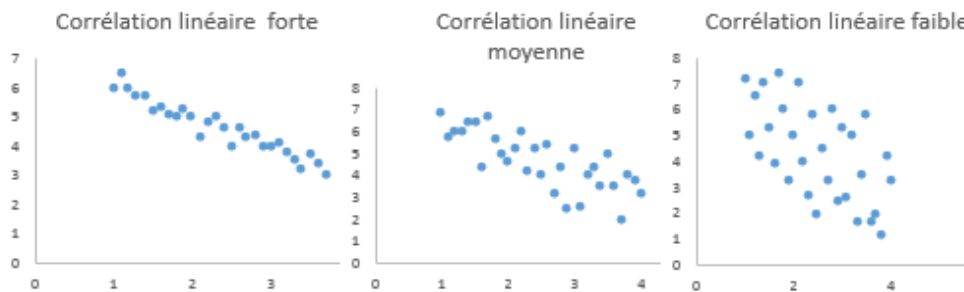
- Si $r_{x,y} = 1$: on dit qu'il y a une parfaite corrélation linéaire positive entre les deux variables.
- Si $r_{x,y} = -1$: on dit qu'il y a une parfaite corrélation linéaire négative entre les deux variables.
- Si $0 < r < 1$: la corrélation est positive, elle est d'autant plus forte que l'on se rapproche de 1.
- Si $-1 < r < 0$: la corrélation est négative, et elle est d'autant plus forte que l'on se rapproche de -1.
- Si $r_{x,y} = 0$: la corrélation est nulle, on dit qu'il y a absence de corrélation linéaire entre les deux variables.
- On dit qu'il y a une forte corrélation linéaire entre les deux variables (ou forte dépendance linéaire) si $\mathbf{r_{x,y}}$ est proche de ± 1 . En revanche, si r est proche de zéro, on dit qu'il y a une faible corrélation linéaire entre les deux variables.

Généralement, les valeurs suivantes sont utilisées pour qualifier l'intensité de la liaison :

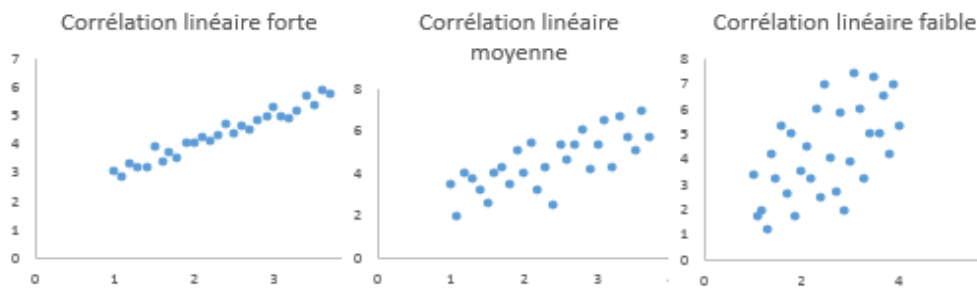
Valeur de r (Coefficient de corrélation)	Force du lien linéaire ou intensité de la liaison
Près de 0	Nulle
$0 < r \leq 0.2$	Faible
$0.2 < r \leq 0.5$	Moyenne
$0.5 < r \leq 0.8$	Forte
$r > 0.8$	Très forte
1	parfaite

Afin de bien voir la différence entre chacun des qualificatifs de corrélation, voici des figures qui les représentent.

Corrélation linéaire négative



Corrélation linéaire positive



Graphique 7-9: Les types de corrélation linéaire entre deux variables

Remarque :

- Le coefficient de corrélation nous donne des informations sur l'existence d'une relation linéaire (sous forme d'une droite) entre les deux grandeurs considérées. Un coefficient de corrélation nul ne signifie pas l'absence de toute relation entre les deux grandeurs. Il peut exister une relation *non linéaire* entre elles. La connaissance de X nous donne des informations sur la valeur de Y.

- Il ne faut pas confondre *corrélation* et *relation causale*. Une bonne corrélation entre deux grandeurs peut révéler une relation de cause à effet entre elles, mais pas nécessairement.
- Lorsqu'on écrit : $r = \sqrt{aa'}$ r = racine a .a', nous avons une expression très positif. Comment trouver alors le signe d'une corrélation ?

Réponse : le sens de la corrélation est donnée par le signe de a et a' :

- Si a et a' sont >0 $\Rightarrow$ le produit a.a' >0 $\Rightarrow$ corrélation positive.
- Si a et a' sont <0 $\Rightarrow$ le produit a.a' <0 $\Rightarrow$ corrélation négative.

- On peut dire d'une corrélation qu'elle est très satisfaisante à partir 0.86.
- On peut dire d'une corrélation qu'elle parfaite à partir de 0.96.

IV. Le coefficient de détermination entre x et y :

D'un autre côté afin de juger la qualité de l'ajustement linéaire, on se base sur le coefficient de détermination qui est défini comme le carré du coefficient de corrélation de x et y :

$$r_{x,y}^2 = a.a'$$

$$r_{x,y}^2 = \frac{Cov(x, y)}{V(x).V(y)}$$

Le r^2 (r^2 %) est le pourcentage de variation totale de Y s'expliquant par la liaison de Y par rapport X. Il s'agit du coefficient de détermination.

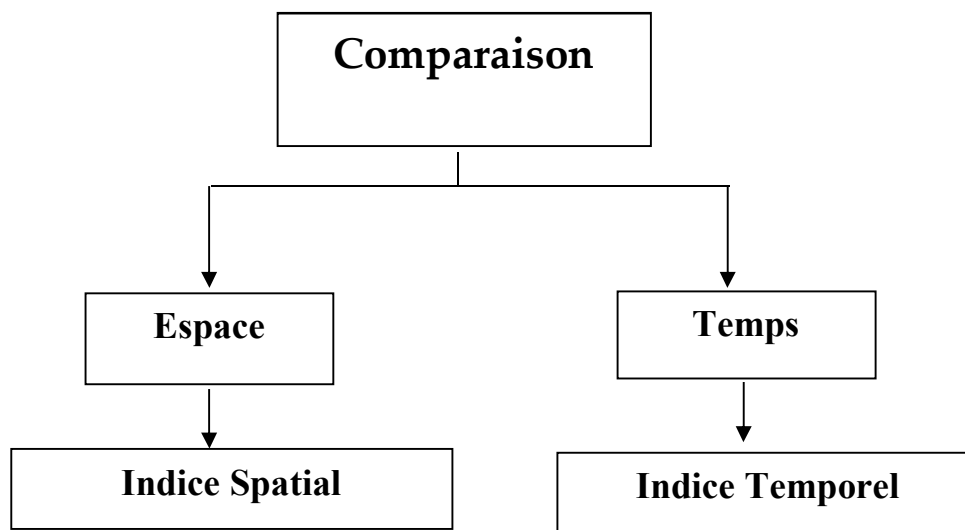
Ce coefficient de détermination est compris entre 0 et 1 :

$$0 \leq r_{x,y}^2 \leq 1$$

- Plus r^2 est proche de 1, plus la qualité de l'ajustement linéaire est bonne.
- Plus r^2 est proche de 0, plus la qualité de l'ajustement linéaire est mauvaise.

CHAPITRE VIII : LES INDICES STATISTIQUES

L'analyse de l'évolution de certaines grandeurs (économiques, sociales, démographiques...) nécessite le recours aux indices. Ces derniers peuvent s'intéresser à l'évolution d'une grandeur simple (grandeur homogène) on parle alors d'indices élémentaires, ou à l'évolution d'une grandeur complexe (grandeur hétérogène ou composite) on a recours dans ce cas aux indices synthétiques.



Les indices statistiques servent également à formuler des prévisions relatives aux différentes activités économiques à venir. Ils trouvent leur application dans plusieurs domaines, notamment :

1. L'analyse de l'évolution des prix d'une période à une autre, afin d'en comprendre les déterminants et de dégager des solutions adaptées aux attentes de la population.
2. L'évaluation des variations du niveau de vie, permettant ainsi de mieux appréhender et encadrer les revendications sociales.
3. La mesure des changements affectant le taux de chômage ainsi que la population active dans une région donnée.

I. Les indices élémentaires :

L'indice élémentaire est un nombre sans unité résultant du rapport de deux valeurs prises par la même grandeur soit à deux dates différentes.

Soit x_t la valeur de la grandeur A à la date t (date courante) et soit x_0 la valeur de la grandeur A à la date 0 (date de base ou de référence). L'indice élémentaire de la grandeur A entre les dates t et 0 sera alors :

$$I_{t/0} = \frac{x_t}{x_0}$$

Cet indice est souvent multiplié par 100 et on dit que cette valeur est à la base 100 en $t = 0$:

$$I_{t/0} = \frac{x_t}{x_0} \times 100$$

Exemple 1 :

La production d'un engrais dans un pays est passée de 210,5 mille tonnes en 2024 à 225,5 mille tonnes en 2025. Ainsi l'indice de la production d'engrais entre 2024 et 2025 sera :

$$I_{2025/2024} = \frac{225,5}{210,5} \cdot 100 = 107,1$$

Ainsi la production de cet engrais a augmenté de 7,1% entre 2024 et 2025.

Exemple 2 :

Le prix d'un même produit X est de 500 DA dans le supermarché et il est de 600 DA dans l'épicerie du quartier. Ainsi l'indice du prix de ce produit entre le supermarché et l'épicerie est de :

$$I_{\text{Supermarché/Epicerie}} = \frac{600}{500} \cdot 100 = 120$$

On dit alors que le produit X est de 20% plus cher dans l'épicerie que dans le supermarché.

Les propriétés des indices :

- La propriété de l'identité : tout indice, qu'il soit élémentaire ou synthétique, vérifie la propriété de l'identité :

$$I_{t/t} = \frac{x_t}{x_t} = 1$$

- La propriété de circularité : grâce à cette propriété on peut changer l'année de référence ou de base dans le calcul d'un indice :

$$I_{t_2/t_1} = \frac{x_{t_1}}{x_{t_2}} = \frac{x_{t_2}}{x_{t_0}} \cdot \frac{x_{t_0}}{x_{t_1}} = I_{t_2/t_0} \cdot I_{t_0/t_1}$$

- La propriété de réversibilité :

$$I_{t/t} = \frac{1}{I_{t_0/1}}$$

II. Les indices synthétiques :

Les indices synthétiques sont un résumé d'un ensemble d'indices élémentaires relatifs à des grandeurs hétérogènes évoluant entre une situation de base 0 et une situation courante t. Le plus souvent ces grandeurs sont des prix (p), des quantités (q) ou des valeurs (v = pq).

On considère un panier composé de k biens achetés chacun en quantité q_i au prix. On s'intéresse à l'évolution de la valeur de ce panier entre les deux dates : 0 et t, pour cela on calcule l'indice de la valeur globale :

$$VG_{t/0} = \frac{VG_t}{VG_0} \cdot 100 = \frac{\sum_{i=1}^k P_t^i \times Q_t^i}{\sum_{i=1}^k P_0^i \times Q_0^i} \times 100$$

Exemple :

On considère un panier du petit déjeuner formé de trois aliments : café, pain au chocolat et jus d'orange. Les quantités consommées et les prix de ces trois aliments entre la date 0 et la date t sont résumés dans le tableau suivant :

	Période 0		Période t	
	Quantité	Prix	Quantité	Prix
Lait	100	80	120	90
Pain au chocolat	180	20	200	24
Jus d'orange	320	40	360	60

L'indice valeur de ce panier du petit déjeuner entre les date 0 et t sera alors :

$$VG_{t/0} = \frac{VG_t}{VG_0} = \frac{90(120) + 24(200) + 60(360)}{80(100) + 20(180) + 40(320)} = 1,525$$

On remarque que la valeur de ce panier a augmenté de 52,5% entre la date 0 et la date t.

Cet indice présente l'inconvénient de ne pas montrer l'origine de l'évolution (hausse ou baisse) qui peut être due soit :

- A une évolution (hausse ou baisse) des prix.

- A une évolution (hausse ou baisse) des quantités.
- A une évolution (hausse ou baisse) à la fois des prix et des quantités.

Afin de cerner l'origine de cette variation on doit fixer les quantités et calculer un indice des prix et ensuite fixer les prix et calculer un indice des quantités. A cet effet, on distingue deux types d'indices selon que l'on se fixe les quantités ou les prix à la date de base 0 ou à la date courante t. On parlera alors respectivement soit d'indices de Laspeyres ou d'indices de Paasche.

1. Les indices de Laspeyres :

On distingue l'indice prix de Laspeyres et l'indice quantité de Laspeyres.

- **L'indice prix de Laspeyres :** Il s'agit de fixer les quantités à la date de référence 0 et calculer ensuite l'évolution d'une grandeur synthétique en tenant compte de l'effet prix uniquement :

$$L^P_{t/0} = \frac{\sum_{i=1}^k P_t^i \times Q_0^i}{\sum_{i=1}^k P_0^i \times Q_0^i} \times 100$$

- **L'indice quantité de Laspeyres :** Il s'agit dans ce cas de fixer les prix à la date de référence (de base) 0 et calculer ensuite l'évolution d'une grandeur synthétique en tenant compte de l'effet quantité uniquement :

$$L^Q_{t/0} = \frac{\sum_{i=1}^k P_0^i \times Q_t^i}{\sum_{i=1}^k P_0^i \times Q_0^i} \times 100$$

Exemple :

Reprenons les données de l'exemple du panier du petit déjeuner, pour calculer les indices quantité et prix de Laspeyres :

$$L^P_{t/0} = \frac{90(100) + 24(180) + 60(320)}{80(100) + 20(180) + 40(320)} = 1,333$$

$$L^Q_{t/0} = \frac{80(120) + 20(200) + 40(360)}{80(100) + 20(180) + 40(320)} = 1,148$$

Remarques :

- On appelle coefficient budgétaire (noté w) associé au bien i , la part de la dépense consacrée à ce bien. Ainsi les coefficients budgétaires d'un bien i , respectivement à la date 0 et à la date t sont :

$$w_0^i = \frac{P_0^i \times Q_t^i}{\sum_{i=1}^k P_0^i \times Q_0^i} \quad ; \quad w_t^i = \frac{P_t^i \times Q_t^i}{\sum_{i=1}^k P_t^i \times Q_t^i}$$

- L'indice synthétique de Laspeyres peut être défini comme étant la moyenne arithmétique des indices élémentaires pondérés par les coefficients budgétaires de la date de base (date 0).

$$L^P_{t/0} = \sum_{i=1}^k w_0^i \times I^i_{t/0}(p) \quad ; \quad L^q_{t/0} = \sum_{i=1}^k w_0^i \times I^i_{t/0}(q)$$

2. Les indices de Paasche :

On distingue l'indice prix de Paasche et l'indice quantité de Paasche.

- **L'indice prix de Paasche :** Il s'agit de fixer les quantités à la date courante t et calculer ensuite l'évolution d'une grandeur synthétique en tenant compte de l'effet prix uniquement :

$$P^P_{t/0} = \frac{\sum_{i=1}^k P_t^i \times Q_t^i}{\sum_{i=1}^k P_0^i \times Q_t^i} \times 100$$

- **L'indice quantité de Paasche :** Il s'agit dans ce cas de fixer les prix à la date courante t et calculer ensuite l'évolution d'une grandeur synthétique en tenant compte de l'effet quantité uniquement :

$$P^q_{t/0} = \frac{\sum_{i=1}^k P_t^i \times Q_t^i}{\sum_{i=1}^k P_t^i \times Q_0^i} \times 100$$

Exemple :

Reprenons les données de l'exemple du panier du petit déjeuner, pour calculer les indices quantité et prix de Paasche :

$$P^P_{t/0} = \frac{90(120) + 24(200) + 60(360)}{80(120) + 20(200) + 40(360)} = 1,329$$

$$P^q_{t/0} = \frac{90(120) + 24(200) + 60(360)}{90(100) + 24(180) + 60(320)} = 1,144$$

Remarque :

- L'indice synthétique de Paasche peut être défini comme étant la moyenne harmonique des indices élémentaires pondérés par les coefficients budgétaires de la date courante (date t) :

$$\frac{1}{P^P_{t/0}} = \sum_{i=1}^k w_t^i \frac{1}{I_{t/0}^i(p)} \quad ; \quad \frac{1}{P^q_{t/0}} = \sum_{i=1}^k w_t^i \frac{1}{I_{t/0}^i(q)}$$

3. Les indices de Fisher :

On peut définir un troisième indice synthétique, appelé indice de Fisher, comme la moyenne géométrique des deux indices de Paasche et de Laspeyres :

$$F^P_{t/0} = \sqrt{P^P_{t/0} L^P_{t/0}} \quad ; \quad F^q_{t/0} = \sqrt{P^q_{t/0} L^q_{t/0}}$$

- On peut établir plusieurs relations entre les indices synthétiques :

$$VG_{t/0} = L^P_{t/0} \cdot P^q_{t/0} = L^q_{t/0} \cdot P^P_{t/0}$$

$$VG_{t/0} = F^P_{t/0} F^q_{t/0}$$

EXERCICES CORRIGES

Exercice n°1 :

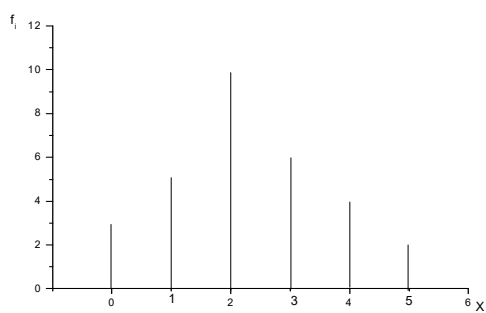
La distribution d'un échantillon d'individus selon le nombre de personnes par ménage est résumée dans le tableau suivant :

Valeurs x_i	Effectifs n_i
0	3
1	5
2	10
3	6
4	4
5	2
Total	30

TAF : Représenter cette distribution graphiquement.

Solution :

Nombre de personnes par ménage est une variable statistique discrète le graphique qui convient est le : **Diagramme en bâtons.**



Exercice n°2 :

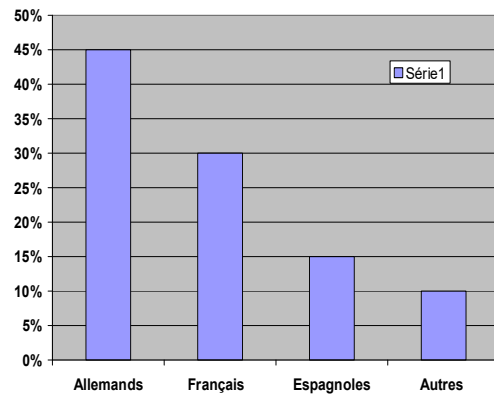
La distribution de personnes dans une station balnéaire selon la nationalité est résumée dans le tableau suivant :

Nationalité	Pourcentage %
Allemands	45
Français	30
Espagnols	15
Autres	10
Total	100

TAF : Représenter par un diagramme en tuyau d'orgue cette distribution.

Solution :

Caractère qualitatif : **Diagramme en tuyau d'orgue.**



Exercice n° 3 :

Une voiture parcourt 100Km/h, puis 160Km/h à 80Km/h. Calculer la vitesse moyenne ?

Solution :

$$\text{Vitesse moyenne} = \frac{\text{Distance totale}}{\text{Temps total}} = \frac{100+160}{\frac{100}{50} + \frac{160}{80}} = 6.5$$

$$H = \frac{\sum n_i}{\sum n_i \cdot \frac{1}{x_i}}$$

Exercice n°4 :

Un étudiant a obtenu les notes suivantes : 8-10-12. On veut calculer la moyenne des écarts entre les notes et la moyenne arithmétique. Quelle est la moyenne appropriée dans ce cas ?

Solution :

$$\bar{X} = \frac{8+10+12}{3} = 10$$

Ecart type à la moyenne
8-10 = -2
10-10 = 0
12-10 = 2

moyenne arithmétique des écarts = $(-2+0+2)/3$

moyenne arithmétique des écarts = 0

On retrouve ici une des propriétés des moyennes arithmétiques :

$$\sum (x_i - \bar{X}) = 0$$

Démonstration : $\sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} = \sum x_i - n \frac{\sum x_i}{n} = 0$

Si on veut calculer la moyenne des écarts, il vaut mieux **calculer la moyenne quadratique** :

$$Q^2 = \frac{(-2)^2 + (0)^2 + (2)^2}{3} = \frac{8}{3}$$

$$Q = \sqrt{\frac{8}{3}} = 1.6$$

Exercice n°5 :

Soit la série suivante relative aux deux variables X (Prix) et Y (Quantités) :

Prix (x)	Quantités (y)
70	72
75	51
77	60
80	62
86	74
33	83
481	372

Travail à faire :

1. Trouver $D_y(x)$
2. Calculer le coefficient de corrélation d'une autre façon (existe-t-il un lien entre y et x).

Solution :

1. $D_y(x)$ a pour équation :

$$Y = ax + b$$

$$a = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$\bar{x} = \frac{\sum x_i}{n} = 481 / 6 = 80$$

$$\bar{y} = 372 / 6 = 62$$

$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
-10	-20	200	100
-5	-11	55	25
-3	-2	6	9
0	0	0	0
6	12	72	36
13	21	273	169
		606	339

$$a = 606 / 339 = 1.79$$

$$b = 62 - (1.79)80$$

$$b = -81$$

Donc

$D_{y(x)}$ a pour équation :

$$y = 1.79x - 81$$

La loi de l'offre pour ce bien

2.

Prix	Qté	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2 (y - \bar{y})^2$
70	42				
75	51				
77	60				
80	62				
86	74				
33	83				
			606	339	1110

$$r^2 = a \cdot a' = \frac{606}{339} \times \frac{606}{1110} \text{ donc } r = 0.98$$

$$a = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{606}{339} = 1.79$$

$$a' = \frac{\sum x_i y_i}{\sum y_i^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2} = \frac{606}{1110} = 0.545$$

On a une très forte corrélation car $r = 0.975$ tend vers 1

Exercice 6 :

Une étude portant sur les petites et moyennes entreprises(PME) a donné les résultats suivants en milliers de DA :

Ecart inter-décile	Etendue générale	Neuvième décile	Premier quartile	médiane	Ecart interquartile	Mode	Ecart type	Chiffre d'affaires moyen	Chiffre d'affaires minimum
2800	4000	5500	3100	3600	100	3550	550	3900	2500

Travail à faire :

1. Quel est le chiffre d'affaire maximum de ces petites et moyennes entreprises (PME).
2. Calculez le troisième quartile puis le premier décile.
3. Classez ces paramètres en deux parties : -paramètres de la tendance centrale et paramètres de dispersion.
4. Calculez le coefficient d'asymétrie de Pearson. Interprétez le résultat.

Solution :

- 1. Le chiffre d'affaire maximum de ces petites et moyennes entreprises (PME) :**

$$E = \text{Max}_{xi} - \text{Min}_{xi} \Leftrightarrow \text{Max}_{xi} = E + \text{Min}_{xi} \Leftrightarrow \\ \text{Max}_{xi} = 4000 + 2500 = 6500 \Leftrightarrow \boxed{\text{Max}_{xi} = 6500 \times 10^3 \text{DA}}$$

- 2. Le troisième quartile :**

$$IQ = Q_3 - Q_1 \Leftrightarrow Q_3 = IQ + Q_1 \Leftrightarrow \\ Q_3 = 100 + 3100 = 3200 \Leftrightarrow \boxed{Q_3 = 3200 \times 10^3 \text{DA}}$$

Le premier décile

$$ID = D_9 - D_1 \Leftrightarrow D_1 = D_9 - ID \Leftrightarrow \\ D_1 = 5500 - 2800 = 2700 \Leftrightarrow \boxed{D_1 = 2700 \times 10^3 \text{DA}}$$

- 3. Classement des paramètres :**

- **Paramètres de la tendance centrale :** Mode – Moyenne arithmétique – Médiane.
- **Paramètres de dispersion :** Etendu – Ecart interquartile – Ecart inter décile et Ecart-type.

- 4. Coefficient d'asymétrie de Pearson :**

$$C_p = \frac{\bar{X} - M_o}{\delta_x} = \frac{3900 - 3550}{550} \Rightarrow \boxed{C_p = +0,63}$$

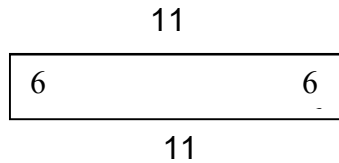
Interprétation :

$$\boxed{C_p > 0}$$

Assymétrie à droite.

Exercice 7 :

Soit un petit jardin sous forme de rectangle, le propriétaire ne peut se souvenir que d'un seul chiffre :



Travail à faire :

1. Quel type de moyenne choisir (il faut utiliser) ?
2. Conclusion.

Solution :

S'il veut entourer son champ de fil de fer, il a intérêt à se souvenir de la moyenne arithmétique car le périmètre est lié à la somme des différents côtés.

$$\bar{X} = \frac{11+6+11+6}{4} = 8.5$$

S'il veut mettre de l'engrais à son jardin, il a intérêt à se souvenir de la moyenne géométrique :

$$G = \sqrt[4]{11^2 * 6^2} = 8.12$$

Moyenne arithmétique du périmètre = $34 = 8.5 * 4 \neq 8.12 * 4$

Moyenne géométrique : surface = $65.93 = 8.12 * 8.12 \neq 8.5 * 8.5$

Conclusion :

D'une manière générale, on retient la moyenne arithmétique quand les variables s'additionnent, et on utilise la moyenne géométrique lorsque les variables se multiplient.

Exercice 8 :

Une voiture roule pendant 200 kilomètres à 50 km/h, puis pendant 100 kilomètres à 100 km/h.

Quelle est sa vitesse moyenne sur son trajet ?

Solution :

Cette vitesse moyenne sera égale au rapport entre la distance parcourue et le temps de trajet :

Soit $200 + 100 = 300$ kilomètres parcourus.

$$\text{Vitesse moyenne} = \frac{\text{Distance totale}}{\text{Temps total}} = \frac{200+100}{\frac{200}{50} + \frac{100}{100}} = \frac{300}{5} = 60 \text{ Km/h}$$

$$H = \frac{\sum n_i}{\sum n_i \cdot \frac{1}{x_i}}$$

Soit :

200 km à 50 km/h durent 4 heures

100 km à 100 km/h durent 1 heure

Donc le trajet dure 5 heures pour 300 kilomètres parcourus : la vitesse moyenne est bien de 60 km/h.

Si on note x_i la vitesse et n_i la distance parcourue à cette vitesse, la vitesse moyenne est donnée dans le cas général par :

$$H(x) = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \frac{n_i}{x_i}} \text{ ou } \frac{1}{H(x)} = \frac{1}{N} \sum_{i=1}^k \frac{n_i}{x_i}$$

La vitesse moyenne est égale à la moyenne harmonique des vitesses pondérées par les distances.

Exercice 9 :

Une grandeur S_0 a augmenté sur 3 années : 10% la première année, puis de 15% la deuxième année et 30% pour la troisième année.

Travail à faire :

Quel est le taux moyen de croissance ?

Solution :

Fin de la 1^{ère} année : S_0 devient $S_1 = S_0 + (S_0 \cdot 10/100) \Rightarrow S_1 = S_0(1+0.10) = 1.10 \cdot S_0$

Fin de la 2^{ème} année S_1 devient $S_2 = S_1 + 0.15 \cdot S_1 \Rightarrow S_1 \cdot 1.15 \Rightarrow (S_1 \cdot (1+0.15))$

Fin de la 3^{ème} année S_2 devient $S_3 = S_2 + 0.3 \cdot S_2 = 1.3 \cdot S_2 \Rightarrow (S_2 \cdot (1+0.3))$

$$S_3 = S_0 \times 1.10 \times 1.15 \times 1.3$$

D'autre part on a: $S_3 = S_0 \cdot (1+G)^3$

$$\Rightarrow S_3 = S_0 \cdot (1+G)^3 = S_0 \times 1.10 \times 1.15 \times 1.3$$

$$\Rightarrow (1+G)^3 = 1.10 \times 1.15 \times 1.3$$

$$\Rightarrow (1+G) = (1.10 \times 1.15 \times 1.3)^{\frac{1}{3}}$$

$$\Rightarrow (1+G) = (1.10 \times 1.15 \times 1.3)^{\frac{1}{3}}$$

Moyenne géométrique $G = \sqrt[3]{1.10 \times 1.15 \times 1.3} = 1.1804$

Remarque : Le taux de croissance moyen de cette grandeur est : 18.04%.

Exercice 10 :

1. Les importations de pièces de rechange d'un pays en **2022** (date 0) sont égales à X_0 millions de DA. Les indices élémentaires d'importation de pièces de rechange sont : $I_{0/1} = 0,8$; $I_{3/2} = 1,15$; $I_{3/1} = 1,61$. Déterminer alors en fonction de X_0 les montants relatifs aux importations de pièces de rechange entre **2022** et **2025**.
2. Calculer l'indice $I_{3/0}$.

Solution :

1. Les importations de pièces de rechange en fonction de la valeur d'importation en **2022** (année de référence) :

Année	Importation
2022 (0)	X_0
2023 (1)	$1,25 X_0$
2024 (2)	$1,75 X_0$
2025 (3)	$2,0125 X_0$

$$2. I_{3/0} = I_{3/1} I_{1/0} = I_{3/1} I_{0/1} = 1,61 \cdot 0,8 = 1,288$$

$$\text{Ou bien : } I_{3/0} = 2,0125 X_0 = X_0$$

Exercice 11 :

Le nombre d'abonnements ADSL d'un pays a connu durant les quatre dernières années, les taux de croissance suivants :

- 2,5% entre 2021 et 2022.
- 1% entre 2022 et 2023.
- 3% entre 2023 et 2024.
- 2% entre 2024 et 2025.

Déterminer l'indice élémentaire du nombre d'abonnements ADSL entre 2025 et 2021.

Solution :

On peut déduire les indices élémentaires du nombre d'abonnements ADSL à partir des taux de croissance :

$$I_{2022/2021} = 1,025$$

$$I_{2023/2022} = 0,99$$

$$I_{2024/2023} = 1,03$$

$$I_{2025/2024} = 1,02$$

$$\text{Ainsi : } I_{2025/2021} = I_{2025/2024} \cdot I_{2024/2023} \cdot I_{2023/2022} \cdot I_{2022/2021}$$

$$I_{2025/2021} = (1,02) \cdot (1,03) \cdot (0,99) \cdot (1,025) = 1,066$$

Exercice 12 :

On considère un panier formé de quatre biens de première nécessité. Le tableau suivant résume les consommations et les prix, unitaires de ces quatre biens entre 2023 et 2024 :

	Année 2023		Année 2024	
	Quantité	Prix unitaire	Quantité	Prix unitaire
Bien 1	40	1	35	1,1
Bien 2	25	3	30	3
Bien 3	50	0,5	40	0,6
Bien 4	10	2	15	2,1

Travail à faire :

1. Calculer l'indice synthétique de la valeur globale. Interpréter.
2. Calculer les indices prix de Laspeyres et Paasche.
3. En déduire les indices quantité de Laspeyres et Paasche.
4. Calculer les indices prix et quantité de Fisher.

Solution :

1.

$$VG_{2024/2023} = \frac{VG_{2024}}{VG_{2023}} \cdot 100 = \frac{\sum_{i=1}^k P_{2024}^i \times Q_{2024}^i}{\sum_{i=1}^k P_{2023}^i \times Q_{2023}^i} \times 100 = \frac{184}{160} = 1,15$$

La valeur globale de ce panier a augmenté de 15% entre 2024 et 2023.

2.

$$L^P_{2024/2023} = \frac{\sum_{i=1}^k P_{2024}^i \times Q_{2023}^i}{\sum_{i=1}^k P_{2023}^i \times Q_{2023}^i} \times 100 = \frac{170}{160} = 1,0625$$

$$P^P_{2024/2023} = \frac{\sum_{i=1}^k P_{2024}^i \times Q_{2024}^i}{\sum_{i=1}^k P_{2023}^i \times Q_{2024}^i} \times 100 = \frac{184}{175} = 1,0514$$

3.

$$VG_{2024/2023} = L^P_{2024/2023} \cdot P^q_{2024/2023}$$

$$\text{d'où : } P^q_{2024/2023} = \frac{VG_{2024/2023}}{L^P_{2024/2023}} = \frac{1,15}{1,0625} = \boxed{1,0824}$$

$$VG_{2024/2023} = L^q_{2024/2023} \cdot P^P_{2024/2023}$$

$$\text{d'où : } L^q_{2024/2023} = \frac{VG_{2024/2023}}{P^P_{2024/2023}} = \frac{1,15}{1,0514} = \boxed{1,0938}$$

4.

$$F^P_{2024/2023} = \sqrt{P^P_{2024/2023} \cdot L^P_{2024/2023}} = \sqrt{1,0514 (1,0625)} = 1,0569$$

$$F^q_{2024/2023} = \sqrt{P^q_{2024/2023} \cdot L^q_{2024/2023}} = \sqrt{1,00824 (1,0938)} = 1,0881$$

Exercice 13 :

La société ELECTROM commercialise des produits électroménagers. Elle s'intéresse en particulier au portefeuille de l'électroménager blanc (réfrigérateur, cuisinière, four, machine à laver et lave-vaisselle). Elle nous communique les statistiques des ventes des années 2020 et 2025 :

	Année 2020		Année 2025	
	Quantité	Prix unitaire 10 ³ DA	Quantité	Prix unitaire DA 10 ³
Réfrigérateur,	20	800	30	1040
Cuisinière	16	250	20	300
Four	5	300	10	420
machine à laver	4	400	10	500
Lave-vaisselle	1	900	5	720

Travail à faire :

1. Calculer les indices élémentaires des prix et des quantités.
2. Calculer les coefficients budgétaires pour les années 2020 et 2025.
3. Calculer les indices prix et quantités de Laspeyres.
4. Calculer les indices prix et quantité de Paasche.

Solution :

1.

Les indices élémentaires des prix et des quantités :

	Indice élémentaires (Quantité)	Indice élémentaires (Prix)
Réfrigérateur	1,5	1,3
Cuisinière	1,25	1,2
Four	2	1,4
machine à laver	2,5	1,25
Lave-vaisselle	5	0,8

2.

Les coefficients budgétaires pour les années 2020 et 2025 :

	Coefficients budgétaires (2020)	Coefficients budgétaires (2025)
Réfrigérateur	66,67%	62,4%
Cuisinière	16,67%	12,0%
Four	6,25%	8,4%
machine à laver	6,67%	10,0%
Lave-vaisselle	3,75%	7,2%

3.

Indices prix de Laspeyres :

$$L^P_{2025/2020} = \sum_{i=1}^k w_{2020}^i \times I^i_{2025/2020}(p) = 0,666(1,3) + \dots + 0,0375(0,8) = 1,267$$

Indices quantité de Laspeyres :

$$L^q_{2025/2020} = \sum_{i=1}^k w_{2020}^i \times I^i_{2025/2020}(q) = 0,666(1,5) + \dots + 0,0375(05) = 1,6876$$

4.

Indices prix de Paasche :

$$\frac{1}{P^P_{2025/2020}} = \sum_{i=1}^k w_{2025}^i \frac{1}{I^i_{2025/2020}(p)} = 0,625\left(\frac{1}{1,3}\right) + \dots + 0,072\left(\frac{1}{0,8}\right) = 0,81$$

$$P^P_{2024/2023} = \frac{1}{0,81} = 1,2346$$

Indices quantité de Paasche :

$$\frac{1}{P^q_{2025/2020}} = \sum_{i=1}^k w_{2025}^i \frac{1}{I_{2025/2020}^i(q)} = 0,625 \left(\frac{1}{1,5} \right) + \dots + 0,072 \left(\frac{1}{5} \right) = 0,6084$$

$$P^q_{2024/2023} = \frac{1}{0,6084} = 1,6437$$

BIBLIOGRAPHIE

المراجع باللغة العربية :

- جاك لوكايون و كريستيان لاپروس : "الإحصاء الوصفي"، ديوان المطبوعات الجامعية الجزائر 1985
- محمد صبحي أبو صالح و عدنان محمد عوض : "مقدمة في الإحصاء"، دار المسيرة عمان 2007.
- عز الدين جوني : "نظرية الإحصاء"، ديوان المطبوعات الجامعية الجزائر 1984.
- عبد العزيز فهمي هيكل : "طرق التحليل الإحصائي"، دار النهضة العربية بيروت.
- عبد الجبار توفيق البياتي : " الإحصاء و تطبيقاته"، إثراء للنشر و التوزيع، عمان 2008.
- جلاطو جيلالي: " الإحصاء مع تمارين و مسائل محلولة"، ديوان المطبوعات الجامعية - الجزائر- 1998.
- حامد نور الدين: " دروس في مقياس الإحصاء الوصفي"، دار اليازوري العلمية للنشر و التوزيع – عمان - 2016.

Références en langue française :

- AIT HABOUCHE-MIHOUB Ouahiba., « **Statistique et mathématique en économie et gestion : Rappel de cours et exercices sur données réelles** », Edition Dar El AdIb, 2013.
- BAGGIO Stéfanie, ROTHEN Stéphane et DELINE Stéphane., « **Statistique descriptive en 20 fiches** », 2^e édition, De Boeck Supérieur, 2022.
- BAILLARGEON G., « **Probabilité, statistique descriptive et technique de régression** », SMG, 1989.
- BARTHE René., « **La Statistique Descriptive en 10 leçons : Méthode progressive "ABCD"** », Edition Economica, 1989.
- CALOT Gérard, « **Cours de statistique descriptive** », édition Dunod, Paris, 1973.
- CHAUVAT Gérard, REAU Jean-Philippe, « **Statistiques descriptives** », ARMAND COLIN, 2002.
- DAGNELIE P., « **Théorie et méthodes statistiques 2 tomes** », Presses agronomiques Gembloux 1985.

- DROESBEKE Jean-Jacques, « **Eléments de Statistique** », OPU, 1988 by Editions de l'Université de Bruxelles (Belgique).
- GIARD Vincent., « **Statistique appliquée à la gestion** », Ed. Economica, 1992.
- GOLDFARB Bernard, PARDOUX Catherine, « **Introduction à la Méthode Statistique, Exercices Corrigés** », DUNOD, 2003.
- GOLDFARB Bernard, PARDOUX Catherine, « **Introduction à la Méthode Statistique** », DUNOD, 2003.
- GRAIS B., « **Statistique descriptive** », Edition Dunod, 1995.
- GRAIS B., « **Exercices corrigés de statistique descriptive avec rappel de cours** », Dunod, 1991.
- GRAIS B., « **Statistique pour économistes** », Economica, 1987.
- GUEGUEN Nicolas, BOULBRY Gaëlle, « **Le petit stat 2026 : L'essentiel en bref** », Collection les petits experts, 48 pages, septembre 2025.
- GUITTON Henri, « **Statistique** », Edition Dalloz, 1976.
- KAZMIER Léonard G, « **Statistiques de la gestion** », McGRAW-HILL Québec 1982.
- KHALDI Khaled, « **Méthodes statistiques** », OPU Constantine 2005.
- LABROUSSE Christian, Statistique, « **Exercices corrigés avec rappel de cours** », Dunod, 292 p., Paris, 1980.
- LECOURTE Jean-Pierre, « **Statistique et probabilités – Cours et exercices corrigés** », Edition Dunod, 2023.
- LEGROS Benjamin, « **L'essentiel des Statistiques et probabilités en économie-gestion** », 2^e édition Dunod, Janvier 2025.
- LETHIELLEUX Maurice, « **Statistique descriptive** », Dunod, 2003.
- MASIERI Walder, « **Méthodes Quantitatives** », 2^{ème} Dunod, Paris 1990.

- MASIERI Walder , « **Statistique et calcul des probabilités** », Dalloz , Paris 2001.
- MONINO Jean-Louis, KOSIANSKI Jean-Michel, LE CORNU François, «**Statistiques descriptives - Travaux dirigés** », DUNOD, 2000.
- MURRAY R. Spiegel, « **Statistique** », Série Schaum, 2^{ème} édition, traduction de la 2^{ème} américaine, France, 1996.
- MURRAY R. Spiegel, « **Théorie et application de la statistique** », McGRAW-HILL Québec 1985.
- PACE P. et CLUZEL R, « **Statistique et probabilité** », Ed Delagrave Paris 1982.
- PY Bernard, « **Exercices corrigés de Statistique Descriptive** », ECONOMICA, 2^{ème} édition, Paris, 1994.
- PY Bernard, « **Statistique Descriptive** », ECONOMICA, 4^{ème} édition, Paris, 2001.
- SALVAORE Dominick, « **Econométrie et statistique appliquée, Cours et problème** », Série Schaum, 297 p., Paris, 1985.
- TENENHAUS Michel, « **Statistique : Méthodes pour décrire, expliquer et prévoir** », DUNOD, 2006.

TABLE DES MATIERES

Avant-propos	I
Sommaire	II
Introduction générale	1
Chapitre I : La présentation des statistiques	2
I : Définitions :	2
1- Les statistiques	2
2- La statistique	2
II: La statistique utile aux économistes	3
1- La statistique est utile aux théoriciens de l'économie	3
2- La statistique est utile aux praticiens de l'économie	4
III: Les limites de la méthode statistique	4
IV: La terminologie statistique	4
1- Population	4
2- Unité statistique ou Individu	5
3- Caractère et modalité	5
4- Effectifs (fréquences absolues)	8
5- Fréquences relatives	8
6- Série statistique	8
7- Classe	9
V: Quelques symboles mathématiques utilisés	9
1- L'opérateur Somme de : Sigma Σ	9
2- L'opérateur de produit : Π	10
Chapitre II : Les représentations graphiques	11
I : Cas d'une variable qualitative :	11
1- Diagramme en secteurs (circulaire)	11
2- Diagramme en tuyau d'orgue	12
3- Diagramme rectiligne	13
II : Cas d'une variable quantitative :	13
A- Séries statistiques discrètes	13
1- Diagramme en bâtons	13
2- Fréquences cumulées croissantes, fonction de répartition et diagramme cumulatif	15
B - Séries statistiques continues	17
1- Histogramme	17
2- Polygone des fréquences	19
3- Courbes des effectifs ou des fréquences cumulées croissantes et décroissantes	20
Chapitre III : Les caractéristiques de tendance centrale ou paramètres de position	23
I : Le Mode :	23
II : La Médiane (Me)	28

III- La moyenne arithmétique ($\bar{X}$)	32
IV- Autres moyennes	38
1- La moyenne géométrique (G)	38
2- La moyenne harmonique (H)	41
3- La moyenne quadratique (Q)	43
4- Généralisation de la notion de moyenne : Moyenne d'ordre r	45
5- Classement des moyennes : Les inégalités entre moyenne	45
 Chapitre II : Les caractéristiques de dispersion	 46
I : L'étendu ou l'intervalle de variation :	47
II : Les écarts inter quartiles	47
1- Les quantiles	47
2- Les quartiles	47
3- Les déciles	49
4- Les percentiles (ou les centiles)	50
5- Les intervalles inter quantiles	51
III- Mesure de la dispersion autour de la moyenne	54
IV- Variance et écart-type	56
1- Variance	56
2- Ecart-type	57
V : Coefficient de variation	59
IV : Coefficient interquartile (relatif)	60
 Chapitre V : Les caractéristiques de forme	 61
I : Les moments d'une série statistique	61
1- Moments non centrés	61
2- Moments centrés	62
II : Les coefficients d'asymétrie	62
1- Le coefficient empirique de Pearson	64
2- Le coefficient d'asymétrie de Yule	64
3- Le coefficient d'asymétrie de Pearson	64
III : Les coefficients d'aplatissement	65
1- Le coefficient d'aplatissement de Pearson	65
2- Le coefficient d'aplatissement de Fisher	66
 Chapitre VI : Les caractéristiques de concentration	 68
I : La détermination algébrique de la concentration	69
1- Notion de médiale (Mle)	69
2- Mesure de la concentration	69
II : La détermination graphique de la concentration	71

1- La courbe de concentration (ou courbe de Lorenz)	71
2- L'indice de Gini	72
Chapitre VII : Corrélation et ajustement linéaire	73
I : Ajustement linéaire d'un nuage de points	73
1- La droite de régression de y sur x	73
1-1 La méthode des moindres carrées	74
1-2 Calcul des paramètres de la droite de régression	74
2- La droite de régression de x sur y	77
II : La covariance entre x et y	78
III : Le coefficient de corrélation linéaire entre x et y	79
IV : Le coefficient de détermination entre x et y	84
Chapitre VIII : Les indices statistiques	85
I : Les indices élémentaires	85
II : Les indices synthétiques	87
1- Les indices de Laspeyres	88
2- Les indices de Paasche	89
3- Les indices de Fisher	90
Exercices corrigés	91
Bibliographie	106
Table des matières	110

