



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

Département des Sciences Économiques

Laboratoire d'Économie et Développement

Polycopié pédagogique

Dossier numéro (à remplir par l'administration) :

Titre

Statistiques I

Cours destiné aux étudiants de Licence : Première année L1

Année : 2024/2025

Avant-propos

Ce polycopié de cours de Statistique I s'adresse principalement aux étudiants de première année en sciences économiques, commerciales et sciences de gestion. Il est conçu pour répondre aux exigences du programme pédagogique de statistique descriptive, tout en étant accessible à toute personne souhaitant découvrir et maîtriser les notions et techniques fondamentales de cette discipline.

Fruit de plusieurs années de travail et d'expérience pédagogique, ce support vise à fournir une introduction claire et rigoureuse aux bases de la statistique descriptive. Il combine théorie et pratique pour permettre aux étudiants de comprendre, d'analyser et d'interpréter des données numériques dans divers contextes.

Le contenu est structuré en plusieurs chapitres organisés de manière progressive. Il débute par une présentation des concepts et définitions de base (comme la population, l'unité statistique et la typologie des variables) et aborde ensuite les outils fondamentaux de la statistique descriptive, tels que les tableaux, les graphiques et les diagrammes. Les chapitres suivants traitent des caractéristiques de position (moyenne, médiane, mode) pour résumer les distributions, des mesures de dispersion (étendue, variance, écart-type) pour évaluer leur variabilité, et des caractéristiques de forme (asymétrie, aplatissement) pour analyser leur structure. Enfin, les notions de corrélation et de régression linéaire permettent d'introduire l'étude des relations entre variables et la construction de modèles simples de prévision.

Chaque chapitre est enrichi de nombreux exemples concrets et contextualisés, qui illustrent les notions abordées, ainsi que d'exercices variés pour vérifier l'acquisition des concepts. Le langage mathématique employé a été soigneusement adapté pour être accessible aux débutants, rendant ainsi l'apprentissage plus fluide et motivant.

Ce polycopié a été conçu dans l'objectif de développer chez les étudiants un raisonnement logique et analytique, tout en leur offrant des outils pratiques pour mieux comprendre les phénomènes économiques et sociaux. Nous espérons qu'il contribuera à démystifier cette discipline et à encourager les apprenants à approfondir leur exploration des méthodes statistiques.

Introduction

La statistique descriptive est une branche fondamentale des statistiques, essentielle pour toute analyse et interprétation de données. Elle constitue le point de départ de nombreuses études, en particulier dans les sciences économiques, commerciales et de gestion, où elle offre des outils indispensables pour collecter, organiser, synthétiser et présenter l'information. En transformant des données brutes en résultats clairs et exploitables, la statistique descriptive permet une meilleure compréhension et une prise de décision éclairée.

Dans un monde où les données jouent un rôle central, chaque décision stratégique repose sur une analyse quantitative rigoureuse. La statistique descriptive se positionne ainsi comme une compétence incontournable, fournissant les bases nécessaires pour appréhender des phénomènes complexes, analyser des tendances économiques ou visualiser des informations sous une forme compréhensible et pertinente.

Ce module a été conçu pour répondre aux besoins spécifiques des étudiants de première année en sciences économiques, commerciales et de gestion, tout en s'adressant également à toute personne souhaitant maîtriser les bases de la statistique descriptive. En adoptant une approche pédagogique progressive et pratique, il combine des notions théoriques avec des exemples concrets afin de faciliter l'apprentissage et de le rendre à la fois clair et engageant.

Le contenu est structuré de manière logique, débutant par une introduction aux concepts fondamentaux et au vocabulaire statistique, pour établir des bases solides. Il explore ensuite les principales méthodes de présentation et d'organisation des données, telles que les tableaux, les graphiques et les diagrammes. Les mesures de tendance centrale (moyenne, médiane, mode) et de dispersion (étendue, variance, écart-type) sont développées pour permettre une analyse approfondie et précise des distributions.

Les chapitres avancés abordent des outils spécifiques, tels que les paramètres de forme et de concentration, pour étudier les asymétries et les disparités, ainsi que les concepts de corrélation et de régression linéaire, ouvrant la voie à l'analyse des relations entre variables et à la prévision. Chaque notion est illustrée par des exemples concrets et accompagnée d'exercices pratiques, permettant aux étudiants d'évaluer et d'affiner leur compréhension.

Ce polycopié vise à rendre la statistique descriptive accessible et compréhensible, en transformant une discipline souvent perçue comme abstraite en un domaine concret et stimulant. Son objectif est de doter les apprenants de compétences analytiques solides et durables. Nous espérons qu'il deviendra un outil précieux, alliant des applications pratiques à une compréhension approfondie des phénomènes économiques et sociaux, indispensables à leur réussite tant académique que professionnelle.

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

Introduction

Pour avoir une idée de la place et de l'importance que revêt la statistique dans la vie de tous les jours, il suffit d'allumer son ordinateur, de parcourir les actualités en ligne ou d'écouter les informations à la radio. On y entend parler des taux de chômage, des prévisions économiques, des chiffres liés à la santé publique, ou encore des tendances dans les réseaux sociaux. Ces informations sont issues d'analyses statistiques qui permettent de comprendre et d'interpréter les phénomènes complexes qui nous entourent. Cela illustre que le monde moderne est largement axé sur le quantitatif et le mesurable, avec une recherche constante de données fiables pour éclairer les décisions.

La statistique, bien qu'étant une discipline relativement nouvelle, s'est imposée comme un outil incontournable dans un contexte où les données jouent un rôle essentiel. Elle permet non seulement de résumer et d'interpréter l'information de manière compréhensible, mais également de faire des prévisions ou des comparaisons utiles dans des domaines aussi variés que l'économie, la santé, l'éducation, ou encore l'environnement. Ainsi, la statistique s'aligne parfaitement avec les besoins d'un monde de plus en plus tourné vers la rationalité et l'objectivité, où chaque décision s'appuie sur des éléments tangibles et mesurables.

1. La collecte de l'information

La collecte de données chiffrées remonte à l'Antiquité, où la nécessité d'obtenir des informations sur la population et ses richesses s'est imposée comme un enjeu essentiel à la pérennité des États. Aujourd'hui, pour collecter des informations relatives à une population statistique, deux méthodes principales sont utilisées :

1.1. La méthode exhaustive (recensement)

La collecte statistique porte sur la totalité de la population, où chaque individu de la population est étudié selon le ou les caractères étudiés.

Exemple :

L'Algérie procède tous les dix ans au recensement général de la population et de l'habitat (RGPH), une opération essentielle pour mieux comprendre la dynamique démographique et les caractéristiques de l'habitat à l'échelle nationale. Depuis l'indépendance, six recensements ont été réalisés : en 1966, 1977, 1987, 1998, 2008 et, plus récemment, en 2022. Ces recensements fournissent une multitude d'informations précieuses. Concernant la population, on peut y trouver des données sur la répartition géographique, les sexes, les âges, les niveaux d'instruction, les catégories d'emploi, la croissance démographique et l'espérance de vie. Du côté de l'habitat, des informations détaillées sont également recueillies, telles que la consistance du parc de logements, le nombre d'occupants par logement, l'âge des habitations, leur état de vétusté, et la présence éventuelle d'habitats informels.

Le RGPH est un outil fondamental pour les autorités publiques, car il leur permet de planifier efficacement les politiques sociales, économiques et urbaines. Il constitue également une ressource précieuse pour les universitaires, les chercheurs, ainsi que pour les acteurs économiques, sociaux et culturels qui s'appuient sur ces données pour leurs travaux et décisions. Cependant, cette opération de grande envergure est complexe et coûteuse, mobilisant d'importants moyens humains, financiers et matériels.

Face aux contraintes budgétaires ou au manque de temps, il peut parfois être difficile de mener un recensement exhaustif. Dans de telles situations, une alternative consiste à effectuer l'étude sur une partie représentative de la population, appelée échantillon. Cette méthode permet d'obtenir des résultats fiables tout en réduisant considérablement les coûts et les efforts nécessaires.

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

1.2. La méthode des sondages (échantillonnage)

Le sondage est une méthode de collecte non exhaustive des données qui consiste à examiner seulement une fraction de la population, appelée échantillon. L'échantillonnage est utilisé pour plusieurs raisons : observer une partie d'une population infinie, réduire les coûts par rapport à un recensement complet, et éviter la destruction d'individus lors de l'observation. Il existe deux méthodes d'échantillonnage.

1.2.1. Echantillonnage aléatoire :

L'échantillonnage aléatoire est une méthode statistique qui sélectionne un sous-ensemble représentatif d'une population, assurant que chaque individu ait une chance égale ou proportionnelle d'être inclus. Cela permet d'obtenir des résultats généralisables. Plusieurs techniques, telles que l'échantillonnage élémentaire, systématique, stratifié, par grappes, par degrés et à probabilité inégale, sont utilisées pour améliorer la précision et la représentativité des estimations selon les besoins spécifiques des études.

a. Echantillonnage élémentaires

Dans une population de taille N où à chaque individu a une probabilité $\frac{1}{N}$ d'être choisi, on procède à un tirage au hasard de n individus (à l'aide d'une table de nombre au hasard par exemple). On distingue deux types de tirage : le tirage exhaustif (sans remise) et le tirage non exhaustif (avec remise).

Exemple :

Supposons que vous souhaitez estimer la consommation moyenne d'eau par ménage dans un quartier composé de $N=50$ ménages. Vous décidez de sélectionner un échantillon de $n=10$ ménages à l'aide d'un tirage aléatoire.

a.1. Tirage exhaustif (sans remise)

Dans un tirage sans remise, chaque ménage ne peut être sélectionné qu'une seule fois. Par exemple, vous notez les numéros des 50 ménages sur des papiers, les mélangez, puis tirez au hasard 10 papiers. Une fois qu'un numéro est tiré, il est exclu des tirages suivants. Ainsi, chaque ménage a une probabilité égale d'être sélectionné, mais une fois tiré, il ne peut plus l'être à nouveau.

Conséquence : Les 10 ménages sélectionnés sont tous différents. Ce type de tirage est souvent utilisé lorsqu'on souhaite éviter les répétitions dans l'échantillon.

a.2. Tirage non exhaustif (avec remise)

Dans un tirage avec remise, après avoir sélectionné un ménage, son numéro est remis dans le lot, ce qui signifie qu'il peut être tiré à nouveau. Par exemple, après avoir tiré un numéro, vous le remettez dans l'ensemble des 50 numéros, mélangez à nouveau, et tirez un autre numéro. Ce processus est répété jusqu'à ce que 10 ménages soient sélectionnés.

Conséquence : Il est possible que certains ménages soient sélectionnés plusieurs fois dans l'échantillon. Ce type de tirage est utilisé lorsque les répétitions ne posent pas de problème ou lorsqu'on souhaite préserver l'indépendance des tirages.

b. Echantillonnage systématique

L'échantillonnage systématique consiste à sélectionner les individus de la population à intervalles réguliers. Après avoir tiré un premier individu au hasard, on inclut dans l'échantillon chaque $k^{\text{ème}}$ individu de la population, où k représente le pas de sélection, calculé comme le rapport entre la taille de la population et celle de l'échantillon.

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

Exemple :

Supposons que nous disposons d'une liste de 250 individus et que nous souhaitons sélectionner un individu sur 10. Pour commencer, on choisit aléatoirement le premier individu parmi les 10 premiers de la liste. Par exemple, si le 5^{ème} individu est sélectionné, les suivants seront le 15^{ème}, le 25^{ème}, le 35^{ème}, et ainsi de suite jusqu'à couvrir l'ensemble de la liste.

c. Echantillonnage stratifié

On tire les individus de l'échantillon dans des groupes homogènes de la population que nous appelons strates. On cherche le nombre optimal d'individus à tirer dans chaque strate. Cette méthode a pour but d'améliorer la précision de l'estimation.

Exemple :

Si on souhaite évaluer le niveau de satisfaction des élèves par rapport à la cantine scolaire. L'école est composée de 600 élèves répartis en trois niveaux :

- **Primaire** : 300 élèves (50 %)
- **Collège** : 200 élèves (33,3 %)
- **Lycée** : 100 élèves (16,7 %)

Étapes de l'échantillonnage stratifié :

c.1. Division en strates homogènes :

Les élèves sont regroupés en fonction de leur niveau scolaire (primaire, collège, lycée). Ces groupes constituent les strates.

c.2. Détermination de la taille de l'échantillon :

Si l'on décide d'interroger un échantillon de 60 élèves, la taille de l'échantillon dans chaque strate sera proportionnelle à la taille de la strate dans la population :

- Primaire : $60 \times 0,50 = 30$ élèves
- Collège : $60 \times 0,333 = 20$ élèves
- Lycée : $60 \times 0,167 = 10$ élèves

c.3. Tirage au sort dans chaque strate :

Au sein de chaque niveau (primaire, collège et lycée), les élèves sont sélectionnés aléatoirement pour former l'échantillon final.

Avantages : Cette méthode garantit que chaque niveau scolaire est représenté dans l'échantillon proportionnellement à son importance dans la population. Cela améliore la précision de l'estimation globale et permet d'analyser les résultats de manière plus détaillée, par niveau scolaire.

d. Echantillonnage par grappes

Un tirage au hasard ayant donné des grappes (groupes d'individus), on examine tous les individus de chaque grappe tirée. Cette méthode est utile si on ne dispose pas de liste de tous les individus de la population.

Exemple :

Si l'on souhaite analyser les habitudes de lecture des élèves dans une ville.

Étapes de l'échantillonnage par grappes :

d.1. Définir les grappes :

La population totale comprend tous les élèves des écoles de la ville, mais il n'existe pas de liste individuelle des élèves. Les écoles sont donc considérées comme des **grappes**, chaque grappe représentant un groupe d'élèves.

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

d.2. Tirage aléatoire des grappes :

Supposons que la ville compte 50 écoles. Pour former un échantillon, on sélectionne au hasard 5 écoles parmi les 50.

d.3. Étudier tous les individus des grappes sélectionnées :

Une fois les 5 écoles sélectionnées, tous les élèves de ces écoles (par exemple, 2 000 élèves au total) sont inclus dans l'étude.

Avantages :

- Cette méthode est pratique lorsque l'on n'a pas de liste individuelle complète de la population.
- Elle réduit les coûts et la logistique, car les données sont collectées dans des groupes localisés.

Inconvénient :

Les individus dans une même grappe peuvent avoir des caractéristiques similaires, ce qui peut réduire la diversité de l'échantillon et introduire un biais si les grappes ne sont pas représentatives de la population entière.

e. Échantillonnage par degrés

L'échantillonnage par degrés, également appelé échantillonnage en cascade, est une méthode de sélection dans laquelle l'échantillon n'est pas directement tiré parmi les individus de la population, mais plutôt en plusieurs étapes ou degrés. Cette approche est souvent utilisée lorsque la population est trop vaste ou difficile à atteindre directement, nécessitant ainsi une sélection progressive et hiérarchique. Chaque étape du processus permet de réduire progressivement le champ des individus à étudier, ce qui facilite la collecte des données.

Exemple :

Si l'on s'intéresse aux habitudes de lecture des écoliers dans une région, l'échantillonnage pourrait se dérouler en trois étapes :

e.1. Sélection des villes : On tire au hasard plusieurs villes dans la région.

e.2. Sélection des écoles : Dans chaque ville tirée, on choisit au hasard quelques écoles.

e.3. Sélection des écoliers : Enfin, on sélectionne des écoliers au hasard dans chaque école choisie.

Cette méthode permet de réduire les coûts et de simplifier la logistique en procédant par étapes successives.

f. Échantillonnage à probabilité inégale

L'échantillonnage à probabilité inégale consiste à tirer des individus avec des probabilités qui varient en fonction d'un caractère connu, tel que la taille d'une sous-population ou l'importance d'une caractéristique spécifique. Cette méthode est utilisée lorsque certains groupes de la population sont sur-représentés ou sous-représentés, et que l'on souhaite ajuster les probabilités de sélection en conséquence.

Exemple :

Imaginons que l'on souhaite étudier la taille des étudiants par sexe dans une université, et que l'on sait a priori qu'il y a plus de garçons que de filles dans cette université. Dans ce cas, l'échantillon sera constitué de manière à tirer un nombre plus élevé de garçons par rapport aux filles, en fonction de la proportion exacte de chaque sexe dans la population totale. Cela permet d'obtenir des résultats plus représentatifs de la structure réelle de la population.

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

1.2.2. Echantillonnages empiriques

L'échantillonnage empirique est caractérisé par la construction de l'échantillon par un choix « raisonné ». Cet échantillon ressemble le plus possible à la population, mais cette ressemblance est due aux connaissances préalables qu'on a sur cette population. On ne peut calculer la précision de cet échantillonnage. Un exemple d'échantillonnage empirique est fourni par ce qu'on appelle la méthode des quotas. Cette méthode se base sur la construction d'un échantillon de taille n dans lequel les proportions des individus selon différents critères sont égales à celle de la population. Ces proportions dans la population sont connues (par un recensement par exemple). Une fois ces quotas déterminés, on est libre quant au choix de l'échantillon (à condition de respecter les quotas).

2. Les sources statistiques

La source statistique fait référence à l'origine des données utilisées dans les analyses statistiques. Elle est cruciale, car elle garantit la transparence et la fiabilité des informations présentées. En indiquant clairement la source, il devient possible pour les chercheurs, les analystes ou toute autre personne intéressée de vérifier l'exactitude des données, d'évaluer leur pertinence et de les approfondir selon leurs objectifs spécifiques. Une source statistique bien définie permet également d'assurer la reproductibilité des résultats et d'éviter les biais ou erreurs liés à des informations non vérifiées.

3. Différence entre la statistique et les statistiques

3.1. La statistique

La statistique désigne une méthode scientifique qui repose sur des méthodes systématiques pour collecter, organiser, résumer et analyser des données, qu'elles soient qualitatives ou quantitatives. Elle permet d'extraire des informations utiles et pertinentes à partir de données brutes, afin de prendre des décisions éclairées, de formuler des hypothèses ou de prédire des tendances futures.

3.2. Les statistiques

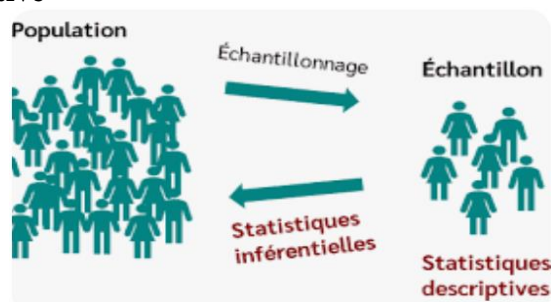
Les statistiques, en revanche, désignent un ensemble de données numériques collectées à partir de diverses sources, telles que les recensements de population, les registres d'état civil ou les enquêtes. Elles peuvent couvrir une large gamme de domaines, comme l'économie, où l'on retrouve des indicateurs relatifs aux prix, aux ventes d'entreprises ou aux taux de croissance économiques. Les statistiques servent ainsi à représenter des phénomènes réels et à fournir des bases factuelles pour l'analyse et la prise de décision.

4. Les concepts de base de la statistique descriptive

4.1. La population statistique

On appelle population statistique l'ensemble des individus sur lesquels porte l'étude statistique.

4.2. L'échantillon : un échantillon est une partie ou un sous ensemble d'individus représentatifs d'une population.



Exemple : Les étudiants de l'université de Bejaia.

La population statistique : l'ensemble des étudiants de l'université de Bejaia

4.3. L'individu (unité statistique)

On appelle individu chaque élément de la population ou de l'échantillon. On utilise également le terme **unité statistique** pour désigner un individu.

Exemple : Dans l'exemple précédent portant sur les étudiants de l'université de Bejaia, chaque **étudiant** de l'université représente un **individu** ou une **unité statistique**.

4.4. Le caractère ou la variable

Le caractère est l'aspect particulier que l'on veut étudier. Pour décrire les individus qui composent une population ou un échantillon, nous pouvons nous intéresser à leur âge, leur sexe, leur taille, leur religion, leur nationalité, etc. Il existe deux types de caractères : le caractère quantitatif et le caractère qualitatif.

a. Le caractère quantitatif

On dit d'un caractère qu'il est quantitatif si ses modalités sont mesurables c'est-à-dire elles s'expriment par un nombre.

Exemple : L'âge, le poids et le nombre de voitures sont des caractères quantitatifs.

En raison de la nature des valeurs que peut prendre un caractère quantitatif dans une série statistique, nous distinguons deux catégories : le caractère quantitatif discret (discontinu) et le caractère quantitatif continu.

• Le caractère quantitatif discret (discontinu)

Un caractère quantitatif est discontinu si l'ensemble des valeurs qu'il peut prendre est fini ou dénombrable. Lorsque ces valeurs sont entières et isolées, on dira que le caractère est discret.

Exemple : Le nombre d'enfants par famille, le nombre de voitures, le nombre d'étudiants et le nombre de tables par classe sont des caractères quantitatifs discrets.

• Le caractère quantitatif continu

Un caractère quantitatif est continu s'il peut prendre n'importe quelle valeur dans un intervalle donné de nombres réels.

Exemple : Le poids, la taille et l'âge sont des caractères quantitatifs continus.

b. Le caractère qualitatif

Un caractère est considéré comme qualitatif lorsque son observation ne peut pas être exprimée par une mesure. Ses différentes modalités sont simplement observées et décrites à l'aide de mots représentant un état, tels que la religion, le sexe, le lieu de résidence ou l'état de santé. Il peut être classé selon deux types d'échelles : l'échelle nominale et l'échelle ordinale.

• Le caractère qualitatif nominal

Il décrit un nom ou une catégorie qui ne suit pas un ordre naturel.

Exemple : Le mode de transport utilisé par les habitants d'une ville :

Voiture personnelle, transport en commun, bicyclette, avion, autres moyens.

• Le caractère qualitatif ordinal

Un caractère qualitatif est ordinal si ses modalités peuvent être classées dans un ordre spécifique ou dans un ordre naturel quelconque.

Exemple :

Classement de l'état de santé à la suite d'une visite médicale :

Excellent, Très bon, Bon, Mauvais, Très mauvais.

4.5. Les modalités

Les modalités sont les différentes situations dans lesquelles les individus peuvent se trouver à l'égard du caractère considéré.

Exemple1 :

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

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

Exemple2 :

Pour étudier la satisfaction des étudiants de l'université de Bejaia concernant leurs dernières vacances d'été, les modalités pourraient être « peu », « modérément », « beaucoup » ou « énormément ».

Exemple3 : Si le caractère retenu ou la variable étudiée est le nombre d'enfants par familles, les modalités de ce caractère peuvent être 0,1, 2, 3, etc.

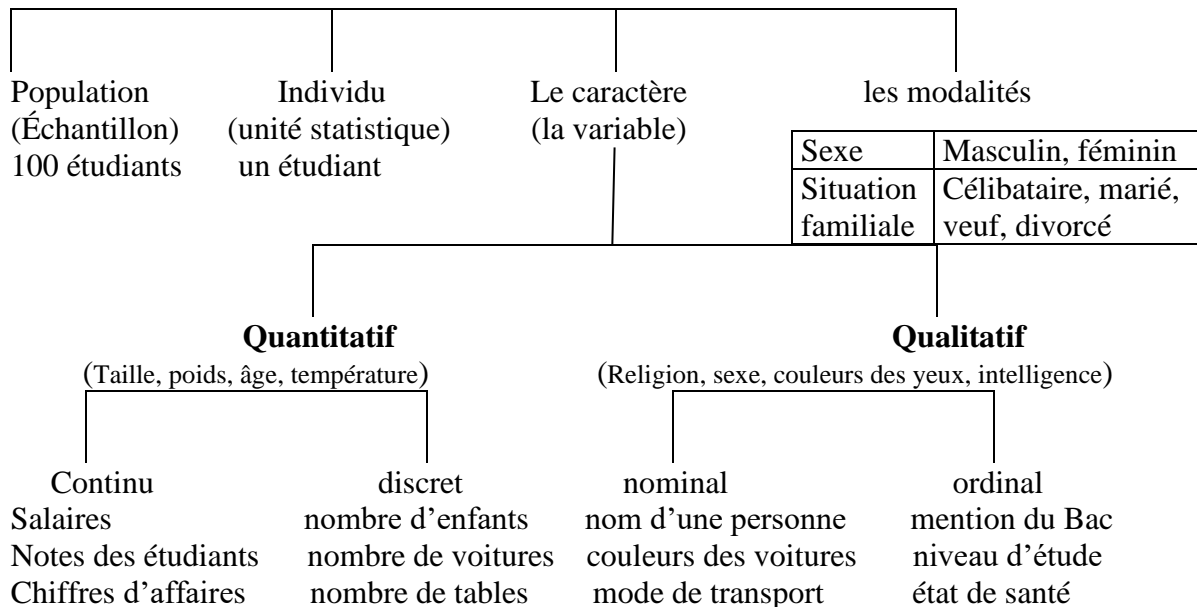
4.6. La série statistique

Une série statistique est une liste de mesures obtenues généralement lors d'une étude ou de relevés de mesures.

Exemple1 : La série suivante provient d'une enquête auprès de quelques personnes pour connaître leur taille en cm.

169 170 180 179 167 158 160 170 175
170 160 158 155 163 172 162 185

4.7. Récapitulatif des concepts de base de la statistique descriptive



Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

5. Exercices corrigés

Exercice 1 :

Les populations suivantes sont-elles bien définies en vue d'une étude statistique :

- a- Le personnel de la banque BNA ?
- b- Les individus de nationalité algérienne ?
- c- Les habitants de Bejaia ?
- d- Les petites annonces déposées sur le panneau d'affichage du quartier.

Corrigé exercice 1 :

a. Le personnel de la banque BNA :

Oui, cette population est bien définie. Elle se compose de toutes les personnes travaillant à la banque BNA, ce qui constitue un groupe homogène et spécifique. Il est facile d'identifier les membres de cette population, et la collecte de données peut être réalisée de manière précise.

b. Les individus de nationalité algérienne :

Non, cette population n'est pas bien définie dans le cadre d'une étude statistique. La nationalité seule ne définit pas d'autres critères spécifiques qui permettraient de limiter ou de décrire davantage la population (âge, lieu de résidence, statut social, etc.). De plus, la population des "individus de nationalité algérienne" peut inclure des personnes résidant à l'étranger, ce qui complique la collecte de données.

c. Les habitants de Bejaia :

Oui, cette population est bien définie. Elle se réfère à toutes les personnes vivant dans la ville de Bejaia, ce qui permet de déterminer clairement les individus à inclure dans l'étude. La délimitation géographique de la ville rend cette population précise et facilement identifiable.

d. Les petites annonces déposées sur le panneau d'affichage de la grande surface A :

Non, cette population n'est pas bien définie. Les petites annonces représentent un ensemble d'informations variées, mais elles ne concernent pas un groupe spécifique de personnes. Elles peuvent couvrir différents sujets, types d'annonces, et auteurs, ce qui rend cette population peu homogène et difficile à analyser de manière statistique.

Exercices 2 :

Pour les sujets d'étude qui suivent, spécifier : la population statistique, l'individu (unité statistique), le caractère et sa nature.

- Étude de la durée de vie des lampes électriques ;
- Étude de l'absentéisme des ouvriers, en jours, dans une usine ;
- Répartition des étudiants selon la mention obtenue sur leur diplôme ;
- Répartition des 20 ouvriers selon leur lieu de résidence ;
- Répartition de 100 familles selon le nombre de filles ;
- Répartition d'un groupe de 30 étudiants selon leur âge.

Corrigé exercice 2 :

La population statistique	L'individu	Le caractère	Sa nature
L'ensemble des lampes	Une lampe	La durée de vie d'une lampe	Quant. continu
L'ensemble des ouvriers	Un ouvrier	Le nombre de jours d'absences	Quant. discret
L'ensemble des étudiants	Un étudiant	La mention du diplôme	Qual. ordinal
Les 20 ouvriers	Un ouvrier	Lieu de résidence	Qual. nominal
Les 100 familles	Une famille	Le nombre de filles par famille	Quant. discret
Les 30 étudiants	Un étudiant	L'âge d'un étudiant	Quant. continu

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

Exercice 3 :

Dans une étude portant sur la consommation d'énergie des ménages dans la ville de Bejaia, un chercheur souhaite analyser l'impact des équipements électroménagers sur la consommation d'électricité. Il choisit de collecter des données auprès de 200 ménages sélectionnés au hasard parmi les 5000 ménages recensés dans la ville.

1. Définir la population statistique de cette étude.
2. Quel est l'échantillon dans cette étude ?
3. Quels sont les individus (ou unités statistiques) dans cet exemple ?
4. Donner deux exemples de caractères que l'on pourrait observer dans cette étude.

Corrigé de l'exercice 3 :

1. **La population statistique** : est l'ensemble des 5000 ménages recensés dans la ville de Bejaia.
2. **L'échantillon** : est constitué des 200 ménages sélectionnés au hasard parmi les 5000 ménages de la ville.
3. **Les individus (unités statistiques)** : sont les ménages.
4. **Deux exemples de caractères que l'on pourrait observer sont :**
 - **Caractère quantitatif discret** : Le nombre d'appareils électroménagers dans chaque ménage.
 - **Caractère qualitatif nominal** : Le type de logement (appartement, maison individuelle).

Exercice 4 :

Dans une enquête sur la satisfaction des étudiants de l'université de Bejaia concernant les infrastructures sportives, les chercheurs collectent des données sur les éléments suivants : l'état des équipements sportifs, la fréquence d'utilisation des installations et le type de sport pratiqué.

1. Quels sont les caractères observés dans cette enquête ? Indiquez si chaque caractère est qualitatif ou quantitatif.
2. Si le caractère "fréquence d'utilisation des installations" est quantitatif, est-il discret ou continu ? Justifiez votre réponse.
3. Pour le caractère "type de sport pratiqué", donnez ses modalités.
4. Si l'on classe les étudiants par "niveau de satisfaction" (Très satisfait, Satisfait, Peu satisfait, Pas satisfait), ce caractère est-il ordinal ou nominal ?

Corrigé exercice 4 :

1. **Les caractères observés dans cette enquête sont les suivants :**
 - **L'état des équipements sportifs** : C'est un caractère **qualitatif ordinal**, car il peut être décrit par des mots (par exemple : bon, moyen, mauvais).
 - **La fréquence d'utilisation des installations** : C'est un caractère **quantitatif**, car il peut être mesuré par des nombres (par exemple : nombre de fois par semaine).
 - **Le type de sport pratiqué** : C'est un caractère **qualitatif nominal**, car il peut être décrit par des catégories ou des noms (par exemple : football, basket-ball, tennis).

Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base

2. Le caractère "fréquence d'utilisation des installations" est discret. En effet, il s'agit d'un nombre de fois par semaine (ou une autre unité de temps), et ce nombre est compté en valeurs entières, donc il est dénombrable et ne prend pas toutes les valeurs possibles dans un intervalle continu. Par exemple, on peut avoir 1, 2, 3 visites, mais pas 2,5 visites.

3. Les modalités du caractère "type de sport pratiqué" peuvent être :

- Football, Basket-ball, Tennis, Natation, Volleyball, Autres sports

Ces modalités représentent les différentes catégories de sports pratiqués par les étudiants.

4. Le caractère "niveau de satisfaction" est ordinal. En effet, les modalités ("Très satisfait", "Satisfait", "Peu satisfait", "Pas satisfait") peuvent être classées dans un ordre spécifique, allant du plus élevé au plus bas. Ce classement reflète une hiérarchie des niveaux de satisfaction, ce qui caractérise un caractère ordinal.

Exercice 5 :

Dans un recensement réalisé dans une école primaire, les données suivantes sont collectées pour chaque élève : âge, sexe, nombre de frères et sœurs, niveau de santé, et mode de transport utilisé pour se rendre à l'école.

1. Listez les caractères observés dans cette enquête. Pour chaque caractère, indiquez s'il est qualitatif ou quantitatif.
2. Pour le caractère "âge", est-ce un caractère quantitatif discret ou continu ?
3. Donnez des exemples de modalités pour le caractère "mode de transport".
4. Le caractère "niveau de santé" peut-il être ordonné ? Si oui, classifiez les modalités possibles.

Corrigé de l'exercice 5 :

1. Les caractères observés dans cette enquête sont :

- **Âge** : C'est un caractère quantitatif car il peut être mesuré en années ou fractions d'années.
- **Sexe** : C'est un caractère qualitatif nominal, car il ne suit pas un ordre naturel (par exemple : garçon, fille).
- **Nombre de frères et sœurs** : C'est un caractère quantitatif discret, car il correspond à un nombre entier et dénombrable.
- **Niveau de santé** : C'est un caractère qualitatif ordinal, car il peut être classé par ordre (par exemple : excellent, bon, moyen, mauvais).
- **Mode de transport utilisé pour se rendre à l'école** : C'est un caractère **qualitatif nominal**, car il désigne des catégories sans ordre particulier (par exemple : à pied, en voiture, en bus).

2. Le caractère âge est un caractère quantitatif continu. En effet, l'âge peut prendre toutes les valeurs possibles dans un intervalle, comme 8 ans, 8,5 ans, ou 9,2 ans, et il peut être exprimé en fractions.

3. Les modalités possibles pour le caractère "mode de transport" sont :

- À pied, En voiture, En bus scolaire, En vélo, En moto.

Ces modalités représentent les différentes façons dont les élèves se rendent à l'école.

4. Oui, le caractère "niveau de santé" peut être ordonné car il peut être classé par ordre croissant ou décroissant. Exemple de classification des modalités possibles :

- Excellent, Bon, Moyen, Mauvais, Très mauvais.

Introduction

Les données d'une série statistique, résultant de l'étude d'un caractère donné, peuvent être présentées principalement de deux façons : sous forme de tableau ou sous forme de graphique. Ces deux méthodes de représentation offrent des moyens efficaces de structurer et de synthétiser les informations, tout en rendant leur analyse et leur interprétation plus simples et accessibles. Elles permettent ainsi de mettre en évidence les tendances, les relations ou les caractéristiques clés des données étudiées de manière claire et concise.

1. Présentation de données sous forme de tableaux :

Un tableau statistique est un tableau constitué initialement de deux colonnes :

- La première colonne correspond aux classes, aux valeurs, aux modalités selon leurs caractères.
- La deuxième colonne correspond aux effectifs (n_i) des classes ou des valeurs, si non elle correspond aux fréquences (f_i)
- Le regroupement des informations chiffrées dans un tableau (ou graphe) exige les éléments suivants : Le titre du tableau, l'unité de mesure et la source des informations recueillies.

Il faut rappeler que le caractère peut être de type qualitatif ou quantitatif.

1.1. Présentation des données d'un caractère qualitatif

Pour présenter les résultats observés, nous identifions d'abord les différentes modalités du caractère étudié, c'est-à-dire les différentes valeurs de la variable considérée, notées par : $X_1, X_2, X_3, \dots, X_k$.

Il s'agit donc d'un caractère ayant k modalités. Ensuite, nous déterminons le nombre d'individus correspondant à chaque modalité, appelé effectif absolu (n_i). Ils sont identifiés par $n_1, n_2, n_3, \dots, n_k$.

Nous pouvons aussi évaluer l'effectif relatif de chacune des modalités. On parlera alors de fréquences relatives. Ces dernières sont représentées par f_1, f_2, \dots, f_k et obtenues grâce au rapport suivant :

Nous pouvons également calculer l'effectif relatif de chaque modalité, que l'on appelle fréquences relatives. Elles sont notées f_1, f_2, \dots, f_k et sont obtenues à partir du rapport suivant : $f_i = \frac{n_i}{N}$

Avec N : effectif total ($N = \sum_{i=1}^k n_i = n_1 + n_2 + n_3 + \dots + n_k$)

L'effectif relatif (ou fréquence) peut aussi être exprimé en pourcentage en multipliant la fréquence relative de chaque modalité par 100.

Exemple :

Une enquête réalisée auprès de 10 passants pour connaître leur moyen de transport préféré a donné la série statistique suivante :

Bus, voiture, bus, vélo, voiture, marche, voiture, bus, voiture, vélo.

Pour présenter les résultats observés, nous commençons par identifier les différentes modalités du caractère en question.

Dans ce cas, le caractère étudié est le moyen de transport, notée X_i . Nous déterminons ensuite le nombre d'individus associés à chacune des valeurs de cette variable, notés n_i .

Enfin, un tableau est construit pour résumer toutes ces informations.

Chapitre 2 : Représentation de données statistiques

T.1 La répartition des 10 passants selon leur moyen de transport

Moyen de transport préféré	Effectifs Absolus n_i	Effectifs Relatifs $f_i = n_i/N$	Effectifs Relatifs $f_i(\%)$
Bus	3	0,3	30
Voiture	4	0,4	40
Vélo	2	0,2	20
Marche	1	0,1	10
Total	10	1	

Source : enquête personnelle

1.2. Présentation des données d'un caractère quantitatif

En raison de la nature des valeurs que peut prendre un caractère quantitatif dans une série statistique, nous distinguons 2 catégories : le caractère quantitatif discret et le caractère quantitatif continu.

a. Caractère quantitatif discret

Un caractère quantitatif est discontinu si l'ensemble des valeurs qu'il peut prendre est fini ou dénombrable. Lorsque ces valeurs sont entières, on dira que le caractère est discret.

Exemple 1 : la distribution de fréquences suivante représente la répartition de 100 familles selon le nombre de fille ($0 \leq X_i \leq 5$).

T.2 : La répartition de 100 familles selon le nombre de filles

Nombre de filles X_i	n_i	$f_i \left(\frac{n_i}{N} \right)$
0	15	15/100=0,15
1	35	35/100=0,35
2	25	25/100=0,25
3	10	10/100=0,10
4	05	05/100=0,05
5	10	10/100=0,10
Total	100	1,00

Exemple 2 :

La série suivante résulte d'une enquête auprès de 30 étudiants pour connaître leur âge :

20 18 21 19 19 18 18 22 20 19 17 18 23 19 21
18 18 19 17 18 19 22 21 21 20 18 17 23 20 22

T.3 : La répartition d'un groupe de 30 étudiants selon leur âge (en années)

Âge (en années)	n_i	f_i	$f_i(\%)$
17	3	0,1	10
18	8	0,26	27
19	6	0,2	20
20	4	0,13	13
21	4	0,13	13
22	3	0,1	10
23	2	0,06	7
Total	30	0,98 ≈ 1	100

Chapitre 2 : Représentation de données statistiques

Remarque :

Il est à noter que la somme des fréquences relatives ne s'élève pas toujours à 1 (ou à 100 en %). Cela est dû aux calculs effectués lors de la détermination des fréquences relatives. En pratique, comme illustré dans le tableau précédent, il suffit d'ajuster la valeur de la fréquence relative d'une ou plusieurs valeurs tout en gardant à l'esprit l'objectif d'atteindre un total de 100.

Reprenons les fréquences relatives du tableau précédent. Nous obtiendrons donc :
10% 27% 20% 13% 13% 10% 7%

La somme de ces fréquences relatives est bien égale à 100 (voir la colonne 4).

b. Caractère quantitatif continu

Un caractère quantitatif est continu s'il peut prendre n'importe quelle valeur dans un intervalle donné de nombres réels.

Par convention : l'intervalle est fermé à gauche et ouvert à droite.

Désignons par e l'extrémité de classe. La i ème classe est alors notée « e_{i-1}, e_i » correspondant à l'intervalle $[e_{i-1} - e_i [$.

L'amplitude a_i d'une classe est la largeur de l'intervalle de cette classe. La i ème classe aura comme amplitude : $a_i = e_i - e_{i-1}$

Exemple :

L'amplitude de la classe $[2 - 6 [$ est égale à : $a_i = 6 - 2 = 4$

Les amplitudes des classes peuvent être égales ou inégales sur l'ensemble de l'intervalle de variation de la variable statistique. Il peut également arriver que les classes extrêmes ne soient pas délimitées.

Par convention :

Pour une classe non délimitée, on utilise comme amplitude celle de la classe adjacente (ou classe voisine).

Étant donné que la variable statistique ne prend plus une valeur précise x_i , mais un ensemble de valeurs comprises dans un intervalle, appelé classe, il est nécessaire de calculer le centre de classe, noté c_i . Ce centre correspond à la moyenne arithmétique des valeurs des extrémités de la classe.

D'où : $c_i = \frac{e_{i-1} + e_i}{2}$, Par exemple le centre de classe de : $[2 - 6 [$ est égale à : $c_i = \frac{2+6}{2} = 4$

Exemple1 :

Le système de pointage dans une entreprise publique a permis d'enregistrer les temps de retard (exprimés en minutes) de la part des ouvriers :

T.4 : Répartition des 140 ouvriers selon leur temps de retard

Temps de retard (mn)	n_i	Centre de classe c_i	Fréquences relatives $f_i\%$
$[4 - 8 [$	5	$(4 + 8)/2 = 6$	3,57
$[8 - 12 [$	20	$(8 + 12)/2 = 10$	14,29
$[12 - 16 [$	15	$(12 + 16)/2 = 14$	10,71
$[16 - 20 [$	60	$(16 + 20)/2 = 18$	42,86
$[20 - 24 [$	25	$(20 + 24)/2 = 22$	17,86
$[24 - 28 [$	15	$(24 + 28)/2 = 26$	10,71
Total	140	-	100,00

Chapitre 2 : Représentation de données statistiques

Exemple2 :

Soit le caractère « nombre d'années d'utilisation d'un certain type d'appareils ». Les modalités retenues sont résumées dans le tableau ci-dessous :

T.5 : Tableau de répartition

e_i	Amplitude (a_i)	Intervalle utilisé	Centre de classe c_i
- 5	5	[0 - 5[2,5
[5 - 10[5	[5 - 10[7,5
[10-20[10	[10-20[15
[20-25[5	[20-25[22,5
[25-30[5	[25-30[27,5
[30-50[20	[30-50[40
+ 50	20	[50-70[60

Nous donnons à la classe « moins de 5 » une amplitude de 5 car la classe adjacente est la classe « 5-10 » d'amplitude égale à 5.

De même, nous affectons pour la classe « plus de 50 » une amplitude de 20 car la classe adjacente est la classe « 30-50 » d'amplitude 20.

2. Représentation graphique des séries statistiques

Considérons une population d'un effectif total N que nous souhaitons étudier en fonction d'un certain caractère K . Nous désignons par n_i l'effectif associé à chaque modalité de ce caractère. Comme mentionné précédemment, les résultats peuvent être présentés sous forme de tableaux. Mais, il est aussi possible d'utiliser des méthodes de représentation graphique. Celles-ci diffèrent selon le caractère qualitatif ou quantitatif de la variable étudiée.

2.1. Cas d'un caractère qualitatif

Dans ce cas, deux types de représentation graphique sont utilisés :

- Représentation cartésienne avec les **diagrammes en colonnes**
- Représentation avec le **diagrammes circulaire**

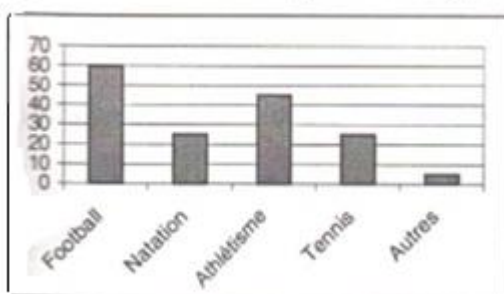
Exemple1 :

Une enquête auprès de 160 jeunes au sujet de leur sport préféré a fourni les résultats suivants :

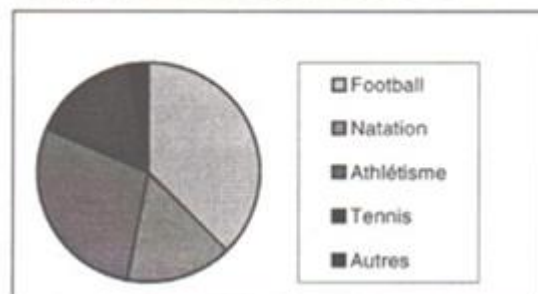
T.6 : La répartition des 160 jeunes selon leur sport préféré

Sport préféré X_i	Nombre de jeunes n_i	f_i	$f_i \times 360^0$
Football	60	0,375	135
Natation	25	0,156	56.25
Athlétisme	45	0,281	101.25
Tennis	25	0,156	56.25
Autres	5	0,031	11.25
Total	160	1	360

Représentation en tuyaux d'orgue



Représentation par secteurs



Chapitre 2 : Représentation de données statistiques

2.2. Cas d'un caractère quantitatif

Nous distinguons 2 catégories de caractères : le caractère quantitatif discret et le caractère quantitatif continu.

a. Cas d'un caractère quantitatif discret (discontinu) :

Un caractère quantitatif est dit discret s'il prend des valeurs isolées et entières.

Exemple :

La répartition d'un échantillon de $N=100$ familles de 5 enfants en fonction du nombre de filles ($0 \leq X_i \leq 5$).

T.7 : Répartition des 100 familles selon le nombre de filles

Nombre de filles	Effectifs (n_i)	Fréquences relatives f_i (n_i/N)	Effectifs cumulés croissants	Effectifs cumulés décroissants
			$n_i \uparrow$	$n_i \downarrow$
0	15	0.15	15	100
1	35	0.35	50	85
2	25	0.25	75	50
3	10	0.10	85	25
4	5	0.05	90	15
5	10	0.10	100	10
Total	100	1.00	-	-

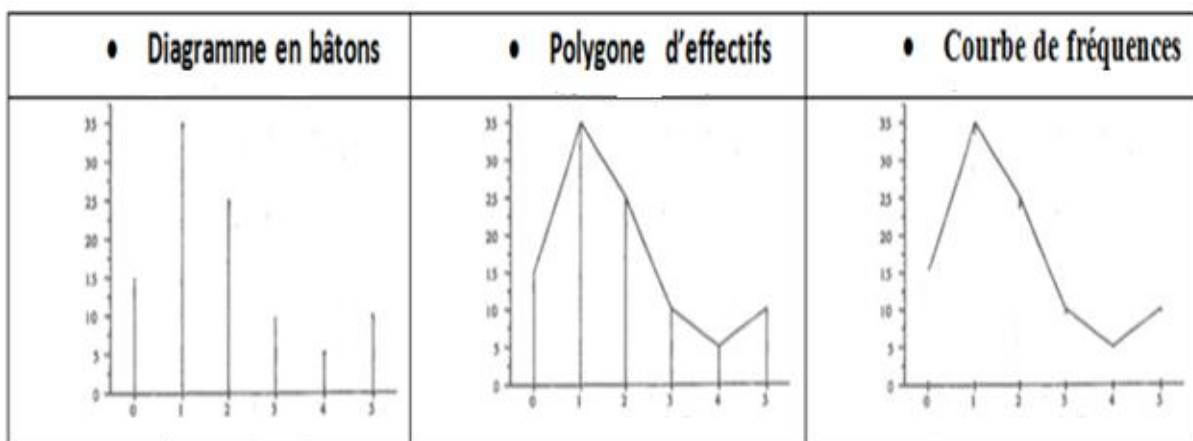
Signification :

Les effectifs cumulés indiquent le nombre d'individus pour lesquels la variable statistique est inférieure à X_{i+1}

Exemple :

15 correspond au nombre de familles ayant eu moins d'une fille, c'est-à-dire aucune fille. De même, 50 indique le nombre de familles ayant eu moins de 2 filles.

Dans le cas d'une série discontinue, il existe plusieurs façons de représenter graphiquement les résultats. Nous rappellerons ici les formes les plus utilisées :



A chaque valeur de la variable X_i portée en abscisse, on fait correspondre un bâton de longueur proportionnelle à l'effectif n_i correspondant à cette valeur.

Chapitre 2 : Représentation de données statistiques

- **Courbe cumulative croissante et décroissante (diagrammes en escaliers)**

Il est également possible de construire des diagrammes en « escaliers » qui représentent les fréquences cumulées d'une série statistique en fonction des valeurs de la variable.

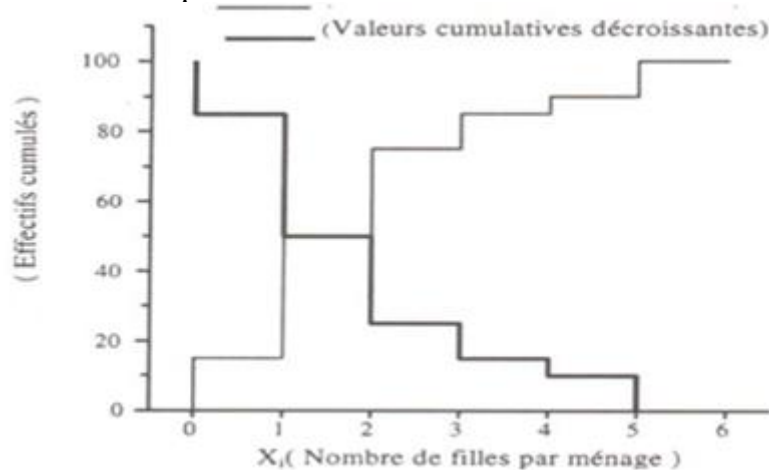
Diagramme ascendant :

À chaque valeur X_i , on monte une marche de l'escalier, dont la hauteur correspond à l'effectif n_i associé à x_i . Chaque palier compris entre deux valeurs consécutives de la variable statistique x_i et x_{i+1} .

Ce diagramme représente en effet l'effectif cumulé n_i^\uparrow pour lequel la valeur prise par la variable statistique est inférieure.

Diagramme descendant :

Le principe est le même que ci-dessus



b. Cas d'un caractère quantitatif continu :

Remarque :

Dans la réalité, une variable n'est pas toujours continue. Le degré de précision de la mesure impose souvent une discontinuité. Or, lorsque la précision de la mesure introduit une série trop grande de valeurs possibles, la présentation dans un tableau de ces valeurs possibles de X_i rendrait celui-ci très difficile à lire. Il convient donc de grouper ces valeurs dans des classes convenablement choisies.

- **Histogramme :** deux cas sont distingués en fonction des amplitudes des classes :

- **Cas d'amplitudes égales :**

Dans le cas où les amplitudes des classes sont égales, il n'est pas nécessaire de rectifier les effectifs, car les hauteurs des rectangles de l'histogramme sont directement proportionnelles aux effectifs.

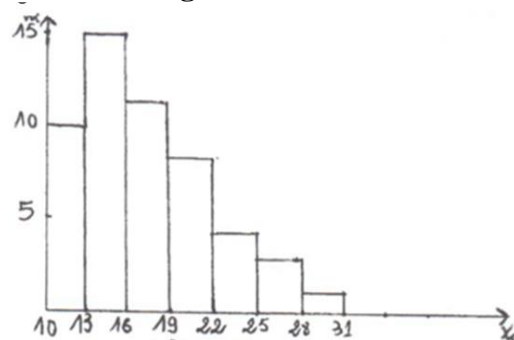
Exemple :

Le tableau suivant présente la répartition des salaires horaires des travailleurs.

T.8 : Distribution des salaires horaires des travailleurs

salaires horaires	n_i
[10-13[10
[13-16[15
[16-19[12
[19-22[07
[22-25[03
[25-28[02
[28-31[01
Total	50

Histogramme



Chapitre 2 : Représentation de données statistiques

➤ Cas d'amplitudes inégales

Si les amplitudes des classes sont inégales, il faut d'abord corriger l'effectif avant de tracer l'histogramme. Pour cela, on applique la formule suivante :

$$n_{ic} = n_i \times \frac{\text{amplitude la plus petite}}{\text{amplitude de la classe considérée}} \quad (n_{ic} \text{ est l'effectif corrigé})$$

Exemple :

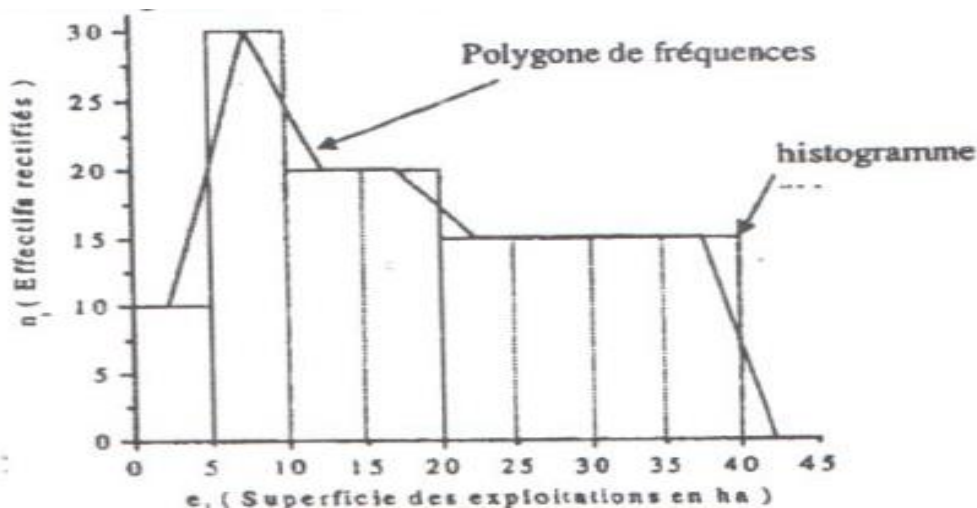
Le tableau suivant donne la répartition des exploitations agricoles selon leur superficie en hectare.

T.9 : Distribution de 140 exploitations agricoles selon leur superficie (en hectare)

Superficie (Hectare)	n_i	a_i	n_{ic}	$n_i \uparrow$	$n_i \downarrow$
[0 – 5[10	5	10	10	140
[5 –10[30	5	30	40	130
[10-20[40	10	$n_{ic} = \frac{40 \times 5}{10} = 20$	80	100
[20-40[60	20	$n_{ic} = \frac{60 \times 5}{20} = 15$	140	60
Total	140	-	-	-	-

Étant donné que les amplitudes des classes sont inégales, il est nécessaire de corriger les effectifs. Cette correction est partielle, car la troisième classe a une amplitude qui est le double de celles de la première et de la deuxième classe, qui ont les plus petites amplitudes. De même, la dernière classe présente une amplitude quadruple par rapport à celles de la première et de la deuxième classe.

L'histogramme sera alors le suivant :



Attention : les effectifs rectifiés ne sont utilisés que pour la construction de l'histogramme.

• Courbe cumulative croissante et décroissante

Dans ce cas, nous traçons une courbe dont seuls quelques points sont connus.

* **Pour la courbe croissante**, on joint les points de coordonnées $(e_i, n_i \uparrow)$ c'est-à-dire :

- En abscisse, la valeur de l'extrémité droite e_i de la classe i
- En ordonnée, la valeur de l'effectif cumulé $n_i \uparrow$ correspond cette classe.

* **Pour la courbe décroissante**, on joint les points de coordonnées $(e_{i-1}, n_i \downarrow)$ c'est-à-dire :

- En abscisse, la valeur de l'extrémité gauche e_{i-1} la classe i ;
- En ordonnée, la valeur de l'effectif cumulé $n_i \downarrow$ correspond cette classe.

Chapitre 2 : Représentation de données statistiques

Exemple 1 :

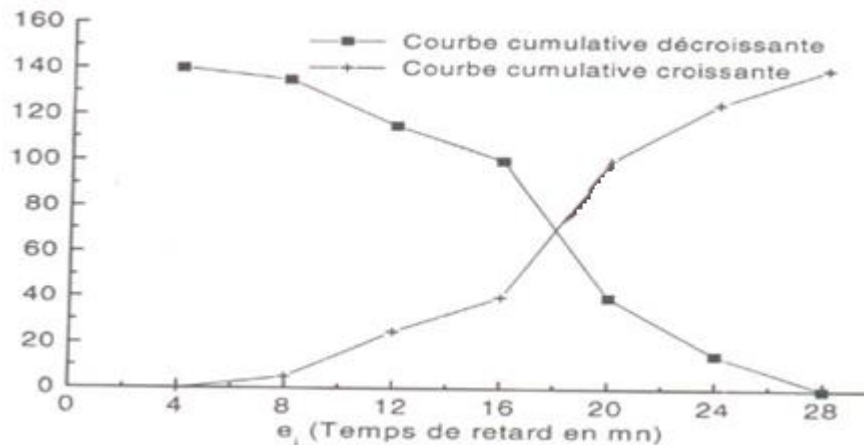
Cas d'une variable dont les classes sont égales.

Reprenons le tableau (T.4) indiquant la répartition des 140 ouvriers selon leur temps de retard (le retard est exprimé en minutes)

T.10 : La répartition des 140 ouvriers selon leur temps de retard

Temps de retard	n_i	$n_i \uparrow$	$n_i \downarrow$
[4 - 8[5	5	140
[8 - 12[20	25	135
[12-16[15	40	115
[16-20[60	100	100
[20-24[25	125	40
[24-28[15	140	15
Total	140	-	-

La courbe cumulative croissante et décroissante



Remarque :

La même chose pour le cas d'une variable dont les classes sont inégales on trace les courbes de fréquences cumulées sans corriger l'effectif.

Attention :

C'est une courbe que l'on trace et non une succession de segments de droite. Par ailleurs, on remarquera que les courbes croissantes et décroissantes sont symétriques par rapport à un axe parallèle à l'axe des abscisses et ayant comme ordonnée $N/2$.

3.Exercices corrigés

Exercice 1 :

On considère l'ensemble des notes obtenues, lors d'un examen noté sur 20, par 50 candidats

10 8 3 12 13 9 12 9 12 11
11 11 8 5 13 14 14 6 12 16
7 11 10 10 2 15 12 10 1 14
11 7 8 10 13 9 13 9 7 13
11 19 9 4 10 8 9 6 7 14

1. Classer ces données (on prendra les classes suivantes : [0-5[, [5-7[, [7-9[, [9-11[, [11-13[, [13-15[, [15-20[.
2. Déterminer la population statistique, le caractère et sa nature.
3. Déterminer la valeur du mode graphiquement et par le calcul.
4. Calculer les effectifs cumulés ainsi que les fréquences relatives cumulées
5. Calculer la médiane
6. Quelle est le pourcentage des candidats ayant une note inférieure à 9 ?
7. Quelle est le pourcentage des candidats ayant au moins une note égale à 13 ?

Chapitre 2 : Représentation de données statistiques

Corrigé de l'exercice 1 :

Le tableau des effectifs

Notes obtenues	n_i	a_i	n_{ic}	$n_i \uparrow$	$n_i \downarrow$	f_i (%)	$f_i \uparrow$	$f_i \downarrow$
[0 - 5[4	5	1,6	4	50	08	08	100
[5 - 7[3	2	3	7	46	06	14	92
[7 - 9[8	2	8	15	43	16	30	86
[9- 11[12	2	12	27	35	24	54	70
[11-13[11	2	11	38	23	22	76	46
[13-15[9	2	9	47	12	18	94	24
[15-20[3	5	1,2	50	03	06	100	06
Total	50	-	-	-	-	100	-	-

1) Classement des données :

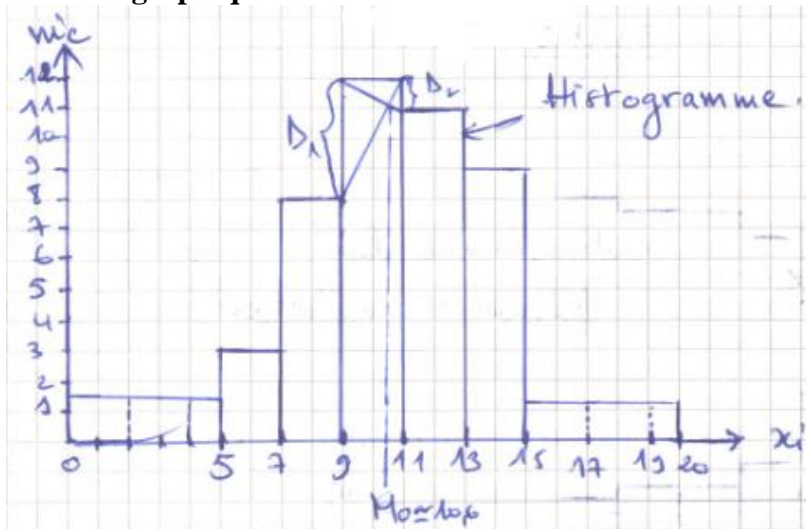
On classe les 50 notes dans les intervalles définis (voir le tableau ci-dessus).

2) La population statistique est : les 50 candidats,

Son caractère : est la note d'un candidat

Son type est : quantitatif continu

3) Le Mode graphiquement



Le Mode analytiquement : la classe modale est : [9- 11[

$$Mo = X_{min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i \quad Mo = 9 + \frac{(12-8)}{(12-8)+(12-11)} \times 2 = 10,6$$

4) Calcul des effectifs cumulés et les fréquences cumulées (voir le tableau ci-dessus).

5) Le Médiane analytiquement :

La classe médiane est : [9 - 11[

$$Me = X_{min} + \frac{N/2 - n_{Me-1}^{\uparrow}}{n_{Me}} \times a_i \quad Me = 9 + \frac{25-15}{12} \times 2 = 10,66$$

6) Proportion des candidats ayant une note inférieure à 9 est 30%, il suffit de lire dans le tableau la fréquence cumulée croissante associée à la classe [7 - 9[

7) Proportion des candidats ayant au moins note égale à 13 est 24%, il suffit de lire dans le tableau la fréquence cumulée décroissante associée à la classe [13 - 15[

Chapitre 2 : Représentation de données statistiques

Exercice 2 :

Dans une bibliothèque, l'ensemble des abonnés a été réparti suivant le nombre d'ouvrages empruntés durant un mois :

X_i	0	1	2	3	4	5	6
$n_i \uparrow$	19	58	115	179	221	254	270

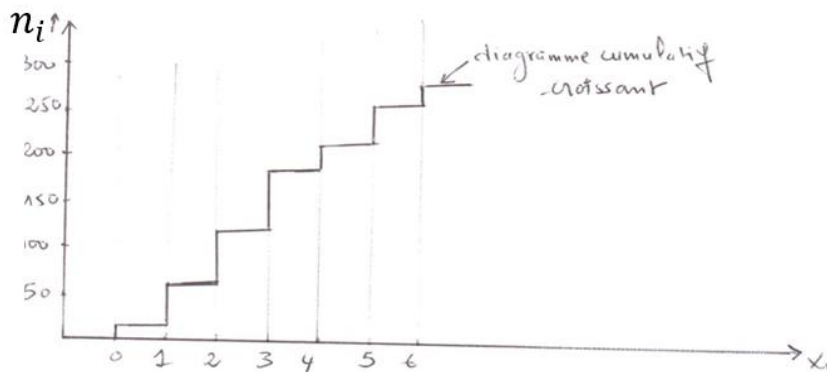
- 1) Déterminer la population, le caractère et sa nature.
- 2) Calculer les effectifs n_i .
- 3) Tracer le diagramme cumulé croissant de cette série.
- 4) Quel est le pourcentage des abonnés dont le nombre d'ouvrages empruntés est :
 - a) Moins de 5 livres ;
 - b) Au plus 3 livres.

Corrigé exercice2 :

Le tableau des effectifs

X_i : le nombre d'ouvrages empruntés	$n_i \uparrow$	n_i
0	19	19
1	58	58-19=39
2	115	115-58=57
3	179	179-115=64
4	221	221-179=42
5	254	254-221=33
6	270	270-254=16
Total	-	270

- 1) **La population statistique est** : les 270 abonnés de la bibliothèque, **son caractère est** : le nombre d'ouvrages empruntés par chaque abonné, **son type est** : quantitatif discret.
- 2) **Calcul des effectifs n_i** (voir le tableau)
- 3) **Le diagramme cumulé croissant**



- 4) **Le pourcentage des abonnés dont le nombre d'ouvrages empruntés est :**

a) Moins de 5 livres : $P = \frac{221}{270} \times 100 = 81.85\%$

b) Au plus 3 livres : $P = \frac{179}{270} \times 100 = 66.29\%$

Chapitre 2 : Représentation de données statistiques

Exercices 3 :

Le tableau ci-dessous donne la répartition des salaires mensuels, en milliers de dinars, des employés d'une entreprise :

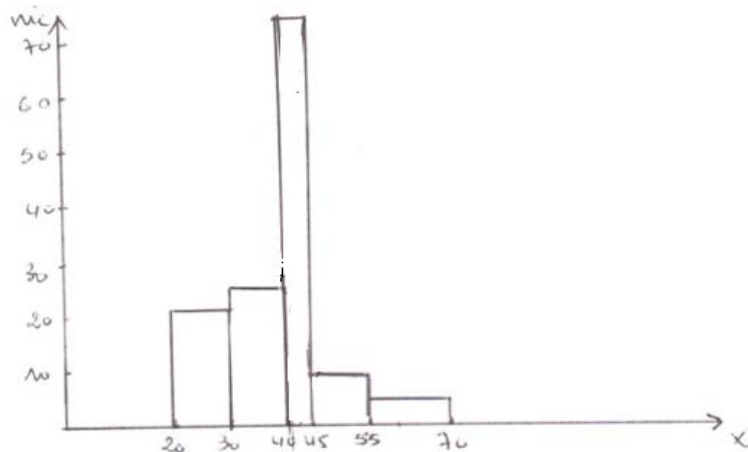
Salaire	[20 - 30[[30 - 40[[40 - 45[[45 - 55[[55 - 70[
Effectif	42	49	74	19	16

- 1) Définir la population statistique, le caractère étudié et sa nature.
- 2) Donner la représentation graphique de cette série.
- 3) Quel est le pourcentage des employés ayant au moins un salaire égal à 40000 DA.

Corrigé exercice 3 :

Salaires (MDA)	n_i	ai	ni_c
[20-30[42	10	21
[30-40[49	10	24,5
[40-45[74	5	74
[45-55[19	10	9,5
[55-70[16	15	5,33
Total	200	-	-

- 1) **La population statistique est :** les 200 employés de l'entreprise,
Son caractère est : le salaire d'un employé.
Sa nature est : quantitatif continu.
- 2) **Le graphique approprié est l'histogramme**



- 3) Le pourcentage des employés ayant au moins un salaire égal à 40000 DA est :

$$p = \frac{(74+19+16)}{200} \times 100 = 53\%.$$

Chapitre 2 : Représentation de données statistiques

Exercice 4 :

Le tableau suivant donne les résultats d'une enquête sur le poids (en Kg) des individus d'une population :

Le poids de l'individu	[45 -50[[50- 55[[55 - 60[[60 - 65[[65 - 70[[70 - 75[[75 - 80[
n_i	14	18	12	19	15	13	9

- 1) Déterminer la population statistique, le caractère et sa nature.
- 2) Calculer les effectifs cumulés ainsi que les fréquences relatives cumulées.
- 3) Tracer les courbes des effectifs cumulés (Croissants et décroissants).
- 4) Déterminer la valeur de la médiane graphiquement, par le calcul et interpréter le résultat.
- 5) Calculer le poids moyen.
- 6) Quelle est le pourcentage d'individus pesant moins de 75Kg ?

Corrigé exercice 4 :

Classes	n_i	$ni \uparrow$	$ni \downarrow$	f_i	$fi \uparrow$	$fi \downarrow$	c_i	$n_i \times c_i$
[45 -50[14	14	100	0,14	0,14	1	47,5	665
[50- 55[18	32	86	0,18	0,32	0,86	52,5	945
[55 - 60[12	44	68	0,12	0,44	0,68	57,5	690
[60 - 65[19	63	56	0,19	0,63	0,56	62,5	1187,5
[65 - 70[15	78	37	0,15	0,78	0,37	67,5	1012,5
[70 - 75[13	91	22	0,13	0,91	0,22	72,5	942,5
[75 - 80[9	100	9	0,09	1	0,09	77,5	697,5
Total	100			1				6140

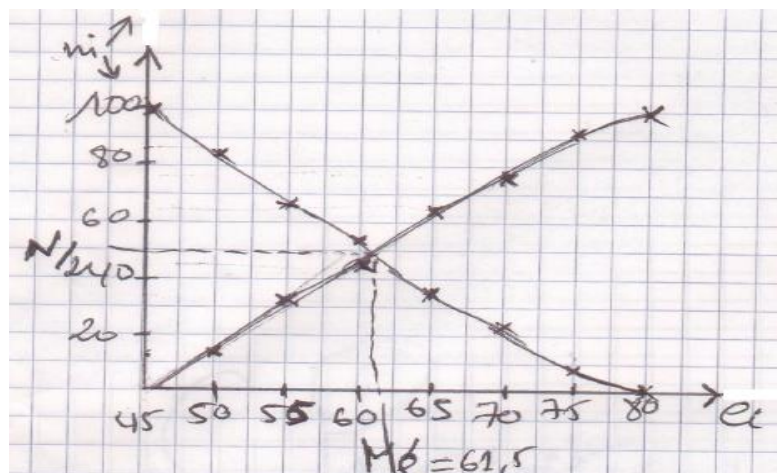
1) **La population statistique** est : les 100 individus,

Son caractère est : le poids d'un individu,

Son type est : quantitatif continu

2) **Calcul des effectifs cumulés et des fréquences cumulées relatives** (voir le tableau ci-dessus).

3) **Les courbes des effectifs cumulés (Croissants et décroissants)**



Chapitre 2 : Représentation de données statistiques

4) La Médiane graphiquement (voir les courbes des effectifs cumulés)

La Médiane par le calcul

La classe médiane est : [60 – 65[

$$M_e = X_{min} + \frac{N/2 - n_{Me-1}^{\uparrow}}{n_{i_{Me}}} \times ai$$

$$M_e = 60 + \frac{50 - 44}{19} \times 5 = 61,57 \text{ kgs}$$

Interprétation :

Il y a 50% des individus qui pèsent moins de 61,57 kgs et 50% des individus qui pèsent plus de 61,57 kgs

5) Calcul du poids moyen

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

$$\bar{X} = \frac{6140}{100} = 61,40 \text{ kgs}$$

5) Proportion d'individus pesant moins de 75 Kgs est 91 %, il suffit de lire dans le tableau la fréquence cumulée croissante associée à la classe [70 - 75[.

Chapitre 3 : Les paramètres de tendance centrale

Introduction

La première étape en statistique consiste à synthétiser l'information d'un tableau à l'aide d'un graphique. Ensuite, on cherche à simplifier davantage cette information en la résumant par une seule valeur numérique. Les mesures de tendance centrale visent à identifier la valeur la plus représentative d'un ensemble de données numériques.

Les caractéristiques de valeur centrale doivent remplir certaines conditions. En 1945, le statisticien britannique « Yule » considère qu'une bonne caractéristique devrait réunir les six qualités suivantes :

- 1- Être objective (2 personnes différentes doivent trouver le même résultat)
- 2- Tenir compte de toutes les observations
- 3- Avoir une signification concrète (interprétation simple et immédiate)
- 4- Être simple à calculer
- 5- Être peu sensible aux fluctuations de l'échantillonnage par rapport à la population mère
- 6- Se prêter facilement au calcul algébrique.

En pratique, aucune caractéristique ne satisfait complètement aux six conditions énoncées ci-dessus. La plus usitée est la moyenne qui se prête le mieux au calcul algébrique.

Exemple :

Examinons la distribution d'une variable caractérisant les éléments des ensembles A et B. Supposons que cette variable soit l'âge et que nous souhaitions comparer les valeurs de cette variable dans chacun des deux groupes. Bien que cette comparaison puisse être réalisée à l'aide de deux polygones sur un même graphique, l'interprétation reste complexe. Pour simplifier cette analyse, on peut utiliser une mesure centrale synthétique, telle que l'âge moyen.

1. La Médiane :

1.1. Définition :

La médiane que l'on note **Me** correspond à la valeur de la variable statistique qui partage la population en deux parties égales, Au préalable, les valeurs de la série statistique sont classées par ordre croissant.

Exemple : Dans une classe, la note médiane à un examen de statistique est 11, dans ce cas-là on peut dire que :

- Au moins la moitié (50%) des élèves a une note inférieure ou égale à 11.
- Au moins la moitié (50%) des élèves a une note supérieure ou égale à 11.

1.2. Déterminations pratique

1.2.1. Cas d'une variable discontinue :

Nous commençons par ordonner la série. Deux cas se présentent :

a. Si le nombre d'observations est impair et aucune valeur n'est répétée

Dans ce cas, la médiane correspond à la $(\frac{n+1}{2})$ ième observation.

Exemple 1 : sept étudiants ont obtenu les notes suivantes : 14, 7, 10, 5, 11, 8, 9

Ordonnons la série par ordre croissant :

$$\begin{array}{ccccccc} \underline{5} & \underline{7} & \underline{8} & & 9 & & \underline{10} & \underline{11} & \underline{14} \\ \mathbf{3 \text{ termes}} & & & & \mathbf{Me} & & \mathbf{3 \text{ termes}} & & \end{array}$$

Le rang de la médiane est donc $\frac{n+1}{2} = \frac{7+1}{2} = 4$ D'où **Me=9 points.**

b. Si le nombre d'observations est pair (2k) et aucune valeur n'est répétée

Dans ce cas, il y a un intervalle appelé « **intervalle médian** », qui laisse autant de termes à gauche qu'à droite : cet intervalle a comme extrémités les valeurs de la **k** ième et de la **(k+1)** ième observations. On prend généralement la moyenne arithmétique de ces 2 bornes.

Chapitre 3 : Les paramètres de tendance centrale

Exemple 2 : soit la série de résultats :

5 8 9 11 14 16

Intervalle médian

La médiane est la moyenne des données occupant les rangs : $\frac{n}{2}$ et $\frac{n}{2} + 1$

C'est-à-dire : $\frac{6}{2} = 3$ et $\frac{6}{2} + 1 = 4$ (donc la médiane occupe donc les rangs 3 et 4).

D'où : $Mé = \frac{9+11}{2} = 10$ points.

Cette méthode de détermination n'est applicable que lorsque le nombre d'observations est faible et que chaque modalité n'est prise qu'une seule fois.

Dans la pratique, la médiane est déterminée soit à partir du graphique cumulé, soit à partir du tableau des effectifs cumulés croissants, en utilisant l'une des deux formules suivantes :

➤ Si le nombre d'observation (N) est impair : $M_e = x_{\frac{n+1}{2}}$

➤ Si le nombre d'observation (N) est pair : $M_e = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$

Exemple 3 :

Les résultats d'une enquête donnent la répartition de 500 familles selon le nombre d'enfants :

T.11 : Distribution des 500 familles selon le nombre d'enfants

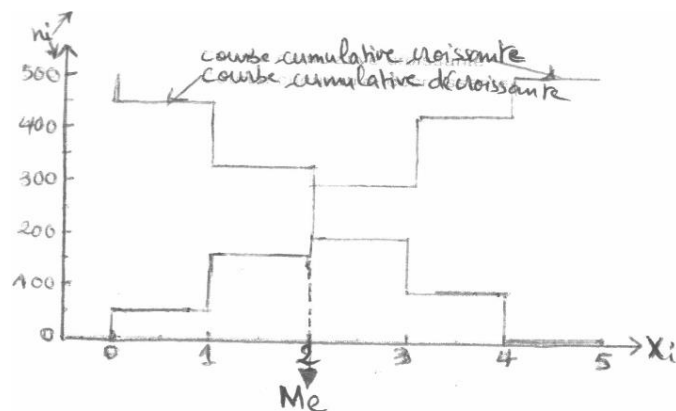
X_i	n_i	$n_i \uparrow$	$n_i \downarrow$
0	50	50	500
1	120	170	450
2	130	300	330
3	110	410	200
4	90	500	90
Total	500	-	

Détermination de la médiane par le calcul :

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

$$M_e = \frac{x_{250} + x_{251}}{2} = \frac{2 + 2}{2} = 2$$

Détermination de la médiane graphiquement :



Chapitre 3 : Les paramètres de tendance centrale

1.2.2. Cas d'une variable continue :

Dans ce cas, les valeurs de la variable statistique étant regroupées en classes, on ne peut, dans un premier temps, que situer la médiane à l'intérieur d'une classe appelée classe médiane. Cette classe médiane se définit à partir des effectifs cumulés.

Pour déterminer la médiane, on regarde quelle est la variable statistique (portée en abscisse) qui correspond à un effectif de $N/2$ (porté en ordonnée).

Exemple : Le tableau suivant retrace les salaires de 50 ouvriers en milliers de dinars (MDA) dans une entreprise.

T.12 : Répartition des 50 ouvriers selon leur salaire en MDA

Classes	n_i	$n_i \uparrow$	f_i	$f_i \uparrow$
[10 - 50 [15	15	0,3	0,3
→ [50 - 100 [20	35	0,4	0,7
[100-200[10	45	0,2	0,9
[200-300[5	50	0,1	1
Total	50	-	1	-

Nous savons que la médiane partage la population (l'effectif total) en deux parties égales, d'où la médiane correspond à la 25^{ème} observation.

Regardons le tableau des effectifs cumulés croissants, le 25^{ème} ouvrier a un salaire compris entre 50 et 100 MDA. Donc la classe médiane appartient l'intervalle [50-100[.

Pour trouver le salaire du 25^{ème} ouvrier, nous procédons de la manière suivante :

$$Me = X_{min} + \frac{\frac{N}{2} - n_{i_{Me-1}}}{n_{i_{Me}}} \times a_i$$

$$Me = 50 + \frac{25 - 15}{20} \times 50 = 75 \text{ MDA}$$

Dans le cas où nous utilisons les fréquences relatives cumulées croissantes, la détermination de la médiane se fait, d'après les formules suivantes :

$$Me = X_{min} + \frac{0,5 - f_{i_{Me-1}}}{f_{i_{Me}}} \times a_i$$

$$Me = 50 + \frac{0,5 - 0,3}{0,4} \times 50 = 75 \text{ MDA.}$$

2. Le Mode (Mo) :

2.1. Définition :

Le Mode noté **Mo** d'une distribution est la valeur de la variable statistique dont l'effectif est plus grand (ou la plus grande fréquence).

Donc Le **mode** d'une série est la valeur la plus fréquente de cette série

2.2. Détermination pratique :

2.2.1. Cas d'une variable discrète : Lorsque la variable statistique est discrète, le mode est défini avec précision

Exemple : Répartition des familles selon le nombre de voitures

Détermination par le calcul		Détermination graphique	
X_i : nombre de voiture	n_i		<p>Le Mode : Mo=1 voiture</p>
1	20		
2	8		
3	2		
Total	30		

Chapitre 3 : Les paramètres de tendance centrale

Remarque :

- Une série peut posséder plusieurs modes s'il y a plusieurs valeurs qui se répètent un même nombre de fois.
- Il peut se trouver également qu'une distribution ne possède aucun mode.

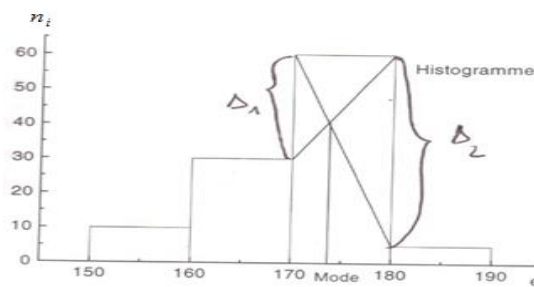
2.2.2. Cas d'une variable continue

Lorsque la variable est continue, la détermination du mode est beaucoup moins précise. En effet, les valeurs prises par la variable étant regroupées en classe, on parle tout simplement de classe modale désignant par-là la classe qui a le plus grand effectif.

Remarque : si les classes sont inégales, la classe modale est définie à partir de l'effectif rectifié.

Exemple1 : La série suivante représente la taille de 105 étudiants (unité de mesure Cm)

e_i	n_i
[150-160[10
[160-170[30
[170-180[60
[180-190[5
Total	105



Il s'agit d'une approximation graphique du mode par la méthode des diagonales. La valeur exacte est donnée par la formule suivante :

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

Avec Δ_1 et Δ_2 représentant la différence entre la fréquence correspondant à la classe modale ([170-180]) et les effectifs des classes voisines.

$$Mo = 170 + \frac{60-30}{(60-30)+(60-5)} \times 10 = 170 + 3,53 \quad \text{D'où : } \mathbf{Mo=173,5 \text{ cm.}}$$

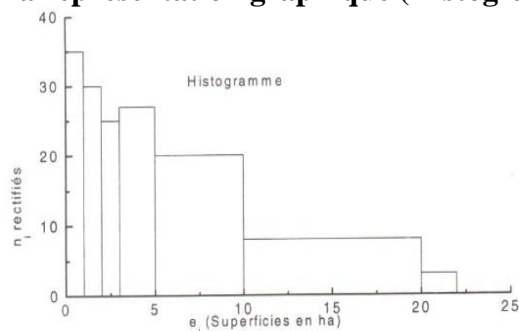
Exemple2 :

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

T.13: la répartition des 330 exploitations selon leur superficie

Classes	n_i	a_i	n_i Corrigés
[0 - 1[35	1	35
[1 - 2[30	1	30
[2 - 3[25	1	25
[3 - 5[54	2	27
[5 - 10[100	5	20
[10-20[80	10	8
[20-22[6	2	3
Total	330	-	-

La représentation graphique (Histogramme)



Chapitre 3 : Les paramètres de tendance centrale

Comme les amplitudes (a_i) sont inégales, donc la classe modale correspond à l'effectif le plus grand des n_i corrigés qui est égal à 35.

D'où la classe modale est : [0 - 1[

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

Remarque : Lorsque les classes sont inégales, a_i correspond à l'amplitude de la classe modale en question.

3. La Moyenne :

3.1. Définition :

La moyenne arithmétique, que le note \bar{X} est la somme des valeurs prises par la variable statistique, divisée par le nombre d'observations.

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

3.2. Détermination pratique

3.2.1. Cas d'une variable discontinue

Dans le cas d'une variable discrète où, à chaque valeur prise par la variable statistique ne correspond qu'un seul individu, il est bien évident qu'il n'y a pas de pondération à faire.

Exemple 1 : 7 ouvriers d'une usine ont reçu en janvier 2004, les primes suivantes : 7500DA, 8300 DA, 9100 DA, 9600 DA, 10700 DA, 10800 DA, 11300 DA

La prime moyenne est :

$$\bar{X} = \frac{7500+8300+9100+9600+10700+10800+11300}{7} = \frac{67300}{7} \quad \text{D'où : } \bar{X}=9614,2\text{DA}$$

La moyenne est donc calculée par cette formule : $\bar{X} = \frac{\sum_{i=1}^n x_i}{N}$

Dans le cas d'une variable discrète où, à chaque valeur de la variable statistique correspond un effectif, on utilise la formule de la moyenne pondérée.

$$\bar{X} = \frac{\sum n_i \times x_i}{N}$$

Exemple 2 : Supposons que les 7 ouvriers aient reçu les primes suivantes : 10700 8300 8300 9600 9600 9600 10700

T.14 : Répartition des ouvriers selon leur prime

X_i : Prime	n_i	$n_i \times x_i$
8300	2	16600
9600	3	28800
10700	2	21400
Total	7	66800

$$\bar{X} = \frac{\sum n_i \times x_i}{N} = \frac{66800}{7} = 9542,8 \text{ DA}$$

3.2.2. Cas d'une variable continue :

Dans ce cas où, à chaque classe de la variable statistique correspond un effectif, nous utilisons la formule de la moyenne pondérée. Seulement, à chaque classe, on associe, pour les calculs, le centre de classe c_i

Exemple 1 :

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

Chapitre 3 : Les paramètres de tendance centrale

T.15 : Répartition des ouvriers selon leur salaire brut mensuel

Salaire (10 ⁴ DA)	n_i	c_i (Centre de classe)	$n_i \times c_i$
[4-5 [27	4,5	121,5
[5-6 [23	5,5	156,5
[6-7 [50	6,5	325
[7-9 [56	8	448
[9-11[37	10	370
Total	193	-	1391

$$D'o\grave{u} : \bar{X} = \frac{\sum n_i \times c_i}{N} = \frac{1391}{193} = 7,20$$

Le Salaire brut hebdomadaire moyen est donc de 72000DA.

3.3. La moyenne géométrique (G) :

Beaucoup moins fréquente que \bar{X} , son utilisation est cependant recommandée lorsque les données sont présentées en taux de croissance, chiffres relatifs, pourcentage.

3.3.1. Moyenne géométrique simple

Soit une population composée de N individus dont les valeurs sont x_1, x_2, \dots, x_n . La moyenne géométrique est dite simple lorsque dans le calcul de la moyenne chaque valeur x_i n'intervient qu'une seule fois. Son calcul est donné par la formule suivante :

$$G = \sqrt[n]{x_1 \times x_2 \times x_3 \times \dots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i} = (\prod_{i=1}^n x_i)^{\frac{1}{n}},$$

Remarque :

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

$$\log G = \log \left(\prod_{i=1}^n x_i \right)^{\frac{1}{n}} = \frac{1}{n} \log \left(\prod_{i=1}^n x_i \right) = \frac{1}{n} \sum_i (\log x_i) = \frac{1}{n} (\log x_1 + \log x_2 + \dots + \log x_n)$$

Exemple :

Supposons que l'on étudie la croissance d'une entreprise qui a eu des augmentations de chiffre d'affaires sur trois ans avec les taux de 10%, 20% et 15%. Nous voulons connaître le taux de croissance annuel moyen.

Calcul des facteurs multiplicatifs

Taux de croissance	$r_1 = +10\%$ (0,1)	$r_2 = +20\%$ (0,2)	$r_3 = +15\%$ (0,15)
Facteur multiplicatif	1,10	1,20	1,15

Lorsque nous ajoutons 1 à chaque taux, nous créons ce qu'on appelle des **facteurs multiplicatifs**. Voici pourquoi et comment cela fonctionne :

- Un taux de croissance de 10% signifie que la valeur a augmenté de 10%.
- En d'autres termes, pour chaque 1 unité (par exemple, 1 dinar ou 1 mètre), on ajoute 0,10 unité. Cela donne un total de $1+0,10=1,10$.
- Le chiffre 1 représente la valeur initiale, et 0,10 représente la croissance.
- Par conséquent, **1,10** est le facteur multiplicatif qui, appliqué à la valeur initiale, donnera la nouvelle valeur après une croissance de 10%.

$$G = \sqrt[3]{1,10 \times 1,20 \times 1,15} = \sqrt[3]{1,518} = (1,518)^{\frac{1}{3}} = 1,147$$

Le taux de croissance moyen est : $Tx = (G - 1) \times 100 = (1,147 - 1) \times 100 = +14,7\%$

Chapitre 3 : Les paramètres de tendance centrale

3.3.2. Moyenne géométrique pondérée

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

$$G = \sqrt[\sum n_i]{x_1^{n_1} \times x_2^{n_2} \times \dots \times x_k^{n_k}}$$

Exemple :

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

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

Réponse : ici, on utilise la formule de la moyenne géométrique pondérée

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

$$\log G = \frac{1}{9}(4 \log 1,50 + 3 \log 1,17 + 2 \log 0,70) = 0,059$$

D'où : $G = 10^{0,059} = 1,145\%$

Donc le taux de croissance annuel moyen est : $Tx = (1,145 - 1) \times 100 = +14,5\%$

3.4. La moyenne harmonique H

La moyenne harmonique est utilisée quand les valeurs statistiques sont données à la forme inverse, c'est à dire les valeurs d'une variable sont données par le calcul d'une unité constante sur une autre variable (km/h, DA/minute, quintaux /hectare, etc.). Elle est particulièrement utile quand nous voulons éviter qu'une valeur extrême fausse le résultat global, comme cela pourrait être le cas avec une moyenne arithmétique.

3.4.1. Moyenne harmonique simple

En considérant N observations x_1, x_2, \dots, x_n , les valeurs de la variable, la moyenne harmonique notée **H** est calculée à partir de la formule suivante :

$$H = \frac{N}{\sum_i \frac{1}{x_i}} = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

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

Quelle est la vitesse moyenne ?

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

3.4.2. Moyenne harmonique pondérée

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

$$H = \frac{\sum n_i}{\sum_i \frac{n_i}{x_i}} = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_k}{x_k}}$$

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

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

Quelle est la vitesse moyenne globale ?

Chapitre 3 : Les paramètres de tendance centrale

Réponse : ici, on utilise la formule de la moyenne harmonique que pondérée

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

3.5. La moyenne quadratique

La moyenne quadratique est utilisée lorsque l'on souhaite prendre en compte les valeurs positives et négatives d'un ensemble de données, mais qu'on veut mesurer leur importance absolue plutôt que simplement les additionner ou soustraire.

3.5.1. La moyenne quadratique simple

En considérant N observations x_1, x_2, \dots, x_n , les valeurs de la variable, la moyenne quadratique notée Q est calculée à partir de la formule suivante :

$$Q = \sqrt{\frac{1}{N} \sum x_i^2}$$

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

Parfois, on cherche à obtenir une mesure de tendance centrale qui reste positive, même lorsque la moyenne arithmétique donnerait un résultat nul. Dans ce cas, on utilise **la moyenne quadratique**, qui consiste à additionner les carrés de toutes les valeurs de la série, puis à prendre la racine carrée de cette somme. Par exemple, dans notre cas :

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

3.5.2. La moyenne quadratique pondérée

Soient $\{x_1, x_2, \dots, x_k\}$ une série de chiffres et $\{n_1, n_2, \dots, n_k\}$ les effectifs correspondants.

La formule de la moyenne quadratique pondérée de cette série est donnée par :

$$Q = \sqrt{\frac{1}{N} \sum n_i \times x_i^2}$$

Exemple : soit le tableau suivant :

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

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

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

4. Les caractéristiques de position

4.1. Définition

Ces caractéristiques correspondent à des nombres occupant une certaine position dans la distribution des valeurs d'une série, celles-ci étant classées dans l'ordre croissant ou décroissant.

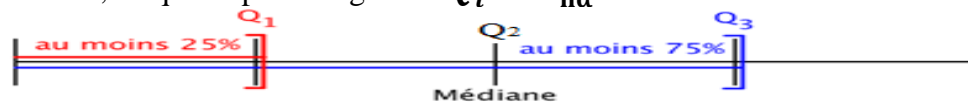
La définition est analogue dans son principe à celle de la médiane. Les paramètres de position séparent les observations ordonnées en deux séries, dans une certaine proportion : ils sont appelés aussi « **quantiles** ».

Chapitre 3 : Les paramètres de tendance centrale

4.2. Principales caractéristiques de position

4.2.1. Les Quartile

Les quartiles divisent l'effectif de la série, préalablement ordonnée par ordre croissant, en quatre parties égales. $Q_i = X_{n\alpha}$



- Le 1^{er} quartile $Q_1 = C_{25}$ est tel que 25% des observations lui sont inférieurs et 75% lui sont supérieures. $Q_1 \left(\alpha = \frac{1}{4} \right)$
- Le 2^e quartile $Q_2 = Me = D_5 = C_{50}$. La médiane apparaît donc comme une caractéristique de position. $Q_2 \left(\alpha = \frac{2}{4} \right)$
- Le 3^e quartile $Q_3 = C_{75}$ est tel que 75% des observations lui sont inférieurs et 25% lui sont supérieures. $Q_3 \left(\alpha = \frac{3}{4} \right)$

4.2.2. Les Déciles

Les déciles, au nombre de neuf, divisent la population en 10 parties égales. Le 1^{er} décile $D_1 = C_{10}$ est la valeur de la variable telle que 10% des observations lui sont inférieurs et 90% ont dessus de lui. $D_1 \left(\alpha = \frac{1}{10} \right)$

- Le 5^e décile $D_5 = C_{50}$ est la médiane. $D_5 \left(\alpha = \frac{5}{10} \right)$
- Le 9^e décile $D_9 = C_{90}$ a 90% des observations au-dessous de lui et 10% lui sont supérieures. $D_9 \left(\alpha = \frac{9}{10} \right)$

4.2.3. Les centiles

Les centiles divisent l'effectif classé dans un ordre croissant en 100 parties égales. Le 1^{er} centile C_1 à 1% des observations au-dessous de lui et 99% lui sont supérieures. $C_1 \left(\alpha = \frac{1}{100} \right)$

- Le 50^e centile C_{50} est la médiane. $C_{50} \left(\alpha = \frac{50}{100} \right)$
- Le 99^e centile P_{99} est tel que 99% des observations lui sont inférieures et 1% lui sont supérieures. $C_{99} \left(\alpha = \frac{99}{100} \right)$

5. Exercices corrigés

Exercice 1 : Cas d'une variable discrète

Calculer les quartiles et les déciles de la distribution des enfants dans 500 familles.

T.16 : Répartition des 500 familles selon le nombre d'enfants

X_i	n_i	$n_i \uparrow$
0	50	50
1	120	170
2	130	300
3	110	410
4	90	500
Total	500	-

Corrigé exercice 1 :

En se référant à la distribution des effectifs cumulés, il est possible d'identifier directement les valeurs associées à ces paramètres. Dans ce cas :

➤ **Le premier quartile:** $\left(\alpha = \frac{1}{4}, n\alpha = 500 \times \frac{1}{4} = 125 \right)$

La 125^{ème} observation se situe à la valeur 1, qui correspond, dans la distribution des effectifs cumulés, à toutes les observations situées entre 50^{ème} et la 170^{ème}.

Chapitre 3 : Les paramètres de tendance centrale

D'où : $Q_1 = X_{n\alpha} = X_{500(\frac{1}{4})} = X_{125} = 1 \text{ enfant}$

➤ **Le deuxième quartile ($\alpha = \frac{2}{4}$, $n\alpha = 500 \times \frac{2}{4} = 250$):**

Pour Q_2 qui est égale à la médiane, la 250^{ème} observation se situe à la valeur 2, d'où :

$$Q_2 = X_{n\alpha} = X_{500(\frac{2}{4})} = X_{250} = 2 \text{ enfants}$$

➤ **Le troisième quartile ($\alpha = \frac{3}{4}$, $n\alpha = 500 \times \frac{3}{4} = 375$):**

Pour Q_3 , la 375^{ème} observation se situe à la valeur 3 puisqu'elle correspond, dans la distribution des effectifs cumulés, à toutes les observations situées entre la 300^{ème} et 410^{ème} observation, d'où :

- $Q_3 = X_{n\alpha} = X_{500(\frac{3}{4})} = X_{375} = 3 \text{ enfants}$

La détermination des déciles et des percentiles s'effectue comme celle des quartiles :

➤ **Le premier décile ($\alpha = \frac{1}{10}$, $n\alpha = 500 \times \frac{1}{10} = 50$):**

- $D_1 = X_{n\alpha} = X_{500(\frac{1}{10})} = X_{50} = 0 \text{ enfant}$

- **Le huitième décile ($\alpha = \frac{8}{10}$):**

- $D_8=3$, puisque la 400^{ème} observation correspondant au 80% des observations qui lui sont inférieures est située entre la 300^{ème} et 410^{ème}.

$$D_8 = X_{n\alpha} = X_{500(\frac{8}{10})} = X_{400} = 3 \text{ enfants}$$

- **Le vingtième centile ($\alpha = \frac{20}{100}$):**

Quant aux centiles, le C20 par exemple est la valeur de la variable telle que 20% des observations lui sont inférieures. Dans ce cas, la 100^{ème} observation se situe à la valeur « 1 » qui correspond dans la distribution des effectifs cumulés, à toutes les observations situées entre la 50^{ème} et la 170^{ème}.

$$C_{20} = X_{n\alpha} = X_{500(\frac{20}{100})} = X_{100} = 1 \text{ enfant}$$

Exercice 2 : Cas d'une variable continue.

Calculer les principaux paramètres de position de la distribution ci-dessous

T.17 : Distribution de 100 ouvriers selon le salaire horaire en DA

Xi : salaire horaire	n_i	$n_i \uparrow$
[20-25[9	9
[25-30[29	38
[30-35[18	56
[35-40[6	62
[40-45[26	88
[45-50[12	100
Total	100	-

Chapitre 3 : Les paramètres de tendance centrale

Corrigé de l'exercice 2 :

Calcul des principaux paramètres de position

➤ La médiane :

$$M_e(\alpha = \frac{1}{2}, n\alpha = 100 \times \frac{1}{2} = 50) \rightarrow M_e \in [30 - 35[$$

$$M_e = X_{min} + \frac{\frac{N}{2} - n_{i_{M_e-1}}}{ni_{M_e}} \times ai \rightarrow M_e = 30 + \frac{50-38}{18} \times 5 = 33,333 DA$$

➤ Le premier quartile:

$$Q_1(\alpha = \frac{1}{4}, n\alpha = 100 \times \frac{1}{4} = 25) \rightarrow Q_1 \in [25; 30[$$

$$Q_1 = X_{min} + \frac{\frac{N}{4} - n_{i_{Q_1-1}}}{ni_{Q_1}} \times ai \rightarrow Q_1 = 25 + \frac{25-9}{29} \times 5 = 27,758 DA$$

➤ Le troisième quartile

$$Q_3(\alpha = \frac{3}{4}, n\alpha = 100 \times \frac{3}{4} = 75) \rightarrow Q_3 \in [40; 45[$$

$$Q_3 = X_{min} + \frac{N \cdot (\frac{3}{4}) - n_{i_{Q_3-1}}}{ni_{Q_3}} \times ai \rightarrow Q_3 = 40 + \frac{75-62}{26} \times 5 = 42,5 DA$$

$$D_1(\alpha = \frac{1}{10}, n\alpha = 100 \times \frac{1}{10} = 10) \rightarrow D_1 \in [25; 30[$$

$$D_1 = X_{min} + \frac{N \cdot (\frac{1}{10}) - n_{i_{D_1-1}}}{ni_{D_1}} \times ai \rightarrow D_1 = 25 + \frac{10-9}{29} \times 5 = 25,1 DA$$

$$D_9(\alpha = \frac{9}{10}, n\alpha = 100 \times \frac{9}{10} = 90) \rightarrow D_9 \in [45; 50[$$

$$D_9 = X_{min} + \frac{N \cdot (\frac{9}{10}) - n_{i_{D_9-1}}}{ni_{D_9}} \times ai \rightarrow D_9 = 45 + \frac{90-88}{12} \times 5 = 45,83 DA$$

Le deuxième quartile : $Q_2 = M_e = 33,333 DA$

Le cinquantième centile : $C_{50} = M_e = 33,333 DA$

Le vingt-cinquième centile : $C_{25} = Q_1 = 27,758 DA$

Le soixante-quinzième centile : $C_{75} = Q_3 = 42,500 DA$

Exercice 3 :

Une société a connu trois phases distinctes dans l'évolution de ses ventes. Pendant les cinq premières années, ses ventes ont augmenté de 40 % par an. Au cours des trois années suivantes, la croissance annuelle a ralenti à 15 %. Enfin, lors des deux dernières années, les ventes ont diminué de 20 % par an.

- Quel est le taux de croissance annuel moyen des ventes sur cette période de 10 ans ?

Corrigé exercice 3 :

Pour résoudre cet exercice, on utilisera la formule du **taux de croissance annuel moyen géométrique** : $Tx = (G - 1) \times 100$

$$G = \sqrt[10]{(1,40)^5 \times (1,15)^3 \times (0,80)^2}$$

$$\log G = \frac{1}{10} (5 \log 1,40 + 3 \log 1,15 + 2 \log 0,80) = 0,07188$$

$$D'où : G = 10^{0,07188} = 1,181\%$$

$$Tx = (1,181 - 1) \times 100 = +18,1\%$$

Le taux de croissance annuel moyen des ventes sur cette période de 10 ans est de 18,1%

Chapitre 3 : Les paramètres de tendance centrale

Exercice 4 :

Une société effectue des livraisons entre Alger et Oran avec 12 véhicules. Lors d'une journée particulière, les véhicules ont roulé aux vitesses moyennes suivantes :

Vitesse moyenne (km/h)	50	65	80
Nombre de véhicules	5	4	3

Quelle est la vitesse moyenne globale des véhicules sur ce trajet ?

Corrigé de l'exercice 4 :

Ici, on utilise la formule de la moyenne harmonique pondérée

$$H = \frac{12}{\frac{5}{50} + \frac{4}{65} + \frac{3}{80}} = 60,29 \text{ km/h}$$

La vitesse moyenne globale est d'environ **60,29 km/h**.

Chapitre 4 : Les paramètres de dispersion

Introduction :

Deux distributions de fréquence peuvent avoir la même valeur pour les paramètres de tendance centrale, comme la moyenne, la médiane ou le mode, tout en présentant des différences importantes dans leur structure. Ces différences peuvent se manifester dans la manière dont les données sont réparties autour de ces valeurs centrales. C'est ici qu'interviennent les paramètres de dispersion, qui permettent de mesurer et de quantifier la variabilité ou l'étalement des données. Ces paramètres offrent ainsi une vision plus complète des distributions en complétant les informations fournies par les indicateurs de tendance centrale.

Exemple : supposons que 10 élèves aient obtenu les résultats suivants lors d'un examen (données classées selon l'ordre croissant) :

2 4 6 6 6 6 6 6 8 10

Nous vérifions aisément que : $\bar{X} = M_e = M_o = 6$

Dans un deuxième examen, les résultats sont les suivants :

1 3 5 5 6 6 7 8 9 11

Encore ici, nous avons : $\bar{X} = M_e = M_o = 6$

L'examen attentif des deux séries fait ressortir une différence entre les deux distributions. **Comment cette différence peut-elle être traduite numériquement ?**

La connaissance de la valeur centrale d'une série de données ne suffit pas à fournir une information complète sur sa structure. Il est donc essentiel de disposer de mesures qui mettent en évidence un aspect déterminant : la **variabilité** ou la **dispersion** des données. Ces mesures permettent d'effectuer des comparaisons significatives entre différentes séries.

Par exemple, bien que deux distributions puissent avoir les mêmes paramètres de tendance centrale, elles peuvent différer considérablement par leur dispersion. Dans le cas présenté, la première distribution est fortement concentrée autour de la valeur 6, alors que la seconde présente une variabilité plus marquée avec des données réparties sur un intervalle plus large. Cela illustre l'importance des paramètres de dispersion pour comprendre pleinement la nature des distributions et leurs différences.

Définition :

Une caractéristique de dispersion est une mesure numérique qui permet d'évaluer dans quelle mesure les observations d'une série s'écartent les unes des autres ou par rapport à leur valeur centrale. Ces caractéristiques apportent des informations essentielles sur la variabilité ou l'étalement des données. Dans ce chapitre, nous examinerons les différentes caractéristiques de dispersion selon le classement suivant :

- **Les différences entre deux valeurs d'une série ;**
- **Les valeurs centrales calculées à partir de variables centrées.**

1. Différences entre deux valeurs d'une série

Les mesures de dispersion basées sur les différences entre deux valeurs d'une série permettent d'évaluer l'amplitude de la variabilité des données. Parmi ces mesures :

1.1. L'étendue

L'étendue constitue l'indicateur de dispersion le plus simple et le plus intuitif.

1.1.1. Définition :

L'étendue, notée E , correspond à la différence entre la valeur maximale et la valeur minimale d'une variable statistique, après avoir classé les données dans l'ordre croissant ou décroissant.

1.1.2. Détermination pratique

a. Variable discontinue

Dans le cas d'une série discontinue, l'étendue ne pose pas de problèmes de calcul.

Nous avons : $E = X_{max} - X_{min}$

Chapitre 4 : Les paramètres de dispersion

Exemple 1 : Reprenons les 2 séries de notes obtenues par 10 étudiants lors de deux examens (séries proposées précédemment).

2	4	6	6	6	6	6	6	8	10
1	3	5	5	6	6	7	8	9	11

- Dans la première distribution, l'étendue est égale à : $E = (10 - 2) = 8$
- Dans la deuxième distribution, l'étendue est égale à : $E = (11 - 1) = 10$.

Exemple 2 : soit la distribution suivante sur la répartition des ménages selon le nombre de filles

T.19 : Distribution des 500 ménages selon le nombre de filles

X_i	n_i
0	50
1	120
2	130
3	110
4	90
Total	500

Dans ce cas, l'étendue est égale à :

$$E = (4 - 0) = 4 \text{ Enfants par famille}$$

b. Variable continue

Dans le cas d'une distribution continue, deux cas se présentent :

- **1^{er} cas : Les classes extrêmes sont bornées** $E = e_k - e_1$

Exemple : le tableau suivant donne la répartition de 50 ouvriers selon leurs salaires :

T.20 : Répartition des 50 ouvriers selon leur salaire en MDA

Classes	n_i
[10 - 50 [15
[50 - 100 [20
[100-200[10
[200-300[5
Total	50

Dans ce cas, l'étendue est égale à :

$$E = 300 - 10 = 290 \text{ MDA}$$

- **2^{ème} cas : Les classes extrêmes ne sont pas bornées**

La détermination de l'étendue pose problème. En pratique, nous bornons les classes extrêmes selon la règle étudiée au chapitre 2 et nous calculons l'étendue comme ci-dessus.

Exemple2 :

La série suivante dans le tableau ci-après donne les âges des participants à une étude:

Classes d'âges (en années)	-25	[25-30[[30-35[[35-40[[40-45[[45-60[+60
Nombre de participants	12	23	42	56	34	16	4

Bornons le 2 classes en tenant compte des classes adjacentes. La classe « moins de 25 » a donc une amplitude de 5 car la classe adjacente est la classe « 25-30 » d'amplitude égale à « 5 ». De même, la classe « plus de 60 » a une amplitude égale à « 15 » car la classe adjacente est la classe « 45-60 » d'amplitude égale à 15

Dans ce cas, l'étendue est égale : $E = (75 - 20) = 55 \text{ ans}$.

Chapitre 4 : Les paramètres de dispersion

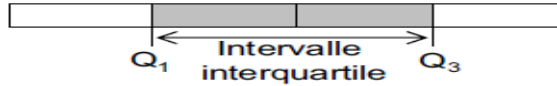
2.2. Écart interquartile Q

L'intervalle interquartile $[Q_1; Q_3]$

L'intervalle interquartile est une mesure de la variation (dispersion) qui n'est pas influencée par les valeurs extrêmes, contrairement à l'étendue.

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

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



2.3 Écart interdécile D

L'intervalle interdécile $[D_1; D_9]$

Avec le même principe que pour l'intervalle interquartile, on peut également s'intéresser à l'intervalle interdécile noté D , dont la formule est la suivante : $D = D_9 - D_1$

Cet intervalle contient 80% des observations en laissant respectivement 10% des observations à droite et à gauche de l'intervalle.

2.4 Écart intercentile C

L'intervalle intercentile $[C_1; C_{99}]$

Avec le même principe que pour l'intervalle interdécile, on peut également s'intéresser à l'intervalle intercentile noté C , dont la formule est la suivante : $C = C_{99} - C_1$

Cet intervalle contient 98% des observations en laissant respectivement 1% des observations à droite et à gauche de l'intervalle.

3. Valeurs centrales de variables centrées

L'opération consiste, ici, à s'appuyer sur des variables centrées (différences des valeurs à leur moyenne ou, très rarement, à leur médiane). Nous retenons ici :

- L'écart absolu moyen
- La variance et l'écart-type (ou standard-déviation)
- Le coefficient de dispersion relative (coefficient de variation).

3.1. Ecart absolu moyen

3.1.1. Définition

L'écart moyen absolu que l'on note EM est la moyenne arithmétique des valeurs absolues des écarts entre la valeur prise par la variable statistique X et une valeur type choisie qui est généralement la moyenne \bar{X} .

3.1.2. Détermination pratique

a. Variable discontinue

- Dans le cas d'une distribution discontinue de N observation, l'écart moyen des

$$\text{valeurs à la moyenne est : } EM = \frac{\sum |x_i - \bar{X}|}{N}$$

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

Exemple1 : La série suivante représente le nombre d'absences d'un échantillon de 6 étudiants, à un cours de mathématique.

4 5 6 2 1 0

Calculons d'abord la moyenne : $\bar{X} = \frac{\sum x_i}{N} = \frac{18}{6} = 3$ absences

D'où :

$$EM = \frac{\sum |x_i - \bar{X}|}{N} = \frac{|4 - 3| + |5 - 3| + |6 - 3| + |2 - 3| + |1 - 3| + |0 - 3|}{6} = \frac{12}{6}$$

$EM = 2$ absences

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

$$EM = \frac{\sum_{i=1}^n n_i |x_i - \bar{X}|}{N}$$

Chapitre 4 : Les paramètres de dispersion

Exemple2 :

La série suivante représente le nombre d'absences, d'un échantillon de 10 étudiants, à un cours de statistique.

9 2 1 1 7 7 7 4 6 7

T.21 : Répartition des 10 étudiants selon le nombre d'absences

X_i	n_i	$n_i \times x_i$	$n_i \times x_i - \bar{X} $
1	2	2	8
2	1	2	3
4	1	4	1
6	1	6	1
7	4	28	8
9	1	9	4
Total	10	51	25

$$\bar{X} = \frac{\sum_{i=1}^n n_i \times x_i}{N} = \frac{51}{10} = 5,1 = 5$$

$$EM = \frac{\sum_{i=1}^n n_i |x_i - \bar{X}|}{N} = \frac{25}{10} = 2,5$$

$EM \approx 3$ absences

b. Variable continue

- Dans le cas d'observation groupées par classe, l'écart absolu moyen s'exprime par :

$$EM = \frac{\sum_{i=1}^n n_i |c_i - \bar{X}|}{N}$$

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

Et n_i : effectif de la classe i

Exemple :

Soit le tableau suivant sur la répartition des ouvriers selon leur salaire brut mensuel.

Salaire (10^4 DA)	n_i	c_i	$n_i \times c_i$	$ c_i - \bar{X} $	$n_i \times c_i - \bar{X} $
[4-5 [27	4,5	121,5	2,7	72,9
[5-6 [23	5,5	126,5	1,7	39,1
[6-7 [50	6,5	325	0,7	35
[7-9 [56	8	448	0,8	44,8
[9-11[37	10	370	2,8	103,6
Total	193	-	1391	-	295,4

Rappelons que la moyenne arithmétique est égale à $\bar{X} = \frac{1391}{193} = 7,2 \times 10^4$ DA.

$$D'où : EM = \frac{\sum_{i=1}^n n_i |c_i - \bar{X}|}{N} = \frac{295,4}{193} = 1,53 \times 10^4 \text{ DA}$$

L'écart absolu moyen et donc de 1530 DA.

3.2. La variance et l'écart-type (ou standard-déviation)

3.2.1. La variance

On appelle la variance de la variable X , la moyenne des carrés des écarts des valeurs observées par rapport à leur moyenne arithmétique \bar{X} .

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

La variance indique dans quelle mesure les données sont étalées ou, au contraire, concentrées autour de la moyenne.

Variance faible : Les valeurs sont proches de la moyenne ; donc, les données sont peu dispersées.

Chapitre 4 : Les paramètres de dispersion

Variance élevée : Les valeurs sont éloignées de la moyenne ; il y a une forte dispersion dans les données.

La variance est toujours positive, et elle peut aussi se calculer à partir d'une formule développée déduite de la première formule :

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

3.2.2. L'écart type

L'écart type noté $\sigma(X)$ est la racine au carrée de la variance. On note : $\sigma(X)$

a. Cas d'une variable discontinue

Exemple :

Si nous reprenons l'exemple des données représentant le nombre d'absences d'un échantillon de 10 étudiants.

T.22: Répartition des 10 étudiants selon le nombre d'absences

X_i	n_i	$n_i \times x_i$	$n_i(x_i - \bar{X})^2$	$n_i \times x_i^2$
1	2	2	33,62	2
2	1	2	9,61	4
4	1	4	1,21	16
6	1	6	0,81	36
7	4	28	14,44	196
9	1	9	15,21	81
Total	10	51	74,9	335

$V(X) = \frac{\sum_{i=1}^n n_i(x_i - \bar{X})^2}{N} = \frac{74,9}{10} = 7,49$. La formule développée $V(X)$ est plus facile.

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

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

$$V(X) = \frac{335}{10} - (5,1)^2 = 7,49$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{7,49} = 2,73$$

$$\sigma(X) \approx 3 \text{ absences}$$

b. Cas d'une variable continue

Dans ce cas, La variance est donnée par la formule suivante :

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

Avec : c_i = centre de la classe et n_i = effectif de la classe i.

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

T.23 : Répartition des salariés selon leur salaire mensuel brut.

Salaire (10^4 DA)	n_i	c_i	$n_i \times c_i$	$n_i \times c_i^2$
[4-5 [27	4,5	121,5	546,75
[5-6 [23	5,5	126,5	695,75
[6-7 [50	6,5	325	2112,5
[7-9 [56	8	448	3584
[9-11[37	10	370	3700
Total	193	-	1391	10639

D'où : $V(X) = \frac{\sum_{i=1}^n n_i \times c_i^2}{N} - \bar{X}^2$

$$V(X) = \frac{10639}{193} - (7,2)^2$$

Chapitre 4 : Les paramètres de dispersion

$$V(X) = 3,28$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{3,28}$$

$$\sigma(X) = 1,81 \times 10^4$$

4. Coefficient de dispersion relative : coefficient de variation

L'écart-type s'exprimant dans la même unité que les modalités du caractère, peut rendre difficiles les comparaisons de dispersion si les variables sont exprimées en unités différentes. Le coefficient de variation (CV) est un nombre sans unité de mesure, c'est le rapport de l'écart type $\sigma(X)$ sur la moyenne \bar{X} de l'échantillon étudié.

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

Le CV est également indépendant des unités choisies, il permet de comparer les séries statistiques exprimées en unités de mesure différentes.

L'usage du CV répond donc à la faiblesse de l'écart type qui est un paramètre simple de dispersion.

Exemple : Supposons que nous comparons deux séries de données :

Série A : Les revenus mensuels (en dinars algériens) de 10 employés dans une entreprise.

- Moyenne : $\bar{X}_A = 50000DA$
- Écart-type : $\sigma_A = 10000DA$

Série B : Les tailles (en centimètres) de 10 personnes dans un groupe.

- Moyenne : $\bar{X}_B = 170cm$
- Écart-type : $\sigma_B = 15cm$

Si nous voulons comparer la dispersion des deux séries, l'écart-type seul est difficile à interpréter, car il dépend des unités de mesure (DA pour les revenus, cm pour les tailles).

En calculant le **coefficient de variation (CV)** :

- Pour la **Série A** :

$$CV_A = \frac{\sigma_A(X)}{\bar{X}_A} \times 100 = \frac{10000}{50000} \times 100 = 20\%$$

- Pour la **Série B** :

$$CV_B = \frac{\sigma_B(X)}{\bar{X}_B} \times 100 = \frac{15}{170} \times 100 = 8,82\%$$

Interprétation :

- La dispersion relative des revenus (Série A) est plus importante (20 %) que celle des tailles (8.82 %).
- Grâce au CV , nous avons pu comparer la variabilité des deux séries même si elles utilisent des unités différentes (DA et cm).

5. Exercices corrigés

Exercice 1 :

Dans une bibliothèque, l'ensemble des abonnés a été réparti suivant le nombre d'ouvrages empruntés durant un mois :

X_i : le nombre d'ouvrages empruntés	0	1	2	3	4	5
Nombre d'abonnés (n_i)	40	53	57	64	42	33

- Déterminer l'étendue
- Calculer le nombre moyen d'ouvrage empruntés et donner son interprétation.
- Calculer l'écart absolu moyen et donner son interprétation
- Calculer la variance, l'écart-type et interpréter les résultats.
- Calculer le coefficient de variation et interpréter les résultats.

Corrigé exercice 1 :

Nombre d'ouvrages	n_i	$n_i \times x_i$	$ x_i - \bar{X} $	$n_i \times x_i - \bar{X} $	$n_i \times (x_i - \bar{X})^2$
0	40	0	2,4	96	230,4
1	53	53	1,4	74,2	103,88
2	57	114	0,4	22,8	9,12
3	64	192	0,6	38,4	23,04
4	42	168	1,6	67,2	107,52
5	33	165	2,6	85,8	223,08
Total	289	692	-	384,4	696,04

1. Détermination de l'étendue :

$$E = X_{max} - X_{min} = 5 - 0 = 5 \text{ ouvrages}$$

2. Calcul du nombre moyen d'ouvrage empruntés

$$\bar{X} = \frac{\sum_{i=1}^n n_i x_i}{N} = \frac{692}{289} = 2,4 \approx 2 \text{ ouvrages}$$

Interprétation : En moyenne, chaque abonné emprunte 2,4 ouvrages durant un mois.

3. Calcul de l'écart absolu moyen :

$$EM = \frac{\sum_{i=1}^n n_i |x_i - \bar{X}|}{N} = \frac{384,4}{289} = 1,33 \approx 1 \text{ ouvrage}$$

Interprétation : En moyenne, chaque abonné s'écarte de 1,33 ouvrage de la moyenne de 2,4.

4. Calcul de la variance et de l'écart-type

➤ Calcul de la variance

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

$$V(X) = \frac{696,04}{289} = 2,41 \text{ (ouvrages)}^2$$

➤ Calcul de l'écart-type

$$\sigma(X) = \sqrt{V(X)} = \sqrt{2,41} = 1,55 \approx 2 \text{ ouvrages}$$

Interprétation :

Un écart-type de 1,55 signifie que le nombre d'ouvrages empruntés varie en moyenne de 1,55 ouvrage autour du nombre moyen d'ouvrages empruntés, soit environ 2,4.

Chapitre 4 : Les paramètres de dispersion

5. Calculer du coefficient de variation CV

$$CV = \frac{\sigma(X)}{\bar{X}} \times 100 = \frac{1,55}{2,4} \times 100 = 64,5\%$$

Interprétation :

- Un (CV = 64,5%) supérieur à 50 % indique une forte dispersion des données par rapport à la moyenne.
- Les observations varient de manière significative autour de la moyenne de 2,4 ce qui signifie que la répartition du nombre d'ouvrages est relativement hétérogène.

Exercice 2 :

Soit la répartition des salaires mensuels des employés d'une usine « A » :

Salaires (milliers DA)	[50-60[[60-70[[70- 85 [[85-90[[90-100[
Nombre d'employés	12	19	24	10	5

1. Déterminer l'étendue de la distribution.
2. Calculer le salaire moyen des employés.
3. Calculer l'écart interquartile, l'écart interdécile et interpréter les résultats.
4. Calculer la variance, l'écart-type et interpréter les résultats.
5. Calculer le coefficient de variation et interpréter les résultats.

Corrigé de l'exercice 2

Salaires (MDA)	n_i	$n_i \uparrow$	c_i	$n_i \times c_i$	$n_i \times c_i^2$
[50-60 [12	12	55	660	36300
[60-70 [19	31	65	1235	80275
[70-85 [24	55	77,5	1860	144150
[85-90 [10	65	87,5	875	76562,5
[90-100[5	70	95	475	45125
Total	70	-	-	5105	382412,5

1. Détermination de l'étendue :

$$E = X_{max} - X_{min} = 100 - 50 = 50 \text{ MDA}$$

2. Calcul du salaire moyen des employés

$$\bar{X} = \frac{\sum_{i=1}^n n_i c_i}{N} = \frac{5105}{70} = 72,93 \text{ MDA}$$

3. Calcul de l'écart interquartile et de l'écart interdécile

➤ L'écart interquartile : $Q = Q_3 - Q_1$

$$\bullet Q_1 = X_{min} + \frac{\frac{N}{4} - n_{i_{Q_1-1}}}{n_{i_{Q_1}}} \times a_i$$

$$\bullet Q_1 = 60 + \frac{17,5 - 12}{19} \times 10 = 62,89 \text{ MDA}$$

$$\bullet Q_3 = X_{min} + \frac{N \cdot (\frac{3}{4}) - n_{i_{Q_3-1}}}{n_{i_{Q_3}}} \times a_i$$

$$\bullet Q_3 = 70 + \frac{52,5 - 31}{24} \times 15 = 83,44 \text{ MDA}$$

$$\text{Donc : } Q = Q_3 - Q_1 = 83,44 - 62,89 = 20,55 \text{ MDA}$$

Interprétation :

La dispersion des 50 % des salaires centrés autour de la médiane est de 20,55 MDA.

➤ L'écart interdécile : $D = D_9 - D_1$

$$\bullet D_1 = X_{min} + \frac{N \cdot (\frac{1}{10}) - n_{i_{D_1-1}}}{n_{i_{D_1}}} \times a_i$$

$$\bullet D_1 = 50 + \frac{7 - 0}{12} \times 10 = 55,83 \text{ MDA}$$

Chapitre 4 : Les paramètres de dispersion

- $D_9 = X_{min} + \frac{N \cdot \left(\frac{9}{10}\right) - n_{i_{D_9-1}}}{n_{i_{D_9}}} \times a_i$
- $D_9 = 85 + \frac{63-55}{10} \times 5 = 89 \text{ MDA}$

Donc : $D = D_9 - D_1 = 89 - 55,83 = 33,17 \text{ MDA}$

Interprétation :

La dispersion des 80 % des salaires centrés autour de la médiane est de 33,17 MDA.

4. Calcul de la variance et de l'écart-type

➤ Calcul de la variance

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

$$V(X) = \frac{382412,5}{70} - (72,93)^2 = 144,85 \text{ (MDA)}^2$$

Interprétation :

Une variance de $144,85 \text{ (MDA)}^2$ est une indication de la variabilité importante des salaires, mais son interprétation directe est difficile car elle s'exprime en unités au carré.

➤ Calcul de l'écart-type

$$\sigma(X) = \sqrt{V(X)} = \sqrt{144,85} = 12,04 \text{ MDA}$$

Interprétation :

➤ Un écart-type de 12.04 MDA signifie que les salaires journaliers des employés s'écartent en moyenne de 12.04 MDA par rapport au salaire moyen (72.93 MDA).

➤ Cela montre une dispersion modérée des salaires autour de la moyenne.

5. Calculer du coefficient de variation CV

$$CV = \frac{\sigma(X)}{\bar{X}} \times 100 = \frac{12,04}{72,93} \times 100 = 16,5\%$$

Interprétation :

➤ Avec un $CV = 16,5\%$, cela signifie que la dispersion relative des salaires représente 16,5 % de la moyenne.

➤ Ce niveau de dispersion est modéré : généralement, un CV inférieur à 20 % est considéré comme un signe d'homogénéité dans la distribution. Cela indique que les salaires des employés de l'usine sont relativement homogènes, bien qu'une certaine variabilité existe.

Chapitre 5 : Les paramètres de forme

Introduction :

L'analyse statistique permet de caractériser une distribution de données non seulement par des mesures de tendance centrale (comme la moyenne) ou de dispersion (comme l'écart-type), mais également par des paramètres de forme. Ces derniers offrent une description qualitative de l'allure de la courbe des effectifs ou des fréquences, sans nécessiter son tracé graphique.

En comparaison avec une distribution « idéale » appelée courbe normale (ou gaussienne), deux types de caractéristiques permettent de décrire les particularités d'une distribution : Les caractéristiques d'asymétrie, qui indiquent si la distribution est symétrique ou non. Les caractéristiques d'aplatissement, qui mesurent si la courbe est plus ou moins concentrée autour de son centre.

Ces caractéristiques sont définies à l'aide des moments statistiques, que nous allons détailler dans la suite.

1. Les moments simples et les moments centrés

1.1. Moments centrés

Nous parlons de moment centré d'ordre k et nous le notons μ_k (k entier positif quelconque). Son expression devient alors :

$$\mu_k = \frac{\sum_{i=1}^n n_i (x_i - \bar{X})^k}{N}$$

Ainsi peut-on établir :

$$\mu_1(X) = 0 \text{ (Somme des variables centrées nulle)}$$

$$\mu_2(X) = V(X) \text{ (Variance).}$$

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

1.2. Moments simples (Moments par rapport à l'origine)

Le principe de calcul est identique au précédent. Seulement, les variables ne sont pas centrées sur la moyenne, mais sur la valeur 0. Ainsi :

Nous parlons de **moment simple d'ordre k** et nous le notons m_k

Nous reconnaitrons les expressions :

$$m_1 = \frac{\sum_{i=1}^n n_i x_i}{N} = \bar{X} : \text{Moyenne arithmétique, moment du 1}^{\text{er}} \text{ ordre.}$$

$$m_2 = \frac{\sum_{i=1}^n n_i x_i^2}{N} : \text{Carré de la moyenne quadratique, moment du 2}^{\text{ème}} \text{ ordre.}$$

La formule générale des moments par rapport à l'origine est donc : $m_k = \frac{\sum_{i=1}^n n_i x_i^k}{N}$

Exemple : La répartition de 100 familles selon le nombre de garçons

X_i	n_i	$n_i \times x_i$	$n_i \times x_i^2$	$n_i \times x_i^3$	$n_i(x_i - \bar{X})^1$	$n_i(x_i - \bar{X})^2$
1	25	25	25	25	-25	25
2	50	100	200	400	0	0
3	25	75	225	675	25	25
Total	100	200	450	1100	0	50

$$m_1 = \frac{\sum n_i x_i}{N} = \frac{200}{100} = 2$$

$$m_2 = \frac{\sum n_i x_i^2}{N} = \frac{450}{100} = 4,5$$

$$m_3 = \frac{\sum n_i x_i^3}{N} = \frac{1100}{100} = 11$$

$$\mu_1(X) = \frac{\sum n_i (x_i - \bar{X})^1}{N} = \frac{0}{100} = 0$$

$$\mu_2(X) = \frac{\sum n_i (x_i - \bar{X})^2}{N} = \frac{50}{100} = 0,5$$

Chapitre 5 : Les paramètres de forme

Remarque :

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

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

$$\mu_3(X) = m_3 - 3m_1 m_2 + 2m_1^3$$

$$\mu_4(X) = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4$$

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

2. Asymétrie

Nous savons qu'une distribution est **symétrique** si les trois caractéristiques de tendance centrale sont confondues :

- Moyenne = Médiane = Mode.
- Q_1 et Q_3 équidistants de la médiane
- D_1 et D_9 équidistants de la médiane

Dans ce cas, la distribution est représentée par un histogramme en forme de cloche. La position relative des éléments permet de calculer les coefficients d'asymétrie, qui sont égaux à zéro en cas de symétrie, et s'éloignent de cette valeur lorsque la distribution devient plus asymétrique.

2.1. Principaux coefficients d'asymétrie.

Nous en retiendrons trois, du nom de leurs auteurs qui en sont les initiateurs :

Fisher, Yule et Pearson.

2.1.1. Coefficient de Fisher

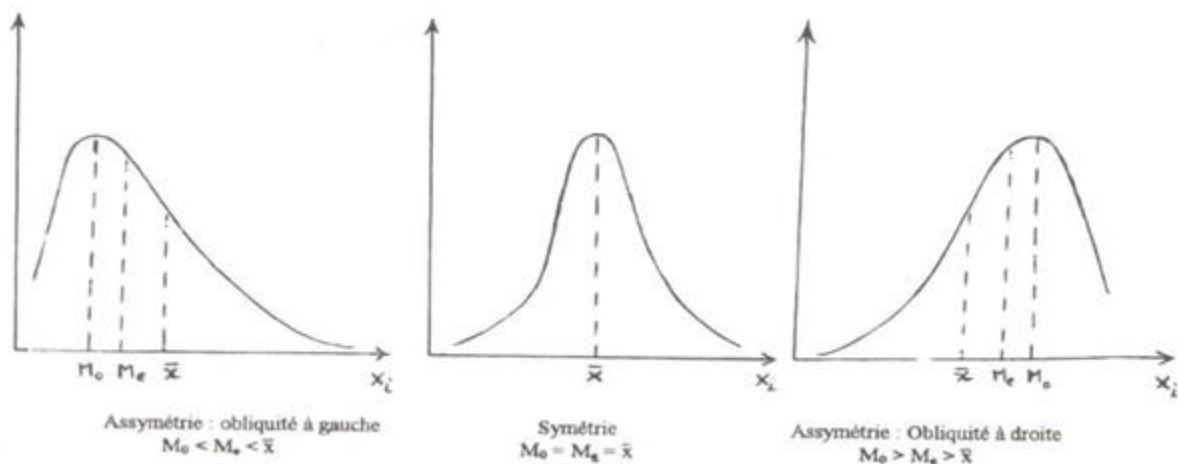
$$F_1 = \frac{\mu_3(X)}{\sigma^3(X)}$$

Nous constatons que le coefficient de Fisher n'est autre que la racine carrée du coefficient β_2 de Pearson (voir ultérieurement). Ce coefficient est sans dimension, invariant par changement d'origine et nul pour les distributions symétriques

* $F_1 = 0$ → Courbe symétrique (les observations sont dispersées de manière égale de part et d'autre d'une valeur centrale)

* $F_1 > 0$ → Étalement à droite (la courbe est oblique à gauche : M_0 est à gauche de la médiane).

* $F_1 < 0$ → Étalement à gauche (la courbe est oblique à droite : M_0 est à droite de la médiane).



2.1.2. Coefficient de Yule :

Ce coefficient est simple se calcule à partir des quartiles.

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

Chapitre 5 : Les paramètres de forme

- $C_y = 0 \rightarrow$ Courbe symétrique
- $C_y > 0 \rightarrow$ La courbe est oblique à gauche (distribution étalée à droite : $M_0 < M_e < \bar{X}$)
- $C_y < 0 \rightarrow$ La courbe est oblique à droite (distribution étalée à gauche : $\bar{X} < M_e < M_0$)

Remarque :

La nullité de ce coefficient ($C_y = 0$) n'est pas un indice de symétrie fiable, si l'asymétrie porte essentiellement sur des « queues » de distribution. Celle-ci est alors révélée par le coefficient de Pearson.

2.1.3. Coefficient de Pearson

Pearson propose deux coefficients pour mesurer l'asymétrie :

a. Un premier assez simple (β_1):

$$\beta_1 = \frac{\bar{X} - M_0}{\sigma(X)}$$

- $\beta_1 = 0 \rightarrow$ Courbe symétrique
- $\beta_1 > 0 \rightarrow$ La courbe est oblique à gauche (distribution étalée à droite : $M_0 < M_e < \bar{X}$)
- $\beta_1 < 0 \rightarrow$ La courbe est oblique à droite (distribution étalée à gauche : $\bar{X} < M_e < M_0$)

b. Un second plus élaboré (β_2):

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

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

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

En effet :

- $\mu_3(x) > 0 \rightarrow$ La courbe est oblique à gauche (distribution étalée à droite : $M_0 < M_e < \bar{X}$)
- $\mu_3(x) < 0 \rightarrow$ La courbe est oblique à droite (distribution étalée à gauche : $\bar{X} < M_e < M_0$)

2.2 Utilisation des coefficients

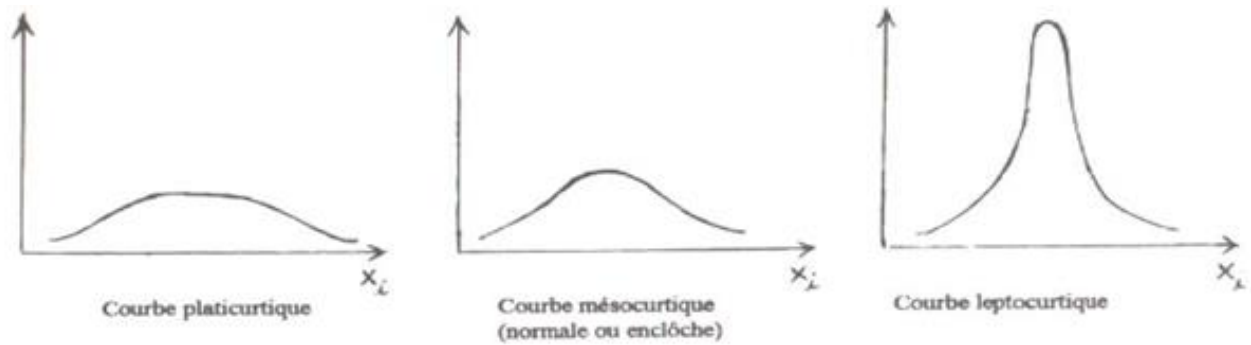
- Les coefficients d'asymétrie ne sont valables que si on possède une population nombreuse et une distribution unimodale.
- On utilise de préférence le coefficient de Fisher F_1 lorsqu'on connaît les cas extrêmes.
- Il est préférable d'utiliser le coefficient de Yule, si on ne connaît pas les cas extrêmes (classes ouvertes), ou si on veut écarter ou on doute des cas extrêmes. Ce coefficient permet d'écarter les cas extrêmes.
- Le coefficient de Pearson β_1 est rarement utilisé, son intérêt est de donner une première idée sur la dissymétrie de la distribution.

3. Aplatissement

Comme pour l'asymétrie, on prend comme référence de la courbe « normale ». Une distribution est plus ou moins aplatie selon que les fréquences des valeurs voisines des valeurs centrales sont plus ou moins élevées par rapport aux autres.

- Une courbe plus aplatie que la courbe normale est dite « **platicurtique** »
- Une courbe moins aplatie que la courbe normale est dite « **leptocurtique** »

Chapitre 5 : Les paramètres de forme



3.1. Principaux coefficients d'aplatissement

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

3.1.1. Coefficient de Pearson

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)}$$

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

C'est pourquoi le coefficient β_3 est toujours supérieur à 1.

- Si $\beta_3 = 3 \rightarrow$ courbe normale
- Si $\beta_3 > 3 \rightarrow$ courbe leptocurtique (moins aplatie que la courbe normale).
- Si $\beta_3 < 3 \rightarrow$ courbe platycurtique (plus aplatie que la courbe normale).

Sa valeur dépasse rarement 5.

3.1.2. Coefficient de Fisher :

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

C'est un coefficient sans dimension, invariant par changement de variable et nul pour les distributions symétriques.

- Si $F_2 = 0 \rightarrow$ courbe normale (mésocurtique).
- Si $F_2 > 0 \rightarrow$ courbe leptocurtique (moins aplatie que la courbe normale).
- Si $F_2 < 0 \rightarrow$ courbe platycurtique (plus aplatie que la courbe normale).

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

Remarque :

La constante 3 est choisie de telle façon que le coefficient soit nul lorsque la distribution est normale. En effet, lorsque la distribution est normale, $\mu_4(X) = 3$ et $\mu_2(X) = 1$ donc $\sigma(X) = 1$.

4. Exercices corrigés

Exercice 01 :

Répartition de 2000 assurés en fonction du nombre X d'accidents d'automobiles signalés à leur compagnie d'assurance en 5 ans. X peut varier théoriquement par valeurs entières de 0 à l'infini.

Nombre d'accident Xi	0	1	2	3	4	5	6 et +
Nombre d'assurés ni	747	718	376	118	31	6	4

- 1) Calculer les moments simples m_1 et m_2 .
- 2) Calculer les moments centrés $\mu_2(x)$ et $\mu_4(x)$.
- 3) Déterminer le mode
- 4) Calculer la moyenne et l'écart type.
- 5) Calculer le coefficient d'asymétrie β_1
- 6) Calculer le coefficient d'aplatissement β_3

Corrigé exercice 01

Xi : nombre d'accident	n_i	$n_i \times x_i$	$n_i \times x_i^2$	$n_i(X_i - \bar{X})^4$
0	747	0	0	747
1	718	718	718	0
2	376	752	1504	376
3	118	354	1062	1888
4	31	124	496	2511
5	6	30	150	1536
6	4	24	144	2500
Total	2000	2002	4074	9558

1) Calcul des moments simples m_1 et m_2

$$m_1 = \frac{\sum n_i x_i}{N} = \frac{2002}{2000} = 1,001 = \bar{X}$$

$$m_2 = \frac{\sum n_i x_i^2}{N} = \frac{4074}{2000} = 2,037$$

2) Calcul des moments centrés $\mu_2(X)$ et $\mu_4(X)$

$$\mu_2(X) = V(X) = m_2 - m_1^2 = \frac{\sum n_i x_i^2}{N} - \bar{X}^2 = \frac{4074}{2000} - (1,001)^2 = 1,035$$

$$\mu_4(X) = \frac{\sum n_i (X_i - \bar{X})^4}{N} = \frac{9558}{2000} = 4,779$$

3) Détermination du mode

Le mode : $M_0 = 0$ accident car : (Max $n_i = 747$)

4) Calcul de la moyenne

$$\bar{X} = \frac{\sum n_i x_i}{N} = \frac{2002}{2000} = 1,001 \approx 1 \text{ accident}$$

Calcul de l'écart type :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{1,035} = 1,02 \approx 1 \text{ accident}$$

5) Calcul du coefficient d'asymétrie de Pearson β_1

$$\beta_1 = \frac{\bar{X} - M_0}{\sigma(X)} = \frac{1,001 - 0}{1,02} = 0,98$$

Le coefficient β_1 est très proche de la valeur 1, ce qui prouve une forte asymétrie. La représentation graphique de la courbe de fréquences montre bien que le M_0 est situé à gauche de médiane puisque l'étalement est à droite ($\beta_1 > 0$)

Chapitre 5 : Les paramètres de forme

6) Calcul du coefficient d'aplatissement de Pearson β_3

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)}$$

$$\mu_4(X) = \frac{9558}{2000} = 4,779$$

$$\sigma^4(X) = (1,02)^4 = 1,082$$

$$\text{D'où : } \beta_3 = \frac{4,779}{1,082} = 4,42.$$

Comme ($\beta_3 > 3$), donc la distribution est moins aplatie que la courbe normale, c'est donc une courbe leptocurtique.

Exercice 02 :

Soit la série suivante concernant la distribution de 100 salariés selon le salaire horaire en DA

Xi en DA	[20-25[[25-30[[30-35[[35-40[[40-45[[45-50[
Effectif	9	29	18	6	26	12

- 1) Calculer le mode
- 2) Calculer le salaire horaire moyen
- 3) Calculer la variance et l'écart-type.
- 4) Calculer les principaux paramètres de position
- 5) Calculer les paramètres d'asymétrie et d'aplatissement.

Corrigé exercice 02

Xi : salaire horaire	n_i	$n_i \uparrow$	c_i	$n_i c_i$	$n_i c_i^2$	$n_i c_i^3$	$n_i c_i^4$	$n_i (C_i - \bar{X})^4$
[20-25[9	9	22,5	202,5	4556,25	102515,63	2306601,56	209368,017
[25-30[29	38	27,5	797,5	21931,25	603109,38	16585507,8	84634,4847
[30-35[18	56	32,5	585	19012,5	617906,25	20081953,1	548,964113
[35-40[6	62	37,5	225	8437,5	316406,25	11865234,4	295,893037
[40-45[26	88	42,5	1105	46962,5	1995906,3	84826015,6	89046,9582
[45-50[12	100	47,5	570	27075	1286062,5	61087968,8	307286,406
Total	100	-	-	3485	127975	4921906,3	196753281	691180,723

1) Calcul du Mode

$$Mo \in [25 - 30[$$

$$M_o = X_{min} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times ai$$

$$M_o = 25 + \frac{(29-9)}{(29-9)+(29-18)} \times 5 = 28,225 \text{ DA}$$

2) Calcul du salaire horaire moyen :

$$\bar{X} = \frac{\sum n_i c_i}{N} = \frac{3485}{100} = 34,85 \text{ DA}$$

3) Calcul de la variance et de l'écart-type.

• Calcul de la variance

$$V(X) = \frac{\sum n_i \times c_i^2}{N} - \bar{X}^2 = \frac{127975}{100} - (34,85)^2 = 65,227$$

• Calcul de l'écart type :

$$\sigma(X) = \sqrt{V(X)} = \sqrt{65,227} = 8,076 \text{ DA}$$

4) Calcul des principaux paramètres de position

$$Q_1 = X_{min} + \frac{\frac{N}{4} - n_{i_{Q_1-1}}}{n_{i_{Q_1}}} \times ai \rightarrow Q_1 = 25 + \frac{25-9}{29} \times 5 = 27,758 \text{ DA}$$

$$Q_2 = X_{min} + \frac{\frac{N}{2} - n_{i_{Q_2-1}}}{n_{i_{Q_2}}} \times ai \rightarrow Q_2 = 30 + \frac{50-38}{18} \times 5 = 33,333 \text{ DA}$$

$$Q_3 = X_{min} + \frac{N(\frac{3}{4}) - n_{i_{Q_3-1}}}{n_{i_{Q_3}}} \times ai \rightarrow Q_3 = 40 + \frac{75-62}{26} \times 5 = 42,5 \text{ DA}$$

Chapitre 5 : Les paramètres de forme

$$D_1 = X_{min} + \frac{\frac{N}{10} - n_{iD_1-1}^\dagger}{n_{iD_1}} \times ai \rightarrow D_1 = 25 + \frac{10-9}{29} \times 5 = 25,172 \text{ DA}$$

$$D_9 = X_{min} + \frac{N \times \left(\frac{9}{10}\right) - n_{iD_9-1}^\dagger}{n_{iD_9}} \times ai \rightarrow D_9 = 45 + \frac{90-88}{12} \times 5 = 45,833 \text{ DA}$$

5) Calculer des paramètres d'asymétrie et d'aplatissement

5.1) Coefficients d'asymétrie

- Coefficient de Yule :

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

$$C_y = \frac{42,5 - 2(33,333) + 27,758}{(42,5 - 27,758)} = \frac{3,592}{14,742} = 0,243$$

$C_y > 0$ Étalement à droite (M_0 est à gauche de la médiane).

- Coefficient de Pearson β_1

$$\beta_1 = \frac{\bar{X} - Mo}{\sigma(X)} = \frac{34,85 - 28,225}{8,076} = 0,82$$

$\beta_1 > 0$ Donc l'étalement à droite (M_0 est à gauche de M_e)

5.2) Coefficients d'aplatissement

- Coefficient de Pearson β_3

$$\beta_3 = \frac{\mu_4(X)}{\mu_2^2(X)} = \frac{\mu_4(X)}{\sigma^4(X)}$$

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

$$\mu_4(X) = \frac{\sum n_i c_i^4}{N} - 4 \left(\frac{\sum n_i c_i}{N} \times \frac{\sum n_i c_i^3}{N} \right) + 6 \left(\left(\frac{\sum n_i c_i}{N} \right)^2 \times \frac{\sum n_i c_i^2}{N} \right) - 3 \left(\frac{\sum n_i c_i}{N} \right)^4$$

$$\mu_4(X) = 1967532,81 - 4(34,85 \times 49219,063) + 6((34,85)^2 \times 1279,75) - 3(34,85)^4$$

$$\mu_4(X) = 6911,73$$

$$\text{Ou bien : } \mu_4(X) = \frac{\sum n_i (C_i - \bar{X})^4}{N} = \frac{691180,723}{100} = 6911,807$$

$$\sigma^4(X) = 4253,88$$

$$\beta_3 = \frac{\mu_4(X)}{\sigma^4(X)} = \frac{6911,73}{4253,88} = 1,62$$

$\beta_3 < 3$ Donc la courbe platicurtique (plus aplatie que la courbe normale).

Chapitre 6 : Les paramètres de concentration

Introduction :

La mesure de la concentration vise à quantifier les effets de la dispersion au sein d'un ensemble de données. En économie, elle joue un rôle primordial, notamment pour analyser la répartition des richesses, des salaires ou encore la répartition des parts de marché entre entreprises. Contrairement à des variables continues comme la taille, cette mesure s'applique principalement à des variables continues et positives dont la somme a une signification économique réelle, comme les salaires. Il existe deux méthodes de détermination de la concentration : par le calcul et par les graphes.

1. Détermination de la concentration par le calcul

La démarche est la suivante :

- 1) On calcule la médiane (Me) de la série
- 2) On calcule la médiale (MI) que nous définissons plus bas.
- 3) On mesure l'écart (ΔM) entre la médiale et la médiane.
- 4) On compare cet écart (ΔM) à l'étendue.

1.1. Détermination de la médiane (Me)

Nous savons effectuer ce calcul (voir chapitre 3), qui passe par la résolution d'une interpolation linéaire.

1.2. Détermination de la médiale (MI)

Définition pratique :

La médiale est la valeur du caractère étudié qui partage la masse globale (la somme des produits $n_i c_i$) en deux parties égales. C'est une caractéristique de valeur centrale.

La médiale de la série (c_i, n_i) est la médiane de la série $(c_i ; n_i c_i)$. ($n_i c_i$: appelée la masse globale).

Si X_i représente le salaire et n_i le nombre de salariés, alors $n_i c_i$ correspond à la masse salariale totale versée. Le salaire médial est ainsi le salaire qui divise la masse salariale globale en deux parties égales. Il indique également la répartition des salariés entre ceux qui se partagent la première moitié de la masse salariale et ceux qui se partagent la seconde moitié.

Le principe du calcul de la médiale est exactement le même que celui de la médiane, mis à part le fait qu'on l'applique, non plus sur la colonne des fréquences cumulées des individus $f_i \uparrow$ mais sur celle des fréquences cumulées de la masse

globale $f'_i \uparrow \left(f'_i = \frac{n_i c_i}{\sum n_i c_i} \right)$

Exemple :

On donne le tableau suivant concernant un échantillon de 40 salariés d'une entreprise selon leurs salaires mensuels en MDA. D'abord, il faut former la colonne des $n_i c_i$ (la masse salariale) et celle des fréquences cumulées des $n_i c_i$ ($f'_i \uparrow$).

Salaires (MDA)	n_i	Centre de Classe c_i	f_i $\left(\frac{n_i}{N} \right)$	$f_i \uparrow$	Masse salariale $n_i c_i$	f'_i $\left(\frac{n_i c_i}{\sum n_i c_i} \right)$	$f'_i \uparrow$
[10-20[5	15	0,125	0,125	75	0,052	0,052
[20-30[7	25	0,175	0,300	175	0,121	0,173
[30-40[12	35	0,300	0,600	420	0,290	0,463
[40-50[10	45	0,250	0,850	450	0,310	0,773
[50-60[6	55	0,150	1,000	330	0,227	1,000
Total	40	-	1	-	1450	1	-

Chapitre 6 : Les paramètres de concentration

La classe médiale est la classe [40-50[

$$Ml = X_{\min} + \frac{0,5 - f'_{i_{Ml-1}}}{f'_{i_{Ml}}} \times ai$$

$$Ml = 40 + \frac{0,5 - 0,463}{0,31} \times 10 = 41,19 \text{MDA}$$

La médiale de la distribution des salaires est donc la valeur du salaire qui partage la masse salariale en deux parties égales : Dès lors, le salaire médial est tel que les salariés qui se situent en deçà, gagnent autant que les salariés qui se situent au-delà.

1.3. L'écart (médiale – médiane)

La médiale est supérieure à la médiane (puisque l'on raisonne « en masse » dans le premier cas).

$$\Delta M = Ml - Me. \text{ Le calcul est immédiat.}$$

1.4. Comparaison de ΔM à l'Étendue.

- Si ΔM est grand par rapport à l'étendue, la concentration est forte (Dans l'exemple des salaires, cela signifierait que l'inégalité entre les salaires est forte).
- Si ΔM est petit par rapport à l'étendue, la concentration est faible (Dans l'exemple des salaires, cela signifierait qu'il n'y a pas de grandes disparités salariales entre les classes des salaires).
- Si ΔM est nul, la médiane est égale à la médiale ; on se trouve donc dans une situation d'égalité parfaite ou d'équi-répartition, si les classes sont bien choisies. (Dans l'exemple des salaires, tous les salariés toucheraient le même salaire).

1.5. Le coefficient de concentration (Cc)

Pour évaluer la concentration, on rapporte l'écart généralement à l'étendue de la série :

$$C_c = \frac{\Delta M}{E} \times 100$$

Plus la valeur de ce coefficient est élevée plus la concentration est importante et inversement, ainsi si $\frac{\Delta M}{E} = 0$ signifie que la concentration est nulle et la distribution parfaitement égalitaire.

Pour notre exemple :

Nous avons les calculs suivants :

$$Me = X_{\min} + \frac{0,5 - f'_{i_{Me-1}}}{f'_{i_{Me}}} \times ai$$

$$\rightarrow Me = 30 + \frac{0,5 - 0,3}{0,3} \times 10 = 36,66 \text{MDA}$$

$$\bullet \Delta M = Ml - Me \rightarrow \Delta M = 41,19 - 36,66 = 4,53 \text{MDA.}$$

Cet écart ΔM traduit la concentration

$$\bullet \text{ On calcule l'étendue : } E = X_{\max} - X_{\min} = 60 - 10 = 50 \text{MDA}$$

• **On compare : ΔM à E :**

On remarque que ΔM est petit par rapport à l'étendue, donc la concentration des salaires est faible. Elle est de l'ordre de 9%. $\left(\frac{\Delta M}{E} = \frac{4,53}{50} = 0,0906 \approx 9\% \right)$

Chapitre 6 : Les paramètres de concentration

2. Détermination de la concentration par le graphe

Cette analyse a été développée par l'italien **Corrado Gini** au cours de ses travaux sur les disparités de revenus. Elle a conduit à la création de l'indice de Gini, un ratio qui quantifie le degré de **concentration** ou **d'inégalité**. Cet indice est dérivé de la courbe de Lorenz, élaborée auparavant par l'économiste américain Max O. Lorenz pour représenter **visuellement la concentration**.

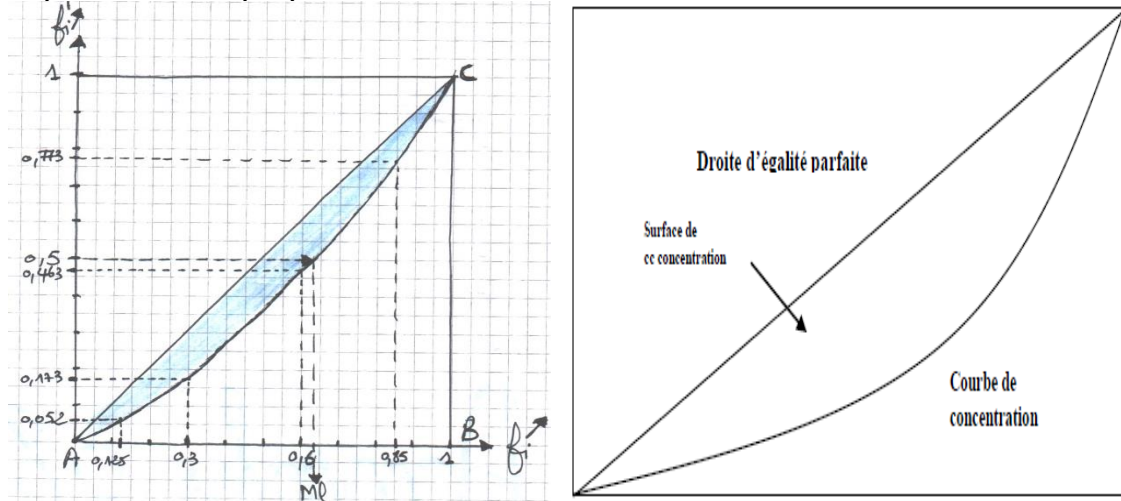
2.1. La courbe de concentration (ou de LORENZ) :

C'est la courbe obtenue en représentant :

***En abscisse**, on utilise les fréquences cumulées des individus ($f_i \uparrow$)

***En ordonnée**, on utilise les fréquences cumulées de la masse globale ($f_i' \uparrow$)

Reprenons l'exemple précédent :



On voit donc que 60% des salariés se partagent seulement 46,3% de la masse salariale. La concentration reste néanmoins faible, si l'on en juge par la surface de l'aire grisée, de façon uniquement visuelle.

La diagonale (AC) représente la ligne d'équité parfaite. Autrement dit, chaque pourcentage de salariés reçoit la même proportion de la masse salariale.

Par exemple : 10 % des salariés perçoivent 10 % de la masse salariale, 20 % en perçoivent 20 %, et ainsi de suite jusqu'à 100 %.

2.2. L'indice de Gini :

L'indice de Gini est un ratio utilisé pour mesurer le degré de concentration ou d'inégalité dans une distribution. Il est défini comme le rapport entre deux surfaces : celle située entre la courbe de Lorenz et la diagonale d'égalité parfaite (surface de concentration), et la surface totale située sous cette diagonale (surface du triangle). Sa valeur varie entre 0 et 1 (d'une concentration nulle à une concentration maximale)

$$IG = \frac{\text{surface de concentration}}{\text{surface du triangle ABC}}$$

$$\text{En effet: la surface du triangle ABC} = \frac{(1 \times 1)}{2} = \frac{1}{2}$$

D'où : l'indice de Gini :

$$IG = \frac{\text{surface de concentration}}{\frac{1}{2}} = 2 \times \text{surface de concentration}$$

I_G Varie de 0 à 1 (d'une concentration nulle à une concentration maximale).

Le problème est de mesurer les surfaces sans avoir recours au calcul intégral. Plusieurs méthodes graphiques sont possibles. La plus simple consiste à compter les carreaux sur le graphique que l'on aura soigneusement construit sur papier millimétré.

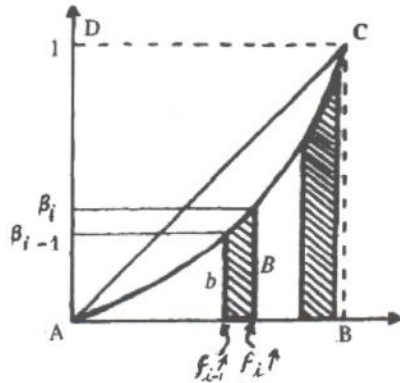
Chapitre 6 : Les paramètres de concentration

Cependant la représentation graphique a essentiellement pour objectif de transmettre un message visuel. Elle n'est que la visualisation de la concentration mesurée par le calcul.

Néanmoins, si l'on tient absolument à calculer une valeur numérique de I_G , on peut se servir (entre autres méthodes d'approximation) de celle donnée par la méthode des « Trapèzes » :

Méthode des Trapèzes

On peut concevoir qu'il existe autant de trapèzes que de classes, comme le montre la figure ci-dessus :



Rappelons que la surface d'un trapèze est donnée par :

$$S = \frac{(b+B)h}{2}$$

La surface de concentration = la surface du triangle ABC - \sum des surfaces des trapèzes

$$\begin{aligned} \text{surface de concentration} &= \frac{1}{2} - \sum S_i \\ &= \frac{1}{2} - \sum \frac{(\text{base inf} + \text{base sup}) \times \text{hauteur}}{2} \\ &= \frac{1}{2} - \sum \frac{(b+B) \times h}{2} \\ &= \frac{1}{2} - \frac{1}{2} \sum (B_{i-1} + B_i) \times (f_i \uparrow - f_{i-1} \uparrow) \\ &= \frac{1}{2} - \frac{1}{2} \sum (B_{i-1} + B_i) \times f_i \end{aligned}$$

Avec : β_i est la valeur de $f_i \uparrow$ de la ligne « i » du tableau.

Et : β_{i-1} est la valeur précédente. ($\beta_{i-1} = 0$ pour la valeur $i=1$).

$$\text{En règle générale : } \beta_i = \sum_{h=1}^i \frac{n_h c_h}{\sum_i n_i c_i}$$

Nous avons : $I_G = 2 \times \text{surface de concentration}$

$$\text{Donc : } I_G = 2 \left(\frac{1}{2} - \frac{1}{2} \sum (\beta_{i-1} + \beta_i) \times f_i \right) \Rightarrow I_G = 1 - \sum (\beta_{i-1} + \beta_i) \times f_i$$

Il suffit de disposer les calculs comme suit :

β_{i-1}	β_i	$\beta_{i-1} + \beta_i$	f_i	$(\beta_{i-1} + \beta_i) \times f_i$
0	0,052	0,052	0,125	0,0065
0,052	0,173	0,225	0,175	0,039375
0,173	0,463	0,636	0,300	0,1908
0,463	0,773	1,236	0,250	0,309
0,773	1,000	1,773	0,150	0,26595
-	-	-	-	0,811625

I_G est donné par : 1 moins la somme des termes de la dernière colonne.

$$\text{Donc : } I_G = 1 - \sum (\beta_{i-1} + \beta_i) \times f_i \Rightarrow I_G = (1 - 0,811625) = 0,188375 \approx 18,84\%$$

Chapitre 6 : Les paramètres de concentration

3. Exercice corrigé :

On donne le tableau suivant concernant un échantillon de 43 locataires selon leurs classes de salaires net en milliers de dinars (MDA) :

Salaires (MDA)	[40 - 60[[60 - 80[[80 - 100[[100-120[[120-140[
Nombre de locataires	5	8	12	10	8

- 1- Calculer la médiane
- 2- Calculer la médiale
- 3- Calculer l'écart ΔM (médiale – médiane)
- 4- Comment elle est la mesure de concentration des salaires ?
- 5- Calculer l'indice de Gini et interpréter le résultat

Le corrigé :

On forme d'abord la colonne des $n_i c_i$ et celle des fréquences cumulées des $n_i c_i$ ($f_i' \uparrow$)

Salaires MDA	n_i	Centre de Classe c_i	f_i $\left(\frac{n_i}{N}\right)$	$f_i \uparrow$	Masse salariale $n_i c_i$	f_i' $\left(\frac{n_i c_i}{\sum n_i c_i}\right)$	$f_i' \uparrow$
[40 - 60[5	50	0,11	0,11	250	0,06	0,06
[60 - 80[8	70	0,19	0,30	560	0,14	0,20
[80 - 100[12	90	0,28	0,58	1080	0,27	0,47
[100-120[10	110	0,23	0,81	1100	0,27	0,74
[120-140[8	130	0,19	1,00	1040	0,26	1,00
Total	43	-	1	-	4030	1	-

1) Calcul de la médiane

La classe médiane est la classe [80-100[

$$Me = X_{\min} + \frac{0,5 - f_{i_{Me-1}} \uparrow}{f_{i_{Me}}} \times ai \qquad Me = 80 + \frac{0,5 - 0,30}{0,28} \times 10 = 94286 \text{DA}$$

2) Calcul de la médiale

La classe médiale est la classe [100-120[

$$Ml = X_{\min} + \frac{0,5 - f'_{i_{Ml-1}} \uparrow}{f'_{i_{Ml}}} \times ai \qquad Ml = 40 + \frac{0,5 - 0,47}{0,27} \times 10 = 102222 \text{DA}$$

3) Calcul de l'écart ΔM (médiale – médiane)

$$\Delta M = Ml - Me = 102222 - 94286 = 7936 \text{DA}$$

Cet écart ΔM traduit la concentration. On le compare à l'étendue.

4) Mesure de la concentration

$$\text{Etendue} : E = X_{\max} - X_{\min} = 140000 - 40000 = 100000 \text{DA}$$

On remarque que : ΔM est petit par rapport à l'étendue, donc la concentration est faible. Dans notre exemple, nous avons $\frac{\Delta M}{E} = \frac{7936}{100000} = 0,079 \approx 8\%$

Comme la concentration des salaires est l'ordre de 8%, donc elle est faible.

5) Calcul de l'indice de Gini

Pour calculer l'indice de Gini, il suffit de disposer les calculs comme suit

β_{i-1}	β_i	$\beta_{i-1} + \beta_i$	f_i	$(\beta_{i-1} + \beta_i) \times f_i$
0	0,06	0,06	0,11	0,0066
0,06	0,20	0,26	0,19	0,0494
0,20	0,47	0,67	0,28	0,1876
0,47	0,74	1,21	0,23	0,2783
0,74	1,00	1,74	0,19	0,3306
-	-	-	-	0,8525

Chapitre 6 : Les paramètres de concentration

$I_G = 2 \times$ surface de concentration

$$I_G = 1 - \sum (\beta_{i-1} + \beta_i) \times f_i$$

$$I_G = 1 - 0,8525 = 0,1475$$

$I_G = 14,75\%$, la concentration est néanmoins faible.

Exercice 2 :

Le tableau suivant donne la répartition du volume des importations de légumes secs réalisées par certaines entreprises nationales au niveau du marché mondial :

Volume des importations (10 ³ Tonnes)	[40-45[[45-55[[55-70[[70-90[[90-300[
Nombre d'entreprises	8	15	47	22	8

- Déterminer la population statistique, le caractère et sa nature.
- Calculer le volume moyen des importations des légumes secs.
- Calculer le volume médian des importations des légumes secs et interpréter le résultat.
- Calculer la médiale.
- Comment elle est la mesure de concentration des importations des légumes secs ?
- Tracer la courbe de Lorenz et calculer l'indice de Gini.

Corrigé de l'exercice 2 :

Le tableau des fréquences :

Volume des importations : (10 ³ Tonnes)	n_i	c_i	f_i	$f_i \uparrow$	$n_i c_i$	f'_i (%) $\left(\frac{n_i c_i}{\sum n_i c_i} \right)$	$f'_i \uparrow$ (%)
[40 - 45[8	42,5	0,08	0,08	340	4,62	4,62
[45 - 55[15	50	0,15	0,23	750	10,20	14,82
[55 - 70[47	62,5	0,47	0,70	2937,5	40,00	54,82
[70 - 90[22	80	0,22	0,92	1760	23,95	78,77
[90-300[8	195	0,08	1,00	1560	21,31	100,00
Total	100	-	1	-	7347,5	100	

1) **La population statistique** : les 100 entreprises nationales

Le caractère : le volume des importations des légumes secs

Sa nature : quantitatif continu.

2) **Calcul du volume moyen des importations des légumes secs :**

$$\bar{X} = \frac{\sum n_i c_i}{N} = \frac{7347,5}{100} = 73,475 \times 10^3 \text{ Tonnes}$$

3) **Calcul du volume la médian des importations des légumes secs :**

La classe médiane est la classe [55-70[

$$Me = X_{\min} + \frac{0,5 - f_{i_{Me-1}}}{f_{i_{Me}}} \times ai$$

$$M_e = 55 + \frac{0,5 - 0,23}{0,47} \times 15 = 63,617 \times 10^3 \text{ Tonnes}$$

Il y a 50% des entreprises nationales qui importent au moins 63617 Tonnes des légumes secs.

Chapitre 6 : Les paramètres de concentration

4) Calcul de la médiale

La classe médiale est la classe [55-70[

$$M_l = X_{\min} + \frac{0,5 - f'_{i_{Ml-1}}}{f'_{i_{Ml}}} \times ai$$

$$M_l = 55 + \frac{50 - 14,82}{40} \times 15 = 68,19 \times 10^3 \text{ Tonnes}$$

5) Mesure de la concentration

On calcule d'abord l'écart ΔM (médiale – médiane)

$$\Delta M = M_l - M_g = 68,19 - 63,62 = 4,57 \times 10^3 \text{ Tonnes}$$

Cet écart ΔM traduit la concentration. On le compare à l'étendue.

$$\text{Etendue} = X_{\max} - X_{\min} = 300 - 40 = 260 \times 10^3 \text{ Tonnes}$$

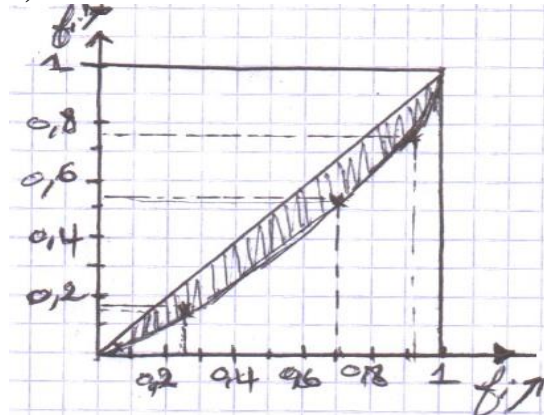
On remarque que :

ΔM est petit par rapport à l'étendue, donc la concentration des importations des légumes secs est faible.

$$\text{Dans notre exemple, nous avons : } Cc = \frac{\Delta M}{E} = \frac{4,57}{260} = 1,75\%.$$

Comme la concentration des importations des légumes secs est l'ordre de 1,75%, donc elle est faible.

6) La courbe de Lorenz



Calcul de l'indice de Gini

Pour calculer l'indice de Gini, il suffit de disposer les calculs comme suit :

β_{i-1}	β_i	$(\beta_{i-1} + \beta_i)(\%)$	$f_i(\%)$	$(\beta_{i-1} + \beta_i) \times f_i$
0	4,62	4,62	8	36,96
4,62	14,82	19,44	15	291,6
14,82	54,82	69,64	47	3273,08
54,82	78,77	133,59	22	2938,98
78,77	100,00	178,77	8	1430,16
-	-	-	100	7970,78

$$I_G = 2 \times \text{surface de concentration}$$

$$I_G = 1 - \sum (\beta_{i-1} + \beta_i) \times f_i$$

$$I_G = 1 - 0,797078 = 0,2029$$

$$I_G = 20,29\%.$$

Interprétation

Avec un $I_G = 20,29\%$, la concentration des importations de légumes secs est faible. Cela signifie que les volumes sont répartis de manière relativement équitable entre les entreprises nationales, sans qu'une seule ou quelques entreprises ne monopolisent le marché.

Chapitre 7 : Les indices

Introduction :

Les indices sont des outils statistiques sans dimension qui servent à comparer des observations quantitatives effectuées dans des contextes variés, tels que des périodes, des lieux, des groupes ou des situations distinctes. En offrant une mesure normalisée, ils permettent d'évaluer facilement les variations et les tendances. Ces indices sont généralement exprimés sous forme de pourcentages, ce qui les rend particulièrement intuitifs pour l'analyse et l'interprétation des données.

1. Les indices élémentaires ou simples :

On fait recours aux indices simples lorsqu'on souhaite mesurer l'évolution d'une seule variable ou d'un seul élément au fil du temps, en comparant sa valeur à une période de base. Cela permet de calculer le changement relatif d'une grandeur (prix, quantité, production, etc.) entre deux périodes.

1.1. Définition :

L'indice élémentaire est le plus simple des indices. Il est égal au rapport de deux valeurs prises par une **même grandeur** mesurées à deux dates différentes ou à deux endroits différents. Lorsqu'il compare des valeurs dans le temps, on parle d'indice chronologique (ou temporel) et lorsqu'il compare des valeurs dans l'espace (lieu), on parle d'indice spatial.

$I_{1/0} = \frac{V_1}{V_0} \times 100$	L'indice 0 correspond à l'année de base ou de référence. L'indice 1 correspond à l'année courante.
--	---

Exemple :

Le prix du café est passé de 180 DA en 2017 à 200 DA en 2018. Déterminez l'indice simple du prix du café pour l'année 2018, en prenant l'année 2017 comme année de base.

Nous calculons alors : $I_{18/17} = \frac{P_{18}}{P_{17}} \times 100 = \frac{200}{180} \times 100 = 111,11\%$.

Le prix du café a augmenté de 11,11% entre 2018 et 2017, soit : $(111,11\% - 100\% = +11,11\%)$.

1.2. Propriétés :

Les indices élémentaires possèdent deux principales propriétés

1.2.1. Réversibilité :

La réversibilité consiste à permuter la valeur courante de la grandeur et sa valeur référence. Un indice est réversible si et seulement si : $I_{1/0} = \frac{1}{I_{0/1}}$

$$D'où : I_{1/0} \times I_{0/1} = 100^2$$

Exemple :

Reprenons le prix du café. Si nous désirons calculer l'indice élémentaire au prix du café, janvier 2017/ janvier 2018.

On a déjà trouvé : $I_{18/17} = 111,11\%$.

Calculons maintenant : $I_{17/18} = \frac{I_{17}}{I_{18}} \times 100 = \frac{180}{200} \times 100 = 90\% = 0,9$

La réversibilité est bien vérifiée car : $I_{18/17} = \frac{1}{I_{17/18}} = \frac{1}{0,9} \times 100 = 111,11\%$

Et $I_{17/18} \times I_{18/17} = 90 \times 111,11 = 9999,9 \approx 100^2$

1.2.2. Transférabilité (circularité) :

C'est la propriété essentielle des indices, qui se traduit par une relation multiplicative.

Un indice est transférable si et seulement si : $I_{2/0} = \frac{V_2}{V_1} \times \frac{V_1}{V_0} = I_{2/1} \times I_{1/0}$

D'où : $I_{2/0} = \frac{I_{2/1} \times I_{1/0}}{100^{2-1}}$ (indice de variation)

La formule générale de cette propriété de Transférabilité s'écrit alors :

$$I_{t/0} = I_{t/t-1} \times I_{t-1/t-2} \times \dots \times I_{1/0}$$

Chapitre 7 : Les indices

Si l'on veut calculer l'indice de la situation 3 en prenant pour base la situation 0, on utilise le produit suivant : $I_{3/0} = I_{3/2} \times I_{2/1} \times I_{1/0}$

Exemple :

En novembre 2013, le prix du fromage est de 500 DA. En novembre 2014, il atteint 600 DA, puis en novembre 2015, il s'élève à 850 DA.

- Calculer l'indice de variation du prix du fromage en novembre 2014, base 100 en novembre 2013.
- Calculer l'indice de variation du prix du fromage en novembre 2015, base 100 en novembre 2014.
- Calculer l'indice de variation du prix du fromage en novembre 2015, base 100 en novembre 2013 et vérifier la condition de transférabilité.

Corrigé :

$$I_{2014/2013} = \frac{600}{500} \times 100 = 120\% \Rightarrow \text{Le prix a donc augmenté de } 20\%$$

$$I_{2015/2014} = \frac{850}{600} \times 100 = 141,67\% \Rightarrow \text{Le prix a donc augmenté de } 41,67\%.$$

$$I_{2015/2013} = \frac{850}{500} \times 100 = 170\% \Rightarrow \text{Le prix a donc augmenté de } 70\%.$$

Nous remarquons que la propriété de transférabilité est bien vérifiée puisque :

$$I_{2015/2013} = \frac{I_{15/14} \times I_{14/13}}{100^{2-1}} = \frac{120 \times 141,67}{100} = 170\%.$$

Conclusion :

Nous constatons que les taux de variation ne s'ajoutent et ne se retranchent jamais. En effet, l'augmentation de novembre 2013 à novembre 2015 (+70%) est **bien différente** de (20%+41,67%) ; ce qui donne une augmentation de 61,67%.

2. Les indices synthétiques

Les indices élémentaires permettent de suivre l'évolution d'une grandeur homogène et bien définie. Cependant, pour analyser des phénomènes plus complexes, tels que le niveau général des prix ou le volume des échanges commerciaux, on utilise des indices synthétiques. Ces derniers résument les variations des grandeurs complexes à l'aide de paramètres de tendance centrale calculés à partir des indices élémentaires associés.

2.1. Les indices synthétiques simples

Définition :

L'indice synthétique simple est un rapport entre deux sommes d'une **même grandeur**, mesurées à deux périodes différentes, l'une de base et l'autre en cours. L'indice est alors déterminé à l'aide de la formule suivante :

$$I_p = I_{1/0}^p = \frac{\sum p_1}{\sum p_0} \times 100 \text{ (Indice prix);}$$

$$I_q = I_{1/0}^q = \frac{\sum q_1}{\sum q_0} \times 100 \text{ (Indice quantité).}$$

Exemple : Le tableau suivant donne les prix des légumes secs achetés par une famille en 2015 et 2016.

Année	2015	2016
Légumes secs	Prix en DA/kg	Prix en DA/Kg
Riz	150	200
Lentilles	250	350
Pois chiche	450	300
Haricot	350	290
Total	1200	1140

- Calculer l'indice de variation des prix des légumes secs en 2016, base 100 en 2015.

Chapitre 7 : Les indices

Corrigé :

Nous calculons alors : $I_p = I_{16/15}^p = \frac{\sum p_{16}}{\sum p_{15}} \times 100 = \frac{1140}{1200} \times 100 = 95\%$

Les prix des légumes secs ont baissé de 5% entre 2016 et 2015, soit : $(100\% - 95\% = 5\%)$.

2.2. Les indices synthétiques pondérés

Ces indices, qui sont plus représentatifs de la réalité, permettent, tout comme les précédents, de décrire l'évolution d'ensembles parfois très complexes. Ils sont calculés à partir des indices élémentaires, souvent en utilisant des moyennes pondérées. Plusieurs indices ont été conçus par des statisticiens, que nous allons définir ci-dessous :

2.2.1. Indices de Laspeyres

Dans ce type d'indice, les coefficients de pondération sont déterminés en fonction de **l'année de base**. L'indice proposé par Laspeyres en 1844 est simplement la moyenne arithmétique pondérée par des coefficients choisis dans l'année de base, ce qui implique que ces coefficients restent fixes.

Trois types d'indice Laspeyres sont utilisés : un indice des prix, un indice des quantités et un indice des valeurs globales. De plus, deux méthodes permettent de calculer les indices de Laspeyres.

A. Première méthode :

- **Indice Laspeyres des prix :** $L_{1/0}^p = \frac{\sum p_1 q_0}{\sum p_0 q_0} \times 100$
- **Indice Laspeyres des quantités :** $L_{1/0}^q = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times 100$
- **Indice Laspeyres des valeurs globales :** $L_{1/0}^{vg} = \frac{\sum p_1 q_1}{\sum p_0 q_0} \times 100$

(La valeur globale étant égale au produit prix par quantité).

B. Deuxième méthode :

Dans ce cas, l'indice Laspeyres est considérée comme **la moyenne arithmétique** des indices élémentaires, pondérée par les **coefficients budgétaires de l'année de base (CB_0)**. Ces coefficients budgétaires correspondent aux poids relatifs de chaque produit dans la consommation totale, ce qui permet de déterminer l'importance relative de chaque poste budgétaire dans la dépense de consommation totale.

- **Indice Laspeyres des prix :** $L_{1/0}^p = \sum_{j=1}^n CB_{j0} \times I_{1/0}^{p_j}$
- **Indice Laspeyres des quantités :** $L_{1/0}^q = \sum_{j=1}^n CB_{j0} \times I_{1/0}^{q_j}$

2.2.2. Indices de Paasche

Dans ce type d'indice, les coefficients de pondération sont choisis dans **l'année courante**. Comme pour **Laspeyres**, on définit trois indices Paasche (un indice des prix, un indice des quantités, un indice des valeurs globales). De même, deux méthodes permettent de calculer les indices de Paasche.

A. Première méthode :

- **Indice Paasche des prix :** $P_{1/0}^p = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times 100$
- **Indice Paasche des quantités :** $P_{1/0}^q = \frac{\sum q_1 p_1}{\sum q_0 p_1} \times 100$
- **Indice Paasche des valeurs globales :** $P_{1/0}^{vg} = \frac{\sum p_1 q_1}{\sum p_0 q_0} \times 100$

Chapitre 7 : Les indices

B. Deuxième méthode :

L'indice de Paasche est considéré comme la **moyenne harmonique** des indices élémentaires pondérés par les coefficients budgétaires de l'année courante (CB_1).

$$\text{➤ Indice Paasche des prix : } P_{1/0}^p = \frac{\sum_{j=1}^n CB_{j1}}{\sum_{j=1}^n \frac{CB_{j1}}{p_j}} = \frac{\sum_{j=1}^n CB_{j1}}{I_{1/0}^p}$$

$$\text{➤ Indice Paasche des quantités : } P_{1/0}^q = \frac{\sum_{j=1}^n CB_{j1}}{\sum_{j=1}^n \frac{CB_{j1}}{q_j}} = \frac{\sum_{j=1}^n CB_{j1}}{I_{1/0}^q}$$

Exemple : (calcul des indices synthétiques selon la première méthode)

Le tableau suivant donne les prix et les quantités des légumes secs achetés par une famille en 2015 et 2016.

Année	2015		2016	
	Légumes secs	Prix en DA/kg	Quantité en kg	Prix en DA/Kg
Riz	150	5	200	8
Lentilles	250	7	350	10
Pois chiche	450	30	300	35
Haricot	350	10	290	10,5

1. Calculer les trois indices de Laspeyres 2016/2015.
2. Calculer les trois indices de Paasche 2016/2015.

Corrigé

Prenons l'année 2015 comme année de base et l'année 2016 comme année en cours. Pour simplifier l'application des formules, organisons les données dans un tableau de calcul.

Légumes secs	p_{15}	q_{15}	p_{16}	q_{16}	$p_{15} \times q_{15}$	$p_{15} \times q_{16}$	$p_{16} \times q_{15}$	$p_{16} \times q_{16}$
Riz	150	5	200	8	750	1200	1000	1600
Lentilles	250	7	350	10	1750	2500	2450	3500
Pois chiche	450	30	300	35	13500	15750	9000	10500
Haricot	350	10	290	10,5	3500	3675	2900	3045
Total	-	-	-	-	19500	23125	15350	18645

1. Les indices de Laspeyres 2016/2015.

$$L_{16/15}^p = \frac{\sum p_{16} q_{15}}{\sum p_{15} q_{15}} = \frac{15350}{19500} \times 100 = 78,7\%$$

$$L_{16/15}^q = \frac{\sum q_{16} p_{15}}{\sum q_{15} p_{15}} = \frac{23125}{19500} \times 100 = 118,6\%$$

$$L_{16/15}^{vg} = \frac{\sum p_{16} q_{16}}{\sum p_{15} q_{15}} = \frac{18645}{19500} = 0,956 \times 100 = 95,6\%$$

2. Les indices de Paasche 2016/2015.

$$P_{16/15}^p = \frac{\sum p_{16} q_{16}}{\sum p_{15} q_{16}} = \frac{18645}{23125} \times 100 = 80,6\%$$

$$P_{16/15}^q = \frac{\sum q_{16} p_{16}}{\sum q_{15} p_{16}} = \frac{18645}{15350} \times 100 = 121,4\%$$

$$P_{16/15}^{vg} = \frac{\sum p_{16} q_{16}}{\sum p_{15} q_{15}} = \frac{18645}{19500} = 0,956 \times 100 = 95,6\%$$

Chapitre 7 : Les indices

Exemple2 : (calcul des indices synthétiques selon la deuxième méthode)

Les prix et les quantités consommées de trois articles 1, 2 et 3 par un consommateur et qui en constituent ses dépenses en 2016 et 2017 sont résumés dans le tableau suivant :

Année	2016		2017	
	Prix (p_{16})	Quantité(q_{16})	Prix(p_{17})	Quantité(q_{17})
Article 1	160	200	190	270
Article 2	250	150	290	160
Article 3	360	100	360	100

- 1- Calculer l'indice de Laspeyres des prix $L_{17/16}^p$
- 2- Calculer l'indice de Paasche des prix $P_{17/16}^p$
- 3- Calculer l'indice des valeurs globales $I_{17/16}^{vg}$

Corrigé :

Prenons l'année 2016 comme année de base et l'année 2017 comme année courante. Pour simplifier l'application des formules, organisons les données dans un tableau de calcul.

Article	p_{16}	q_{16}	$p_{16} \times q_{16}$	$p_{17} \times q_{16}$	Coefficients budgétaires de l'année de base 2016 (CB16)
Article 1	160	200	32000	38000	$\frac{32000}{105500} = 0,30$
Article 2	250	150	37500	43500	$\frac{37500}{105500} = 0,36$
Article 3	360	100	36000	36000	$\frac{36000}{105500} = 0,34$
Total	-	-	105500	117500	1,00

1-Calcul de l'indice de Laspeyres $L_{17/16}^p$

Calculons d'abord les indices élémentaires de prix :

$$\text{Article 1 : } I_{17/16}^p = \frac{190}{160} \times 100 = 118,8\%$$

$$\text{Article 2 : } I_{17/16}^p = \frac{290}{250} \times 100 = 116\%$$

$$\text{Article 3 : } I_{17/16}^p = \frac{360}{360} \times 100 = 100\%$$

Il est également nécessaire de disposer les pondérations des coefficients budgétaires pour l'année de base 2016 (CB16). Celles-ci sont obtenues en calculant le rapport entre la consommation de chaque article et la consommation totale de l'année 2016. (Voir le tableau ci-dessus).

D'où l'indice Laspeyres des prix :

$$L_{17/16}^p = \sum_{j=1}^3 CB_{j16} I_{17/16}^{pj} \quad (\text{l'article : } j=1,2,3)$$

$$L_{17/16}^p = (0,30 \times 118,8) + (0,36 \times 116) + (0,34 \times 100) = 111,4\%$$

Ce résultat pourrait être vérifié par la première méthode en appliquant directement la formule de Laspeyres.

$$L_{17/16}^p = \frac{\sum p_{17} \times q_{16}}{\sum p_{16} \times q_{16}} = \frac{117500}{105500} \times 100 = 111,4\%.$$

Chapitre 7 : Les indices

2- Calcul de l'indice de Paasche $P_{17/16}^p$

L'indice de Paasche étant la **moyenne harmonique** des indices élémentaires de prix pondérés par les coefficients budgétaires de l'année courante 2017, calculons alors ces coefficients pour l'année 2017 (**CB17**).

Article	p_{17}	q_{17}	$p_{17} \times q_{17}$	$p_{16} \times q_{17}$	Coefficients budgétaires de l'année courante (CB17)
Article 1	190	270	51300	43200	$\frac{51300}{133700} = 0,38$
Article 2	290	160	46400	40000	$\frac{46400}{133700} = 0,35$
Article 3	360	100	36000	36000	$\frac{36000}{133700} = 0,27$
Total	-	-	133700	119200	1,00

D'où l'indice Paasche des prix :

$$P_{17/16}^p = \frac{\sum_{j=1}^3 CB_{j17}}{\sum_{j=1}^3 \frac{p_j}{I_{17/16}^p}}, \quad (\text{l'article : } j=1,2,3)$$

$$P_{17/16}^p = \frac{1}{\frac{0,38}{118,6} + \frac{0,35}{116} + \frac{0,27}{100}} = 112,36\%$$

Ce résultat peut également être obtenu directement à l'aide de la formule de Paasche :

$$P_{17/16}^p = \frac{\sum p_{17} \times q_{17}}{\sum p_{16} \times q_{17}} = \frac{133700}{119200} \times 100 = 112,2\%$$

3- Calcul de l'indice des valeurs globales $I_{17/16}^{vg}$

L'indice des valeurs globales est obtenu en faisant le rapport entre la consommation totale de l'année 2017 sur la consommation totale de l'année 2016.

$$I_{17/16}^{vg} = \frac{\sum p_{17} \times q_{17}}{\sum p_{16} \times q_{16}} = \frac{133700}{105500} \times 100 = 126,73\%$$

La dépense du consommateur a augmenté de 26,73%.

Nous savons que la division de l'indice des valeurs globales par l'indice Laspeyres des prix est un **indice de Paasche de quantité** (de volume).

$$\text{Vérifions cela : } P^q = \frac{I^{vg}}{L^p} \times 100 = \frac{126,73}{111,4} \times 100 = 113,7\%$$

Ce résultat est obtenu également à partir de la formule suivante :

$$P_{17/16}^q = \frac{\sum q_{17} \times p_{17}}{\sum q_{16} \times p_{17}} = \frac{133700}{117500} \times 100 = 1,137 \times 100 = 113,7\%$$

L'interprétation de ces résultats permet de conclure, qu'entre 2016 et 2017 les quantités ont augmenté de 13,7% alors que la dépense du consommateur augmentait de 26,7%, la différence s'expliquant par la hausse des prix de 11,4%.

De la même façon, la division de l'indice des valeurs globale par l'indice Paasche des prix est un **indice de Laspeyres de quantités** (de volume).

$$L^q = \frac{I^{vg}}{P^p} \times 100 = \frac{126,7}{112,2} \times 100 = 112,9\%$$

Ce résultat peut être vérifié à l'aide de l'application directe de la formule de l'indice de Laspeyres de volume.

En effet :

Chapitre 7 : Les indices

$$L_{17/16}^q = \frac{\sum q_{17} \times p_{16}}{\sum q_{16} \times p_{16}} = \frac{119200}{105500} \times 100 = 112,9\%$$

Entre 2016 et 2017 les quantités ont augmenté de 12,9% alors que la dépense du consommateur augmentait de 26,7%, la différence trouve son explication dans la hausse des prix évaluée à 12,2%.

Dans cet exercice, nous relevons la complexité des calculs. Cependant dans la plupart des cas concrets, nous disposons généralement des pondérations déjà calculées par des comptables.

2.2.3. Indices de Fisher

L'indice de Fisher est la **moyenne géométrique** simple des indices de Laspeyres et de Paasche.

$$F_{1/0} = \sqrt{L_{1/0} \times P_{1/0}}$$

Il existe trois indices de Fisher : (un indice des prix, un indice des quantités, un indice des valeurs globales).

Les indices de valeur globale de Laspeyres et Paasche étant égaux, l'indice de Fisher sera ainsi égal à cette valeur commune.

- **Indice Fisher des prix** : $F_{1/0}^p = \sqrt{L_{1/0}^p \times P_{1/0}^p}$
- **Indice Fisher des quantités** : $F_{1/0}^q = \sqrt{L_{1/0}^q \times P_{1/0}^q}$
- **Indice Fisher des valeurs globales** : $F_{1/0}^{vg} = \sqrt{L_{1/0}^{vg} \times P_{1/0}^{vg}}$

Exemple :

Reprenons les données de l'exemple précédent portant sur les prix et les quantités des légumes secs achetés par une famille en 2015 et 2016.

- Calculer l'indice de Fisher des prix, des quantités et des valeurs globales pour l'année 2016 en base 100 en 2015.

Réponse :

$$F_{16/15}^p = \sqrt{L_{16/15}^p \times P_{16/15}^p} = \sqrt{78,7 \times 80,6} = 79,6\%$$

$$F_{16/15}^q = \sqrt{L_{16/15}^q \times P_{16/15}^q} = \sqrt{118,6 \times 121,4} = 119,9\%$$

$$F_{16/15}^{vg} = \sqrt{L_{16/15}^{vg} \times P_{16/15}^{vg}} = \sqrt{95,6 \times 95,6} = 95,6\%$$

2.3. Propriétés :

2.3.1. Réversibilité

- Les indices de Laspeyres et de Paasche ne sont pas réversibles :

$$L_{0/1} \neq \frac{1}{L_{0/1}} \quad P_{0/1} \neq \frac{1}{P_{0/1}}$$

- L'indice de Fisher est réversible : $F_{0/1} = \frac{1}{F_{0/1}}$

2.3.2. Transférabilité (circularité)

Aucun des 3 indices synthétiques n'est transférable.

- $L_{2/0} \neq L_{2/1} \times L_{1/0}$
- $P_{2/0} \neq P_{2/1} \times P_{1/0}$
- $F_{2/0} \neq F_{2/1} \times F_{1/0}$

Chapitre 7 : Les indices

3. Exercices corrigés

Exercice 1 :

Soit un budget de consommation comprenant trois biens dont les prix et les quantités sont respectivement sur une base 100 l'année 1 :

	Bien X		Bien Y		Bien Z	
	Prix	Quantité	Prix	Quantité	Prix	Quantité
Année 1	60	100	90	70	110	50
Année 2	80	120	100	80	130	85
Année 3	100	90	110	85	105	100

- 1- Calculer les indices de prix de Laspeyres pour les années 2 et 3 en base 100 l'année 1
- 2- Calculer les indices de Paasche pour les années 2 et 3 en base 100 l'année 1
- 3- Vérifier les propriétés citées- ci-dessus (Réversibilité et Transférabilité)

Corrigé exercice 1 :

1-Calcul des indices de Laspeyres

$$- L_{2/1}^p = \frac{\sum p_2 q_1}{\sum p_1 q_1} \times 100 = \frac{80 \times 100 + 100 \times 70 + 130 \times 50}{60 \times 100 + 90 \times 70 + 110 \times 50} \times 100 = 120,79\%$$

$$- L_{3/1}^p = \frac{\sum p_3 q_1}{\sum p_1 q_1} \times 100 = \frac{100 \times 100 + 110 \times 70 + 105 \times 50}{60 \times 100 + 90 \times 70 + 110 \times 50} \times 100 = 128,93\%$$

2- Calcul des indices de Paasche

$$- P_{2/1}^p = \frac{\sum p_2 q_2}{\sum p_1 q_2} \times 100 = \frac{80 \times 120 + 100 \times 80 + 130 \times 85}{60 \times 120 + 90 \times 80 + 110 \times 85} \times 100 = 120,63\%$$

$$- P_{3/1}^p = \frac{\sum p_3 q_3}{\sum p_1 q_3} \times 100$$

$$- P_{3/1}^p = \frac{100 \times 90 + 110 \times 85 + 105 \times 100}{60 \times 90 + 90 \times 85 + 110 \times 100} \times 100 = 119,96\%$$

3- Vérification des propriétés

-Si la réversibilité est vérifiée, nous aurons :

$$L_{1/2}^p = \frac{1}{L_{2/1}^p} \quad \text{Or :}$$

$$L_{1/2}^p = \frac{\sum p_1 q_2}{\sum p_2 q_2} \times 100 = \frac{60 \times 120 + 90 \times 80 + 110 \times 85}{80 \times 120 + 100 \times 80 + 130 \times 85} \times 100 = 82,89\%$$

$$L_{2/1}^p = 120,79\% \quad \frac{1}{L_{2/1}^p} = \frac{1}{1,2079} \times 100 = 82,78\%$$

$$L_{1/2}^p \neq \frac{1}{L_{2/1}^p} \quad \text{car} \quad (82,89\% \neq 82,78\%).$$

Par contre, nous constatons que :

$$L_{1/2}^p = \frac{1}{P_{2/1}^p}$$

$$L_{1/2}^p = 82,89\%$$

$$\text{On a : } P_{2/1}^p = \frac{\sum p_2 q_2}{\sum p_1 q_2} = 120,63\% \quad \text{et} \quad \frac{1}{P_{2/1}^p} = \frac{1}{1,2063} \times 100 = 82,89\%$$

$$\text{D'où : } L_{1/2}^p = \frac{1}{P_{2/1}^p}$$

De la même façon, nous pouvons vérifier pour l'indice de Paasche que : $P_{1/2}^p \neq \frac{1}{P_{2/1}^p}$

Conclusion

Les indices Laspeyres et Paasche ne sont pas réversibles.

Chapitre 7 : Les indices

Si la transférabilité est vérifiée, nous aurons :

$$L_{3/1} = L_{3/2} \times L_{2/1}$$

$$L_{3/1} = \frac{\sum p_3 q_1}{\sum p_1 q_1} = 128,93\%$$

$$\begin{aligned} L_{3/2} \times L_{2/1} &= \frac{\sum p_3 q_2}{\sum p_2 q_2} \times \frac{\sum p_2 q_1}{\sum p_1 q_1} \\ &= \frac{100 \times 120 + 100 \times 80 + 105 \times 85}{80 \times 120 + 100 \times 80 + 130 \times 85} \times 120,79 = 125,32\% \end{aligned}$$

Comme $128,93\% \neq 125,32\%$, donc l'indice Laspeyres n'est pas transférable.

Les calculs étant faits, nous pouvons également le vérifier pour l'indice Paasche.

En effet :

$$P_{3/1} \neq P_{3/2} \times P_{2/1} \text{ Puisque :}$$

$$P_{3/1} = 119,96\%$$

$$P_{3/2} \times P_{2/1} = \frac{\sum p_3 q_3}{\sum p_2 q_3} \times \frac{\sum p_2 q_2}{\sum p_1 q_2} = \frac{28850}{28700} \times 120,63 = 121,26\%$$

$$\text{Donc : } P_{3/1} \neq P_{3/2} \times P_{2/1}$$

L'indice Paasche n'est donc pas transférable.

Exercice 2 :

Le tableau ci-dessous indique les prix et les quantités relatifs aux produits A et B pour les années 1990, 1991 et 1992.

Produits	1990		1991		1992	
	Prix	Quantités	Prix	Quantités	Prix	Quantités
A	8,00	23	9,50	37	11,4	18
B	4,69	35	6,20	40	7,75	45

- 1- Calculer les coefficients budgétaires pour les trois années.
- 2- Calculer pour ces produits les indices élémentaires (prix et quantités) de l'année 1992 en considérant l'année de base 1990, puis l'année de base 1991 et ceux de l'année 1991 en considérant l'année 1990 comme base.
- 3- Vérifier la réversibilité des indices élémentaires de prix et quantités de l'année 1991 base 100 de 1990 du produit A.
- 4- Vérifier la transférabilité de l'indice élémentaire de prix du produit A.
- 5- Calculer pour ces produits les indices de Laspeyres, de Paasche (prix et quantités) de l'année 1992 en considérant l'année de base 1990, puis l'année de base 1991, et ceux de l'année 1991 en considérant l'année de base 1990.
- 6- Calculer l'indice de Fisher de prix et de quantités pour l'année 1992 base 1990.
- 7- Vérifier la réversibilité de l'indice de Laspeyres de prix de l'année 1992 base 1991 du produit B.
- 8- Vérifier la transférabilité de l'indice Paasche de quantités du produit B.

Chapitre 7 : Les indices

Corrigé de l'exercice 2 :

1) Calcul des coefficients budgétaires

Les coefficients budgétaires de chaque produit s'obtiennent en calculant le quotient de la valeur (prix \times quantité) de ce produit à une période donnée (courante ou de base) sur la somme des valeurs de tous les produits à la même période (voir le tableau)

Prod	P_{90}	q_{90}	P_{91}	q_{91}	P_{92}	q_{92}	$P_{90} \times q_{90}$	$P_{91} \times q_{91}$	$P_{92} \times q_{92}$	CB90	CB91	CB92
A	8	23	9,5	37	11,4	18	184	351,5	205,2	0,528	0,586	0,371
B	4,69	35	6,2	40	7,75	45	164,15	248	348,75	0,472	0,414	0,629
Total	-	-	-	-	-	-	348,15	599,5	553,95	1	1	1

2) Calcul des indices élémentaires (prix et quantités)

Les indices élémentaires des prix, base 100 en 1990 s'obtiennent en divisant le prix du produit en 1992 par son prix en 1990. De même pour les indices élémentaires des quantités.

➤ Calcul des indices élémentaires des prix produit A

$$I_{92/90}^P = \frac{11,4}{8} \times 100 = 142,5\%$$

$$I_{92/91}^P = \frac{11,4}{9,5} \times 100 = 120\%$$

$$I_{91/90}^P = \frac{9,5}{8} \times 100 = 118,75\%$$

➤ Calcul des indices élémentaires des quantités produit A

$$I_{92/90}^q = \frac{18}{23} \times 100 = 78,26\%$$

$$I_{92/91}^q = \frac{18}{37} \times 100 = 48,65\%$$

$$I_{91/90}^q = \frac{37}{23} \times 100 = 160,87\%$$

➤ Calcul des indices élémentaires des prix produit B

$$I_{92/90}^P = \frac{7,75}{4,69} \times 100 = 165,24\%$$

$$I_{92/91}^P = \frac{7,75}{6,2} \times 100 = 125\%$$

$$I_{91/90}^P = \frac{6,2}{4,69} \times 100 = 92,67\%$$

➤ Calcul des indices élémentaires des quantités produit B

$$I_{92/90}^q = \frac{45}{35} \times 100 = 128,57\%$$

$$I_{92/91}^q = \frac{45}{40} \times 100 = 112,5\%$$

$$I_{91/90}^q = \frac{40}{35} \times 100 = 114,28\%$$

Chapitre 7 : Les indices

3) La réversibilité des indices élémentaires de prix et quantités de l'année 1991 base 100 de 1990 du produit A.

➤ La réversibilité des indices élémentaires de prix

$$I_{90/91}^P = \frac{8}{9,5} \times 100 = 84,21\% \quad ; \quad \frac{1}{I_{91/90}^P} = \frac{1}{118,75} \times 100 = 84,21\%$$

comme $I_{90/91}^P = \frac{1}{I_{91/90}^P}$ donc la condition de la réversibilité est bien vérifiée

➤ La réversibilité des indices élémentaires de quantités

$$I_{90/91}^q = \frac{23}{37} \times 100 = 62,16\% \quad ; \quad \frac{1}{I_{91/90}^q} = \frac{1}{160,87} \times 100 = 62,16\%$$

comme $I_{90/91}^q = \frac{1}{I_{91/90}^q}$ donc la condition de la réversibilité est bien vérifiée

4) La transférabilité des indices élémentaires de prix du produit A.

$$\frac{I_{92/91}^P \times I_{91/90}^P}{100^{2-1}} = \frac{120 \times 118,75}{100^{2-1}} = 142,5\% \quad ; \quad I_{92/90}^P = 142,5\%$$

Comme: $\frac{I_{92/91}^P \times I_{91/90}^P}{100^{2-1}} = I_{92/90}^P$ donc la condition de la transférabilité est bien vérifiée

5) Calcul des indices de Laspeyres et Paasche

On procède aux calculs indiqués dans le tableau ci-dessous

Produit	$P_{90} \times q_{91}$	$P_{90} \times q_{92}$	$P_{91} \times q_{90}$	$P_{91} \times q_{92}$	$P_{92} \times q_{90}$	$P_{92} \times q_{91}$
A	296	144	218,5	171	262,2	421,8
B	187,6	211,05	217	279	271,25	310
Total	483,6	355,05	435,5	450	533,45	731,8

➤ Les indices de Laspeyres des prix

$$L_{92/90}^P = \frac{\sum P_{92} \times q_{90}}{\sum P_{90} \times q_{90}} \times 100 = \frac{533,45}{348,15} \times 100 = 153,224\%$$

$$L_{92/91}^P = \frac{\sum P_{92} \times q_{91}}{\sum P_{91} \times q_{91}} \times 100 = \frac{731,8}{599,5} \times 100 = 122,068\%$$

$$L_{91/90}^P = \frac{\sum P_{91} \times q_{90}}{\sum P_{90} \times q_{90}} \times 100 = \frac{435,5}{348,15} \times 100 = 125,089\%$$

➤ Les indices de Laspeyres des quantités

$$L_{92/90}^q = \frac{\sum q_{92} \times p_{90}}{\sum q_{90} \times p_{90}} \times 100 = \frac{355,05}{348,15} \times 100 = 101,981\%$$

$$L_{92/91}^q = \frac{\sum q_{92} \times p_{91}}{\sum q_{91} \times p_{91}} \times 100 = \frac{450}{599,5} \times 100 = 75,062\%$$

$$L_{91/90}^q = \frac{\sum q_{91} \times p_{90}}{\sum q_{90} \times p_{90}} \times 100 = \frac{483,6}{348,15} \times 100 = 138,905\%$$

Chapitre 7 : Les indices

➤ Les indices de Paasche des prix

$$P_{92/90}^P = \frac{\sum P_{92 \times q_{92}}}{\sum P_{90 \times q_{92}}} \times 100 = \frac{553,95}{355,05} \times 100 = 156,02\%$$

$$P_{92/91}^P = \frac{\sum P_{92 \times q_{92}}}{\sum P_{91 \times q_{92}}} \times 100 = \frac{553,95}{450} \times 100 = 123,1\%$$

$$P_{91/90}^P = \frac{\sum P_{91 \times q_{91}}}{\sum P_{90 \times q_{91}}} \times 100 = \frac{599,5}{483,6} \times 100 = 123,96\%$$

➤ Les indices de Paasche des quantités

$$P_{92/90}^q = \frac{\sum q_{92 \times p_{92}}}{\sum q_{90 \times p_{92}}} \times 100 = \frac{553,95}{533,45} \times 100 = 103,842\%$$

$$P_{92/91}^q = \frac{\sum q_{92 \times p_{92}}}{\sum q_{91 \times p_{92}}} \times 100 = \frac{553,95}{731,8} \times 100 = 75,696\%$$

$$P_{91/90}^q = \frac{\sum q_{91 \times p_{91}}}{\sum q_{90 \times p_{91}}} \times 100 = \frac{599,5}{435,5} \times 100 = 137,657\%$$

6) Les indices de Fisher de prix et de quantités pour l'année 1992 base 1990

➤ Les indices de Fisher de prix pour l'année 1992 base 1990

$$F_{92/90}^P = \sqrt{L_{92/90}^P \times P_{92/90}^P} = \sqrt{153,224\% \times 156,02\%} \times 100 = 154,61\%$$

➤ Les indices de Fisher des quantités pour l'année 1992 base 1990

$$F_{92/90}^q = \sqrt{L_{92/90}^q \times P_{92/90}^q} = \sqrt{101,981\% \times 103,842\%} \times 100 = 102,907\%$$

7) La réversibilité de l'indice de Laspeyres de prix de l'année 1992 base 1991 du produit B

L'indice de Laspeyres est réversible si: $L_{92/91}^P = \frac{1}{L_{91/92}^P}$

$$L_{92/91}^P = 122,068\% \quad ; \quad L_{91/92}^P = \frac{\sum P_{91 \times q_{92}}}{\sum P_{92 \times q_{92}}} \times 100 = \frac{450}{553,95} \times 100 = 81,234\%$$

$$\frac{1}{L_{91/92}^P} = \frac{1}{81,234} \times 100 = 123,101\%$$

comme $L_{92/91}^P \neq \frac{1}{L_{91/92}^P}$ donc l'indice de laspeyres n'est pas réversible

8) La transférabilité de l'indice Paasche de quantités du produit B

L'indice de Paasche est transférable Si: $P_{92/90}^q = \frac{P_{92/91}^q \times P_{91/90}^q}{100^{2-1}}$

$$P_{92/90}^q = 103,842\% \quad ; \quad \frac{P_{92/91}^q \times P_{91/90}^q}{100^{2-1}} = \frac{75,696 \times 137,657}{100^{2-1}} = 104,201\%$$

comme $P_{92/90}^q \neq \frac{P_{92/91}^q \times P_{91/90}^q}{100^{2-1}}$ donc l'indice de Paasche n'est pas transférable

Chapitre 7 : Les indices

Exercice 3 :

Pour trois produits P1, P2, P3, le tableau ci-dessous donne les indices élémentaires $I(x)$ de 2021 par rapport à 2020 et les coefficients budgétaires correspondants :

Produits Pi	Indices élémentaires $I(x_j)_{21/20}$	Coefficients budgétaires	
		CB_{90}^j	CB_{91}^j
P1	110	0,20	0,15
P2	125	0,50	0,30
P3	95	0,30	0,55

Calculer les indices de Laspeyres et de Paasche de 2021 par rapport à 2020 sur l'ensemble de ces trois produits.

Corrigé de l'exercice 3

Définition : Laspeyres est la moyenne arithmétique des indices élémentaires de prix pondéré par les coefficients budgétaires de l'année de base.

➤ **L'indice de Laspeyres**

$$L(x)_{21/20} = \sum_{j=1}^3 CB_{90}^j I(x_j)_{21/20} = 0,2 \times 110 + 0,5 \times 125 + 0,3 \times 95 = 113$$

Définition : Paasche est la moyenne harmonique des indices élémentaires de prix pondéré par les coefficients budgétaires de l'année courante.

➤ **L'indice de Paasche**

$$P(x)_{21/20} = \frac{\sum_{j=1}^n CB_{91}^j}{\sum \frac{CB_{91}^j}{I(x_j)_{21/20}}}$$

$$P(x)_{21/20} = \frac{0,15 + 0,3 + 0,55}{\frac{0,15}{110} + \frac{0,3}{125} + \frac{0,55}{95}} = 104,7$$

Chapitre 8 : Corrélation et régression linéaire

Introduction :

L'étude des relations entre variables constitue une étape essentielle dans de nombreux domaines tels que la santé, l'économie ou l'administration. Elle permet, par exemple, d'analyser comment les investissements influencent les profits d'une entreprise ou de comprendre l'impact des revenus sur les dépenses alimentaires.

Plusieurs questions se posent dans ce contexte :

- Existe-t-il une relation entre ces variables, et si oui, est-elle linéaire ?
- Peut-on exprimer cette relation sous forme d'une équation mathématique ?
- Quelle est l'intensité de cette relation ?
- Peut-on utiliser cette équation pour prédire des valeurs futures ?

Ces problématiques relèvent de l'analyse de la régression, qui se divise en deux grandes catégories :

- **Régression simple** : utilisée pour prédire une variable dépendante à partir d'une seule variable indépendante.
- **Régression multiple** : utilisée pour prédire une variable dépendante en tenant compte de plusieurs variables indépendantes.

1. Régression linéaire simple

1.1 Définition :

La régression linéaire consiste à ajuster une droite à un nuage de points pour résumer, interpréter et prévoir les variations d'une variable dépendante en fonction d'une variable indépendante.

1.2. Détermination de la droite de régression

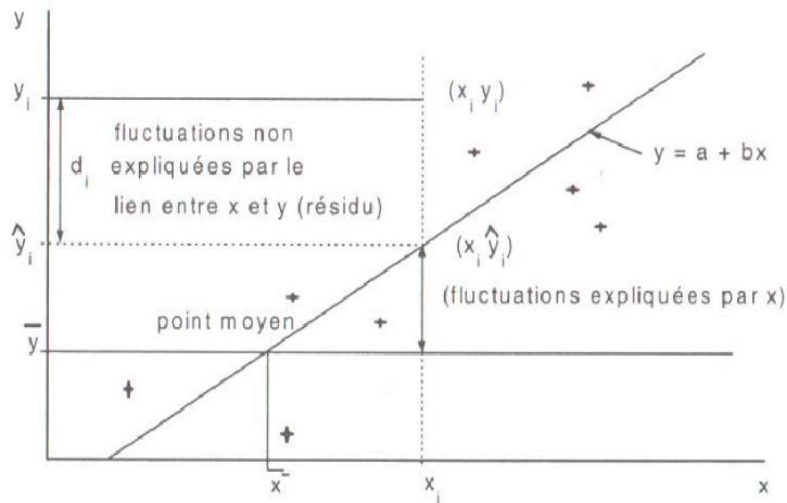
La méthode pour y arriver consiste à considérer n couples de données émanant de l'étude de deux variables statistiques X et Y .

X : Variable indépendante (Variable explicative)	x_1	x_2	x_n
Y : Variable dépendante) (Variable à expliquer)	y_1	y_2	y_n

Le problème consiste à déterminer la droite qui représente le mieux la relation entre X et Y . Cette droite est notée $\hat{y} = a + bx$ (Équation linéaire simple)

La technique utilisée pour déterminer la droite la plus représentative [celle pouvant s'ajuster à un ensemble (x_i, y_i)] est connue sous le nom de **méthodes des moindres carrés**. Le principe de cette méthode est de rendre minimale la somme du carré de la distance verticale (d_i) qui sépare chaque point (x_i, y_i) de la droite elle-même.

Chapitre 8 : Corrélacion et régression linéaire



Ce qui conduit à minimiser G , en calculant les dérivées partielles de l'expression par rapport à (a) et par rapport à (b) . G étant la somme du carré de la distance verticale (d_i).

$$\text{Ce qui entraîne : } G = \sum_{i=1}^n d_i^2 = d_1^2 + d_2^2 + \dots + d_n^2$$

$$\text{Avec : } d_i = y_i - \hat{y}_i = y_i - (a + bx_i) \text{ puisque } \hat{y} = a + bx_i$$

$$\text{La droite cherchée est encore celle qui minimise } G = \sum_{i=1}^n (y_i - a - bx_i)^2$$

Nous appliquons les principes d'optimisation du calcul différentiel, ce qui donne :

$$G' = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \text{ (Par rapport à } a) \dots \dots \dots (1)$$

$$\text{et } G' = -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \text{ (Par rapport à } b) \dots \dots \dots (2)$$

$$\text{A partir de la relation (1), on a : } \sum_{i=1}^n y_i = na + b \sum_{i=1}^n x_i$$

En divisant par n , l'équation fait ressortir une propriété importante de la droite des moindres carrés, qui passe par le point moyen. Ceci permet d'écrire : $\bar{Y} = a + b\bar{X}$

$$\text{Et à partir de la relation (2), on a : } \sum_{i=1}^n x_i y_i = a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2$$

L'utilisation de la méthode des moindres carrés et la simplification des calculs permettent d'obtenir pour les valeurs (a) et (b) de la droite $\hat{Y} = \hat{a} + \hat{b}X$, les équations suivantes :

$$\hat{b} = \frac{\frac{\sum x_i x_i - \bar{X}\bar{Y}}{N}}{\frac{\sum x_i^2 - \bar{X}^2}{N}} = \frac{Cov(X,Y)}{V(X)} \text{ (}\hat{b} \text{ : pente de la droite de régression).}$$

Dans ce cas, nous avons fait intervenir l'expression de la **covariance** entre X et Y notée $cov(X, Y)$ et $\hat{a} = \bar{y} - \hat{b}\bar{x}$ (\hat{a} : constante).

Chapitre 8 : Corrélation et régression linéaire

Exemple :

Dans une étude de marché, la relation de régression intéressante dans une analyse économique est celle qui explique le prix Y_i (variable expliquée et dépendante) en fonction de l'offre X_i (variable explicative et autonome), ce qui permet d'écrire l'équation de la droite $\hat{Y} = \hat{a} + \hat{b}X$ représente la droite retenue.

2. Corrélation linéaire

La notion de nuage statistique ou de diagramme de dispersion n'est rien d'autre que la représentation graphique dans le plan cartésien de l'ensemble de points ou de couples de données (X_i, Y_i) . A partir de cette représentation graphique, il est facile d'avoir une idée sur le degré de liaison entre les deux variables.

2.1. Définition :

Après avoir souligné l'importance de l'ajustement d'une droite de régression, nous abordons maintenant la corrélation, qui mesure le degré de dépendance entre deux variables, en fonction de la dispersion des points autour de la droite de régression.

2.2. Détermination du coefficient de corrélation

Pour le calcul du coefficient de corrélation noté r , nous utilisons la formule la plus simple et la plus pratique qui est la suivante : $r = \frac{Cov(X,Y)}{\sigma(X).\sigma(Y)}$

Avec $\sigma(X)$: représentant l'écart-type de la variable X

$\sigma(Y)$: représentant l'écart-type de la variable Y

La définition du coefficient de corrélation nous permet d'établir que :

1. $-1 \leq r \leq +1$
2. $r < 0$ indique une **corrélation négative**. Ce qui signifie que la relation est négative entre X et Y lorsque X augmente la variable Y diminue.
3. $r > 0$, il y a une **corrélation positive**, ce qui signifie qu'à l'augmentation de la variable X , la variable Y augmente également. Dans ces cas, la relation est relative : si la corrélation linéaire est faible, les points du nuage sont dispersés autour de la droite de régression.
4. Un r proche de -1 ou +1 indique **une forte dépendance entre les variables**. Par exemple, une corrélation positive très forte ($r \approx +1$) signifie que les variations de Y sont largement expliquées par celles de X .
5. $r = \pm 1$ indique une **corrélation maximum**. Cela signifie que la droite de régression s'ajuste parfaitement aux données recueillies. Nous parlerons alors **de liaison fonctionnelle**.

Chapitre 8 : Corrélation et régression linéaire

6. $r = 0$ indique l'absence de corrélation linéaire entre X et Y . Dans ce cas, les droites de régression sont perpendiculaires, et les pentes sont nulles, signifiant que les variations de X n'ont aucun impact sur celles de Y .

Remarque 1 :

Il peut également exister une relation positive entre X et Y , mais celle-ci **n'est pas linéaire**. Dans ce cas, le nuage de points suggère un ajustement curviligne (ajustement sous forme de courbe).

Remarque 2 :

Il est important d'être prudent dans l'interprétation des résultats. Un coefficient de corrélation de 0,99 indique une forte dépendance entre les variables, mais cela ne prouve pas une relation de cause à effet. Ce coefficient suggère une possible relation, que l'économiste ou le statisticien doit ensuite analyser.

La valeur de $100r^2$ représente le pourcentage de la variation totale de Y expliqué par la relation entre Y et X . **Il s'agit du coefficient de détermination (R^2).**

Exemple 1 :

Pour un coefficient de corrélation $r = 0,9$, le coefficient de détermination (R^2) donne $r^2 100 = 81\%$. Ainsi, 81% de la variation de Y se trouve expliquée par le lien entre X et Y .

3. Exercices corrigés

Exercice 1 :

Une entreprise veut mener une étude sur la liaison entre les dépenses (hebdomadaires) mensuelles en publicité et le volume des ventes qu'elle réalise. Nous avons obtenu au cours des six derniers mois les données suivantes :

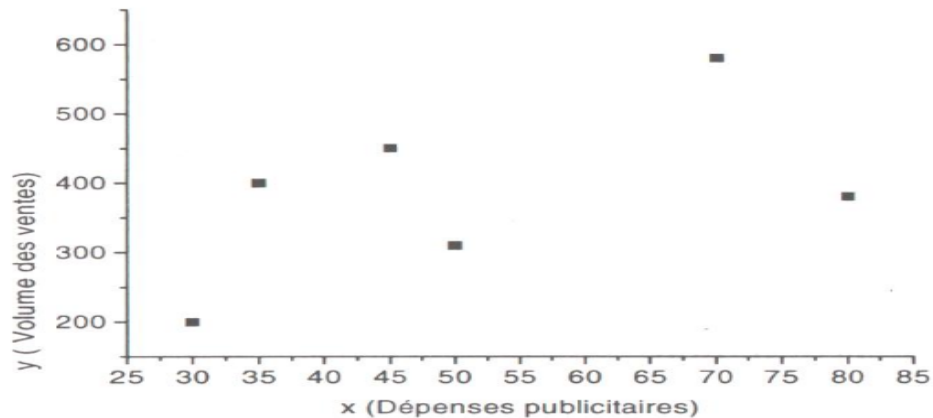
X_i : Dépenses publicitaires (en MDA)	70	80	30	50	35	45
Y_i : Volume des ventes (en MDA)	580	380	200	310	400	450

- 1- Tracer le nuage de points et ajuster la droite de régression.
- 2- Calculer le coefficient de corrélation et interpréter le résultat.

Chapitre 8 : Corrélation et régression linéaire

Corrigé de l'exercice 1:

1. Le nuage de points et ajustement de la droite de régression : $\hat{Y} = \hat{a} + \hat{b}X$



A partir de cette représentation, nous soupçonnons l'existence d'une relation entre X et Y , mais la corrélation n'est pas très forte. Nous le vérifierons avec la question (2).

X_i	Y_i	$X_i Y_i$	X_i^2	Y_i^2
70	580	40600	4900	336400
80	380	30400	6400	144400
30	200	6000	900	40000
50	310	15500	2500	96100
35	400	14000	1225	160000
45	450	20250	2025	202500
310	2320	126750	17950	979400

$$\bar{x} = \frac{\sum x_i}{N} = \frac{310}{6} = 51,6$$

$$\bar{y} = \frac{\sum y_i}{N} = \frac{2320}{6} = 386,66$$

$$V(X) = \frac{\sum x_i^2}{N} - \bar{x}^2 = \frac{17950}{6} - (51,6)^2 = 329,1$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{329,1} = 18,14$$

$$V(Y) = \frac{\sum y_i^2}{N} - \bar{y}^2 = \frac{979400}{6} - (386,66)^2 = 13727,3$$

$$\sigma(Y) = \sqrt{V(Y)} = \sqrt{13727,3} = 117,16$$

$$Cov(X, Y) = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{126750}{6} - 51,6 \times 386,66 = 1173,34$$

Nous obtenons alors pour les valeurs \hat{a} et \hat{b} de la droite $\hat{Y} = \hat{a} + \hat{b}X$, les valeurs suivantes :

$$\hat{b} = \frac{Cov(X, Y)}{V(X)} = \frac{1173,34}{329,1} = 3,56$$

Chapitre 8 : Corrélation et régression linéaire

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = 386,66 - (3,56) \times (51,6) = 387 - 180,96 = 387 - 181 \approx 203$$

$$\text{D'où : } \hat{Y} = 3,56X + 203 \quad \text{On a : } \sigma(X) \times \sigma(Y) = 2125,28$$

2. Calcul et interprétation du coefficient de corrélation :

$$r = \frac{\text{Cov}(X, Y)}{\sigma(X) \times \sigma(Y)} = \frac{1173,34}{18,14 \times 117,16} = \frac{1173,34}{2125,28} = 0,55$$

Cette série montre une corrélation positive modérée entre le volume des ventes et les dépenses en publicité, avec 30,25 % de la fluctuation totale de Y expliquée par la relation entre X et Y. ($r^2 \times 100 = 0,55^2 \times 100 = 30,25\%$).

Exercice 2 :

Le tableau suivant donne la distance de freinage d'un véhicule automobile roulant sur route sèche, en fonction de sa vitesse.

Vitesse en (Km/h)	40	50	60	70	80	90	100	110	120
Distance en m	8	12	18	24	32	40	48	58	72

- 1- Calculer la $V(X)$, $V(Y)$ et la $\text{cov}(X, Y)$.
- 2- Calculer l'équation de la droite de régression permettant d'estimer la distance de freinage en fonction de la vitesse du véhicule ($\hat{Y} = \hat{a} + \hat{b}X$).
- 3- Tracer sur un graphique les points (x_i, y_i) et la droite d'ajustement.
- 4- Calculer le coefficient de corrélation linéaire et interpréter le résultat.
- 5- Estimer, à l'aide de cette équation, la distance de freinage d'un véhicule roulant à 130 km/h.

Corrigé exercice 2 :

Tableau des calculs intermédiaires :

Vitesses en Km/h x_i	Distance en m y_i	x_i^2	y_i^2	$x_i \times y_i$
40	8	1600	64	320
50	12	2500	144	600
60	18	3600	324	1680
70	24	4900	576	1680
80	32	6400	1024	2560
90	40	8100	1600	3600
100	48	10000	2304	4800
110	58	12100	3364	6380
120	72	14400	5184	8640
Total	720	63600	14584	29660

Chapitre 8 : Corrélacion et régression linéaire

1. Calcul de : $V(X)$, $V(Y)$ et la $cov(X, Y)$.

On calcul d'abord les moyennes :

$$\bar{X} = \frac{\sum x_i}{N} = \frac{720}{9} = 80$$

$$\bar{Y} = \frac{\sum y_i}{N} = \frac{312}{9} = 34,67$$

$$V(X) = \frac{\sum x_i^2}{N} - \bar{X}^2 = \frac{63600}{9} - (80)^2 = 666,67$$

$$\sigma(X) = \sqrt{V(X)} = \sqrt{666,67} = 25,82$$

$$V(Y) = \frac{\sum y_i^2}{N} - \bar{Y}^2 = \frac{14584}{9} - (34,67)^2 = 418,43$$

$$\sigma(Y) = \sqrt{V(Y)} = \sqrt{418,43} = 20,46$$

$$cov(X, Y) = \frac{\sum x_i y_i}{N} - \bar{X} \bar{Y} = \frac{29660}{9} - (80) \times (34,67) = 521,96$$

2. Calcul de l'équation de la droite de régression $\hat{Y} = \hat{a} + \hat{b}X$

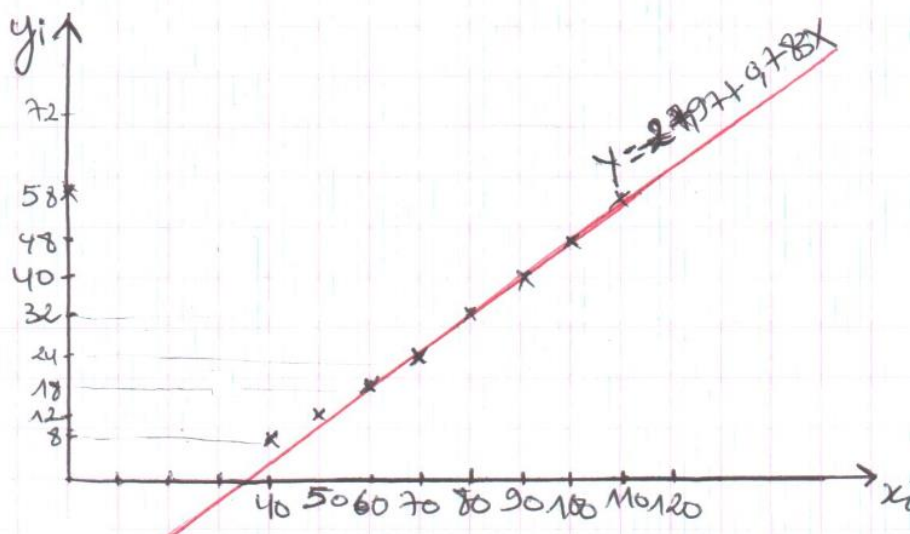
Selon la méthode des moindres carrés, nous avons :

$$\hat{b} = \frac{COV(X, Y)}{V(X)} = \frac{521,96}{666,67} = 0,783$$

$$\hat{a} = \bar{Y} - \hat{b} \bar{X} \rightarrow \hat{a} = 34,6 - 0,783 \times 80 = -27,97$$

La droite de régression s'écrit alors : $\hat{Y} = -27,97 + 0,783X$

3. Le graphique des points (x_i, y_i) et la droite d'ajustement



4. Calcul du coefficient de corrélation linéaire r :

$$r = \frac{Cov(X, Y)}{\sigma(X) \cdot \sigma(Y)}$$

$$r = \frac{521,96}{25,82 \times 20,46} = 0,988$$

Comme ($r = 0,988$) est très proche de 1, on dit qu'il existe une corrélation positive très forte entre les deux variables (vitesse et distance de freinage d'un véhicule).

5. Estimation de la distance de freinage d'un véhicule roulant à 130 km/h

Elle peut être estimée par l'équation : $\hat{Y} = -27,97 + 0,783 \times (130) = 74,21m$

Conclusion

Ce polycopié de Statistique I a été conçu comme un outil pédagogique structuré, clair et accessible. Il s'adresse principalement aux étudiants de première année en sciences économiques, commerciales et de gestion, ainsi qu'à toute personne souhaitant acquérir une maîtrise solide des bases de la statistique descriptive. Son objectif est d'offrir aux apprenants les outils nécessaires pour comprendre, analyser et appliquer efficacement les concepts fondamentaux de cette discipline, tout en restant aligné sur les exigences du programme officiel.

En adoptant une démarche progressive, ce document débute par une présentation des concepts fondamentaux et du vocabulaire statistique. Cette base théorique permet de structurer une approche analytique rigoureuse. Par la suite, il explore les méthodes essentielles pour organiser et présenter les données, telles que l'utilisation de tableaux, de graphiques et de diagrammes. Les notions clés, comme les mesures de tendance centrale (moyenne, médiane, mode) et les indicateurs de dispersion (étendue, variance, écart-type), y sont développées de manière détaillée afin d'assurer une interprétation claire et synthétique des données.

Ce polycopié ne se limite pas à une présentation théorique des notions. Chaque chapitre est enrichi d'exemples concrets, illustrant l'application des concepts dans des situations pratiques. En complément, des outils plus avancés sont introduits, notamment les paramètres de forme pour analyser les asymétries des distributions, ainsi que les mesures de concentration, telles que la courbe de Lorenz et les indices associés. Ces éléments permettent d'apporter une lecture approfondie des phénomènes économiques et sociaux. Les derniers chapitres abordent la corrélation et la régression linéaire, ouvrant la voie à l'étude des relations entre variables et à l'élaboration de prévisions.

Afin de consolider l'apprentissage, le document propose une série d'exercices pratiques accompagnés de solutions détaillées. Ces exercices représentent une opportunité pour les étudiants d'évaluer leur compréhension et de s'entraîner dans un cadre académique, tout en favorisant l'acquisition de compétences applicables à des contextes réels.

Pour conclure, ce polycopié vise à cultiver chez les étudiants une aptitude au raisonnement logique et analytique, tout en rendant l'apprentissage de la statistique à la fois accessible et captivant. Nous espérons qu'il permettra de lever les appréhensions autour de cette discipline fondamentale et encouragera les apprenants à approfondir leur maîtrise des méthodes statistiques, afin de mieux analyser et interpréter les phénomènes complexes qui les entourent.

Références bibliographiques :

- Ayache, A., & Hamonier, J. (2014). *Cours : Statistique descriptive et calcul de probabilités*. Université de Lille. Disponible sur : http://math.univ-lille1.fr/cours_SDCP_Janv19.
- Baccini, A. (2010). *Statistique descriptive élémentaire*. Publications de l'Institut de Mathématiques de Toulouse, UMR CNRS.
- Baccus, A. (2011). *Statistique descriptive : Méthodes et exercices corrigés*. Dunod.
- Bernard, P.-Y. (1994). *Exercices corrigés de statistique descriptive* (2e éd.). Economica.
- Bernard, P.-Y. (1996). *Statistique descriptive : Nouvelle méthode pour bien comprendre et réussir* (4e éd.). Economica.
- Bru, J.-P., Guégan, D., & Haccoun, A. (2016). *Statistique : Méthodologie et applications* (3e éd.). Dunod.
- Caille, J.-P., & Déaux, O. (2010). *Statistique descriptive et probabilités*. Ellipses.
- Chauvat, G., & Réau, J.-P. (2008). *Statistiques descriptives* (5e éd.). Armand Colin.
- Chaumont, L. (2010). *Statistique descriptive et prévision*. Cours.
- Dakhmouche, M. (2010). *Introduction à la statistique descriptive*. École Préparatoire en Sciences Économiques, Commerciales et des Sciences de Gestion de Constantine. Disponible sur : <https://www.cours-gratuit.com/manuel>.
- Dreesbeke, J.-J., Fine, J., & Saporta, G. (2007). *Précis de statistique*. Technip.
- Dubos, J. (1984). *Statistique descriptive en science économique* (2e éd.). Dunod.
- Duthil, G., & Vanhaecke, D. (1998). *Initiation à la statistique descriptive*. Ellipses.
- Freedman, D. A., Pisani, R., & Purves, R. A. (2007). *Statistics* (4th ed.). W. W. Norton & Company.
- Grais, B. (1998). *Exercices corrigés : Statistique descriptive avec rappels de cours*. Dunod.
- Hamdani, H. (2001). *Statistique descriptive avec initiation aux méthodes d'analyse de l'information économique*. OPU.
- Hubler, J. (2011). *Statistique descriptive appliquée à la gestion et à l'économie* (3e éd.). Bréal.
- Leboucher, L., & Voisin, M.-J. (2015). *Introduction à la statistique descriptive* (3e éd.). Cepaduès.
- Lethielleux, M. (2003). *Statistique descriptive*. Éditions Dunod, collection « Express ».
- Malika Boukella-Bouzouane. (2001). *Statistique descriptive : Rappels de cours avec exercices corrigés*. Casbah.
- Monino, J.-L., Kosianski, J.-M., & Cornu, F. (2010). *Statistique descriptive* (4e éd.). Dunod.
- Moore, D. S., McCabe, G. P., Alwan, L. C., & Craig, B. A. (2016). *Introduction to the Practice of Statistics* (9th ed.). W. H. Freeman.
- Olivier, E. (2008). *L'essentiel de statistique descriptive*. Gualino. ISBN : 978-2-297-01103-7.
- Piller, A. (2004). *Statistique descriptive*. Éditions Premium.
- Rabaud, M. (2018). *Statistique descriptive : Bases et applications*. De Boeck Supérieur.
- Saporta, G. (2011). *Probabilités, analyse des données et statistique* (2e éd.). Technip.
- Weiss, N. A. (2020). *Introductory Statistics* (11th ed.). Pearson.
- حسني، محمد عبد الغني. (2016). *مبادئ الإحصاء الوصفي والاستدلالي*. مكتبة الأنجلو المصرية.
- عيسى، عبد الكريم أحمد. (2008). *مدخل إلى الإحصاء الوصفي وتطبيقاته*. دار الحامد للنشر والتوزيع.

Notations et abréviations

Les abréviations et notations employées tout au long de ce polycopié sont détaillées ci-après :

n_i	Effectif ou fréquence absolue
n_{ic}	Effectif corrigé (rectifié)
$n_i \uparrow$	Effectifs cumulés croissants
$n_i \downarrow$	Effectifs cumulés décroissants
f_i	Fréquence relative
$f_i \uparrow$	Fréquences relatives cumulées croissantes
$f_i \downarrow$	Fréquences relatives cumulées décroissantes
a_i	Amplitude
c_i	Centre de classe
N	Effectif total
X	Variable statistique
x_i	$i^{\text{ème}}$ modalité
M_e	Médiane
M_o	Mode
\bar{X}	Moyenne arithmétique d'une série X
G	Moyenne géométrique
H	Moyenne harmonique
Q	Moyenne quadratique
m_k	Moment simple
μ_k	Moment centré
$V(X)$	Variance de X
$\sigma(X)$	Ecart-type de X
CV	Coefficient de variation
T_x	Taux de croissance moyen
I_p	Indice simple des prix
I_q	Indice simple des quantités
L_p	Indice synthétique de Laspeyres des prix
L_q	Indice synthétique de Laspeyres des quantités
P_p	Indice synthétique de Paasche des prix
P_q	Indice synthétique de Paasche des quantités
r	Coefficient de corrélation
R^2	Coefficient de détermination

Table des matières

Avant-propos	
Introduction	1
Chapitre 1 : Statistique Descriptive : Vocabulaire et Concepts de Base	2
Introduction	2
1. La collecte de l'information	2
1.1. La méthode exhaustive (recensement)	2
1.2. La méthode des sondages (échantillonnage)	3
1.2.1. Échantillonnage aléatoire	3
1.2.2. Échantillonnages empiriques	6
2. Les sources statistiques	6
3. Différence entre la statistique et les statistiques	6
3.1. La statistique	6
3.2. Les statistiques	6
4. Les concepts de base de la statistique descriptive	6
4.1. La population statistique	6
4.2. L'échantillon	6
4.3. L'individu (unité statistique)	7
4.4. Le caractère ou la variable	7
4.5. Les modalités	7
4.6. La série statistique	8
4.7. Récapitulatif des concepts de base de la statistique descriptive	8
5. Exercices corrigés	9
Chapitre 2 : Représentation de Données Statistiques	12
Introduction	12
1. Présentation de données sous forme de tableaux	12
1.1. Présentation des données d'un caractère qualitatif	12
1.2. Présentation des données d'un caractère quantitatif	13
2. Représentation graphique des séries statistiques	15
2.1. Cas d'un caractère qualitatif	15
2.2. Cas d'un caractère quantitatif	16
3. Exercices corrigés	19
Chapitre 3 : Les Paramètres de Tendance Centrale	25
Introduction	25
1. La Médiane	25
1.1. Définition	25
1.2. Déterminations pratique	25
1.2.1. Cas d'une variable discontinue	25
1.2.2. Cas d'une variable continue	27
2. Le Mode	27
2.1. Définition	27
2.2. Détermination pratique	27
2.2.1. Cas d'une variable discrète	27
2.2.2. Cas d'une variable continue	28
3. La Moyenne	29
3.1. Définition	29
3.2. Détermination pratique	29
3.2.1. Cas d'une variable discontinue	29
3.2.2. Cas d'une variable continue	29
3.3. La moyenne géométrique	30

3.3.1. Moyenne géométrique simple.....	30
3.3.2. Moyenne géométrique pondérée.....	31
3.4. La moyenne harmonique.....	31
3.4.1. Moyenne harmonique simple.....	31
3.4.2. Moyenne harmonique pondérée.....	31
3.5. La moyenne quadratique.....	32
3.5.1. La moyenne quadratique simple.....	32
3.5.2. La moyenne quadratique pondérée.....	32
4. Les caractéristiques de position.....	32
4.1. Définition.....	32
4.2. Principales caractéristiques de position.....	33
4.2.1. Les Quartile.....	33
4.2.2. Les Déciles.....	33
4.2.3. Les centiles.....	33
5. Exercices corrigés.....	33
Chapitre 4 : Les Paramètres de Dispersion.....	37
Introduction.....	37
1. Différences entre deux valeurs d'une série.....	37
1.1. L'étendue.....	37
1.1.1. Définition.....	37
1.1.2. Détermination pratique.....	37
2.2. Écart interquartile.....	39
2.3. Écart interdécile.....	39
2.4. Écart intercentile.....	39
3. Valeurs centrales de variables centrées.....	39
3.1. Ecart absolu moyen.....	39
3.1.1. Définition.....	39
3.1.2. Détermination pratique.....	39
3.2. La variance et l'écart-type (ou standard-déviaton).....	40
3.2.1. La variance.....	40
3.2.2. L'écart type.....	41
4. Coefficient de dispersion relative : coefficient de variation.....	42
5. Exercices corrigés.....	43
Chapitre 5 : Les Paramètres de Forme.....	46
Introduction.....	46
1. Les moments simples et les moments centrés.....	46
1.1. Moments centrés.....	46
1.2. Moments simples (Moments par rapport à l'origine).....	46
2. Asymétrie.....	47
2.1. Principaux coefficients d'asymétrie.....	47
2.1.1. Coefficient de Fisher.....	47
2.1.2. Coefficient de Yule.....	47
2.1.3. Coefficient de Pearson.....	48
2.2. Utilisation des coefficients.....	48
3. Aplatissement.....	48
3.1. Principaux coefficients d'aplatissement.....	49
3.1.1. Coefficient de Pearson.....	49
3.1.2. Coefficient de Fisher.....	49
4. Exercices corrigés.....	50

Chapitre 6 : Les Paramètres de Concentration	53
Introduction.....	53
1. Détermination de la concentration par le calcul.....	53
1.1. Détermination de la médiane.....	53
1.2. Détermination de la médiale.....	53
1.3. L'écart (médiale – médiane)	54
1.4. Comparaison de ΔM à l'Étendue.....	54
1.5. Le coefficient de concentration.....	54
2. Détermination de la concentration par le graphe.....	55
2.1. La courbe de concentration (ou de LORENZ)	55
2.2. L'indice de Gini.....	55
3. Exercice corrigé.....	57
Chapitre 7 : Les Indices	60
Introduction.....	60
1. Les indices élémentaires ou simples.....	60
1.2. Propriétés.....	60
1.2.1. Réversibilité.....	60
1.2.2. Transférabilité (circularité)	60
2. Les indices synthétiques.....	61
2.1. Les indices synthétiques simples.....	61
2.2. Les indices synthétiques pondérés.....	62
2.2.1. Indices de Laspeyres.....	62
2.2.2. Indices de Paasche.....	62
2.2.3. Indices de Fisher.....	66
3. Exercices corrigés.....	67
Chapitre 8 : Corrélation et Régression Linéaire	73
Introduction.....	73
1. Régression linéaire simple.....	73
1.1. Définition.....	73
1.2. Détermination de la droite de régression.....	73
2. Corrélation linéaire.....	75
2.1. Définition.....	75
2.2. Détermination du coefficient de corrélation.....	75
3. Exercices corrigés.....	76
Conclusion	80
Références bibliographiques	81
Notations et abréviations	82
Tables des matières	83