Université Mouloud Mammeri de Tizi-Ouzou

Faculté des Sciences Economique, Commerciale et des Sciences de Gestion

Département des Sciences Commerciales

POLYCOPIE

COURS DE STATISTIQUE I

Licence 1 (Socle commun) - Semestre 1

Réalisé par Dr. BABOU OMAR

Maitre de conférences B

Année universitaire 2025/2026

Université Mouloud Mammeri de Tizi-Ouzou

Faculté des Sciences Economique, Commerciale et des Sciences de Gestion

Département des Sciences Commerciales

POLYCOPIE

COURS DE STATISTIQUE I

Licence 1 (Socle commun) - Semestre 1

Réalisé par Dr. BABOU OMAR

Maitre de conférences B

Année universitaire 2025/2026

Syllabus

Intitulé de la matière : Statistique I

Domaine : Sciences Economique, Commerciale et des Sciences de Gestion

Niveau : 1ère année Licence

Filière : Toutes les filières Socle commun

Niveau : Licence 1 Semestre : 1

Année universitaire 2025/2026

Identification de la matiére

Unité d'enseignement : Méthodologique

Crédits : 5 ; Coefficient : 3

Module annuel dispensé en deux semestres et sous forme de cours et travaux dirigés Stat I et Stat II

Volume horaire hebdomadaire : 3 heures

Cours magistral : 1h30 - Travaux dirigés : 1h30

Responsable de la matière

Nom et prénom : BABOU Omar (chargé de cours et TD de statistique depuis 2017)

Grade : Maître de conférences B

Affiliation : Département des sciences commerciales

Courriels : o_babou@yahoo.fr ou omar.babou@ummto.dz

Téléphone : 0556120549

Description de la matière

Connaissances préalables recommandées

Connaissances et notions élémentaires en mathématiques, notamment les opérations et les règles dispensées en cycle de seconde, voire de moyen.

Objet de l'enseignement

L'objet de la statistique descriptive est de savoir collecter les données, concernant un phénomène social ou économique donné à étudier, les ordonner, les classifier, les résumer et les présenter, de façon claire et lisible, sous forme de tableaux et/ou graphes afin de pouvoir en calculer des paramètres divers permettant de décrire et d'analyser les variations du phénomène en question.

Objectifs de l'enseignement

Compte tenu de son objet, le présent cours, conçu conformément au programme ministériel (Arrêté n° 808 du 22 juillet 2022), vise à ce que l'étudiant acquiert des compétences en la matière, à savoir : - maitriser les concepts clés de la statistique descriptive, - résumer et présenter des données sous forme de tableaux et graphes, - calculer et analyser les différents paramètres statistiques (de tendance central, de dispersion, de forme et de concentration), - analyser et quantifier la relation entre deux variables et mesurer leur corrélation, - calculer les indices de la vie économique et la compréhension de leur signification et leur utilité et leur usage, - s'initier à l'usage des logiciels statistiques utilitaires, notamment le logiciel de base Excel, pour les graphiques et le calcul des paramètres.

Les chapitres dispensés en cours magistral font l'objet chacun d'un traitement appliqué, sous forme de séries d'exercices, en séances de travaux dirigés. Lors de ces dernières des éclaircissements et des informations supplémentaires, surtout d'ordre pratique, sont fournis aux étudiants. Une bibliographie révisée et mise à jour est fournie en annexe, dans le but de permettre aux étudiants d'approfondir leurs connaissances. Les chapitres du présent cours sont également publiés en ligne, sur la plateforme Moodle (E learning).

<u>Mode d'évaluation</u>

Noté sur 20 : contrôle continu (40%) et examen (60%). Moyenne du module = (Note Examen x 0,6) + (Note TD X 0,4)

Plan du cours

INTRODUCATION GENERALE

Dans notre vie quotidienne nous somme souvent confronter aux informations chiffrées (nombre d'habitants d'un espace géographique, nombre d'admis au baccalauréat, taux de chômage, la répartition des hommes et des femmes selon l'âge dans une région, etc.) qui nous parviennent de plusieurs sources d'informations. Ces ***statistiques*** qu'il faut différencier de ***la statistique*** (intitulé de cette matière d'étude) sont justement les résultats de l'application des méthodes de la statistique.

Ainsi, la statistique est une méthode scientifique, qui nous permet de recueillir, de résumer, de classer et ranger des données ou des informations statistiques qualitatives ou quantitatives, en vue de les analyser et de prendre les décisions judicieuses.

Étymologiquement le mot statistique vient : « De l'allemand Staatskunde, dérivé de l'italien statista (homme d'État, statiste), la statistique représentant l'ensemble des connaissances que doit posséder un homme d'État. » (1785).

Recensements en Chine au XXIIIe siècle av. J.-C. ou en Égypte au XVIIIe av. J.-C, système de recueil se poursuivant jusqu'au XVIIe. Rôle prévisionnel des statistiques au XVIIIe siècle avec la construction des premières tables de mortalité avec Antoine Deparcieux, l'Essai sur les probabilités de la durée de vie humaine (1746). Rôle démographique au XIXe siècle, le Baron de Reiffenberg présentait en 1842 à l'Académie ses calculs rétrospectifs de population chez des peuples gaulois, d'après des chiffres donnés par Jules César dans sa conquête des gaules. (Verlant, B. Saint-Pierre, G. ; 2008)

Au jour d'aujourd'hui, nous distinguons deux types de statistiques : la statistique descriptive qui a pour objet de recueillir des observations portant sur des sujets présentant une certaine propriété et de traduire ces observations par des nombres qui permettent d'avoir des renseignements sur cette propriété. Le but de la statistique descriptive est de structurer et de représenter l'information contenue dans les données. Par contre statistiques inférentielle (probabilités) permet de confirmer ou d'infirmer une hypothèse avec une marge d'erreur la plus petite possible et/ou prédire un événement à l'aide d'outils

Au cours de ce premier semestre d'études, destiné aux étudiants de première année de la faculté des sciences économiques, commerciales et des sciences de gestion, nous nous intéressons à la statistique descriptive.

Par ailleurs, il y a lieu de souligner que ce cours est enseigné dans les différents domaines à l'université ; des sciences humaines et sociales aux sciences de la nature et de la vie en passant par les sciences de l'ingénieur.

CHAPITRE I.
GENERALITES SUR LA STATISTIQUE

L'objet de ce chapitre est d'initier l'étudiant aux techniques de collectes de l'information statistique ainsi qu'a la maitrise du vocabulaire usuelle en statistique.

Section 1. Collecte de données :

La collecte de données est une phase primordiale d'une étude empirique ou d'un travail de recherche durant laquelle le chercheur ou statisticien récolte des informations qui seront analysées pour confirmer (ou non) des hypothèses de départ, et répondre à une problématique. (L'entreprise fabrique-t-elle un produit ou un composant dans les normes ? quels sont les déterminants de l'emploi ou du chômage en Algerie ? s'agit-il du niveau d'étude, de l'état matrimonial, du sexe, etc.). Cette brève section permet à l'étudiant d'avoir un aperçu sur collecte de données.

1.1. Collecte de données : les étapes primordiales

La collecte de données peut s'effectuer à l'aide de plusieurs techniques et aide le statisticien à comprendre le phénomène, le fait, ou le sujet qu'il étudie.

C'est justement la pertinence des données récoltées qui détermine leur bonne interprétation lors de l'analyse puis de la conclusion et qui apportera des réponses adéquates à la problématique de recherche malgré la complexité du phénomène étudié

Prenons l'exemple de la taille des enfants de 7 ans résidant en Algérie. Il semble difficile matériellement de recenser les valeurs prises par la variable aléatoire « taille » dans cette population. On peut au mieux définir un « échantillon représentatif » et tirer des conclusions après avoir analysé les mesures.

Il est clair que la conclusion va dépendre fortement de la façon dont on a formé l'échantillon. Comment en définir les modalités : 1ère lettre du nom, jour de naissance, ? Doit-on effectuer les mesures ou peut-on se contenter de celles inscrites sur le carnet de santé ?

Ce sont ces questions que le statisticien doit se poser avant de réaliser son étude

Des questions similaires se posent dans le domaine industriel lorsque l'on cherche à contrôler un processus de fabrication. Il est en effet souvent impossible de contrôler toutes les unités (individus) d'un lot de fabrication (qui peut être considéré comme une population). L'échantillonnage passe alors par de multiples questions : à quel moment doit-on prélever l'échantillon (début, fin de fabrication), quelle taille d'échantillon, quelle fréquence de prélèvement ? Qui doit prélever ? et comment ?

1.2.<u>Techniques de collecte des données</u>

L'entretien directif, semi-directif ou libre : les différents types d'entretiens permettent de récolter des informations à travers une discussion avec une ou plusieurs personnes.

La technique de l'entretien est très utile dans la collecte de données informatives sur des sujets très précis. Il s'agit surtout de s'entretenir avec des personnes qui ont une expertise ou une expérience particulière sur votre sujet.

L'observation : que l'observation soit participante, non participante, structurée ou non structurée, elle permet de collecter des données intéressantes.

La technique de l'observation s'avère utile quand le chercheur étudie un phénomène, un sujet, réel, qui est observable.

Le questionnaire : Le chercheur peut avoir recours au questionnaire s'il veut interroger un grand nombre de personnes. Avec un échantillon représentatif de la population ciblée par son sujet, le chercheur peut collecter des données statistiques qui lui fourniront des informations après analyse.

Ces informations doivent l'aider à dresser un ou plusieurs constats qui répondent à sa problématique et ses hypothèses de départ.

Le sondage : Le sondage permet au statisticien de poser une question générale à un échantillon dit représentatif.

En réalité, dans la plupart des cas, il est difficile d'obtenir l'information à partir de la population dans son ensemble. On utilise alors un échantillon pour tirer des conclusions sur la population.

Cette technique est utile pour questionner une population autour d'une grande question. La collecte d'une opinion d'individus autour d'une question centrale peut permettre d'expliquer une situation, un phénomène, ou un fait. Dans une démarche statistique, on définit généralement un plan d'échantillonnage qui permet de préciser la façon dont l'échantillon a été obtenu. Un plan couramment utilisé est le plan d'Echantillonnage Aléatoire Simple et Indépendant (plan EASI) où tous les échantillons possibles ont une même chance d'être tirés et sont indépendants les uns des autres.

<u>Section 2. Le vocabulaire statistique</u>

Nous allons nous intéresser dans ce qui suit aux notions de bases de la statistique ; notamment, la population et l'unité statistique, le caractère, la nature et les modalités de caractère. Ce vocabulaire vient à l'origine des travaux sur la démographie pour embrasser tous les champs de la statistique (Leboucher & Voisin, 2011).

2.1. Population et unité statistique

Une population statistique est l'ensemble des individus ou des éléments sur lesquels on effectue des observations (étude statistique).

Exemples : les étudiants d'une faculté, les entreprises d'une région, les arbustes d'une pépinière.

Taille d'une population : La taille d'une population est le nombre d'individus qui la composent.

Individu (ou unité statistique) : Chaque individu ou élément de la population statistique étudiée (Baccini, 2010).

Par exemple si l'on veut faire une étude chiffrée sur l'ensemble des étudiants composant un amphithéâtre, la population statistique étudiée est cet ensemble d'étudiants, alors que chaque étudiant composant cette population représente l'unité ou l'individu statistique.

En statistique, on s'intéresse aux populations de grande taille. Cependant, comme nous l'avons expliqué plus haut ; quand une population est de très grande taille, on peut être amené à restreindre l'étude à une partie seulement de la population : il s'agit d'un échantillon de cette population.

2.2. **Le Caractère étudié : définition, nature et modalités**

Le caractère statistique : C'est ce qui est observé ou mesuré sur les individus d'une population statistique. Autrement dit c'est la caractéristique étudiée par le statisticien. Par exemple si l'étude porte sur les étudiants d'une faculté on peut s'intéresser à plusieurs caractéristiques pour chaque individu de cette population : âge, taille, commune de résidence, filière du bac, moyenne au bac, nombre de frères et sœurs, etc.

Par ailleurs, on distingue deux types de caractères :

- Qualitatif
- Quantitatif

Caractère qualitatif : Un caractère statistique est qualitatif si ses valeurs, ou modalités, s'expriment de façon littérale ou par un codage sur lequel les opérations arithmétiques telles que moyenne, somme, … n'ont pas de sens. En effet, les opérations arithmétiques ne sont pas possibles sur les modalités d'un caractère qualitatif, et débouchent sur des résultats irrationnels et vides de sens (Py, 1996).

Nous distinguons deux types de caractères qualitatifs

Caractère qualitatif nominal : C'est une variable qualitative dont les modalités ne sont pas ordonnées. Si par exemple on s'intéresse à la répartition des étudiants selon la commune de résidence, à priori l'ordre importe peu.

Caractère qualitatif ordinal : C'est une variable qualitative dont les modalités sont naturellement ordonnées. A titre d'exemple si on s'intéresse à la satisfaction des clients par rapport à un produit on peut ordonner cette satisfaction comme suit : très satisfait, satisfait, neutre, insatisfait et très insatisfait.

Caractère ou variable quantitatif : une variable est dite quantitative si ses différentes valeurs sont mesurables. Exemple si la question est de savoir l'âge d'un individu ou sa taille la réponse est certainement chiffrée.

De plus, une variable statistique est quantitative si ses valeurs sont des nombres exprimant une quantité, sur lesquels les opérations arithmétiques (somme, moyenne, produit, etc...) ont un sens.

Nous distinguons deux types de variables ou de caractère quantitatif

Une variable quantitative est dite discrète **(VSD)** si l'étendue des valeurs possibles est dénombrable, c'est-à-dire si les valeurs peuvent être énumérées sous la forme d'une liste de chiffre (0, 1, 2, 3…). Autrement dit, une variable statistique est discrète lorsqu'elle ne peut prendre que certaines valeurs isolées dans son intervalle de variable. Et comme le précise Grais (2000) ces valeurs sont en général des nombres entiers.

Exemple : Nombre d'objets dans un dépôt, nombre de mots dans une phrase, nombre de raisins sur une grappe…

Une Variable statistique continue : Une variable quantitative est dite continue si les valeurs possibles ne sont pas dénombrables.

Exemple : poids d'un produit, âge d'un groupe d'individus, taille des arbustes, distance au lieu du travail…

En pratique les variables statistiques continues sont souvent regrouper en classes. La classe désigne un intervalle borné de modalités susceptibles d'être ne prises par les effectifs « ni » parmi les « N » individus de la population étudiée. Cet intervalle est délimité par des extrémités qu'on appelle « bornes » ou « limites » de classes selon le cas.

Par convention, en statistique, on considère des intervalles semi-ouverts ou ouverts à droite, afin d'éviter le chevauchement des classes. La distance entre les deux extrémités de l'intervalle, mesurée par la différence « borne supérieure moins la borne inférieure », s'appelle ***l'amplitude de classe et notée « ai »***. De plus, par convention, chaque classe est résumée ou représentée par son centre qu'on appelle aussi « centre de classe », noté « xi » tel que ce centre de classe est la moyenne des deux extrémités de la classe (Boukella-Bouziane, M., 2001)

Modalité du caractère :

Ce sont toutes les possibilités ou catégories que peut prendre un caractère. A titre indicatifs le caractère sexe possède deux modalités ; féminin et masculin de même que le caractère " état matrimoniale" possède quatre modalités : célibataire, marié, divorcé, veuf ou veuve.

Il convient de préciser que chaque individu ne peut être associer qu'a une seule modalité. Des cas particuliers existent en statistique (question à choix multiples) mais ne font pas l'objet d'une attention particulière dans ce cours. En effet, comme l'explique Grais (2000, p.13) les modalités d'un caractère doivent être à la fois incompatible et exhaustives ; incompatible dans le sens ou un individu ne peut appartenir à la fois à deux ou plusieurs modalité et exhaustive dans le sens ou tous les cas ont été prévus. Dans l'exemple de l'état matrimonial il parait impossible pour un individu d'appartenir à deux ou plusieurs catégories de même que toutes les possibilités ont été énumérée.

<u>Exercices pour révision</u>

<u>Exercice 1</u> : Répondre par vrai ou faux aux affirmations suivantes :

- La population statistique désigne un ensemble d'êtres humains.
- A chaque individu d'une population statistique on peut associer une ou plusieurs modalités.
- Lorsque les différentes modalités d'un caractère ne sont pas mesurables, on parle d'un caractère quantitatif.
- Le recensement est une enquête exhaustive effectué auprès de toute la population. Par contre le sondage est une enquête partielle effectué auprès d'un échantillon représentatif.
- Lorsque les valeurs prises par un individu statistique appartiennent à un intervalle de valeurs, on parle de variable statistique continue.
- A chaque individu de la population on peut associer un ou plusieurs caractères.

<u>Exercice 2</u> : Classer les caractères suivant selon leur nature :

La couleur des cheveux, le lieu de naissance, le bénéfice net d'une entreprise, la langue maternelle, le nombre de véhicules possédé par les ménages, les catégories socioprofessionnelles, le nombre de pièces d'un appartement et l'âge des habitants d'une commune.

<u>Exercice 3 :</u> Déterminer la population statistique, l'unité statistique, le caractère étudié et la nature du caractère des données statistiques suivante.

- Le sport pratiqué par chacun des élèves d'une école.
- Le nombre d'enfant par ménage dans un quartier.
- La taille des arbustes d'une pépinière en cm.

<u>Exercice 4 :</u> Une enquête menée auprès de 160 jeunes au sujet de leur sport préféré a fourni les résultats suivants : 60 jeunes préfèrent le football, 25 la natation, 45 l'athlétisme, 25 le tennis et 5 le handball.
Questions :

1. Déterminer la population statistique.
2. Quel est le caractère étudié ? préciser sa nature.
3. Quelles sont les différentes modalités du caractère ?
4. Construire un tableau statistique.
5. Calculer les fréquences relatives.

<u>Corrigé des exercices</u>

<u>Exercice 1</u> :

- La population statistique désigne un ensemble d'êtres humains. **Faux**
- A chaque individu d'une population statistique on peut associer une ou plusieurs modalités. **Faux**
- Lorsque les différentes modalités d'un caractère ne sont pas mesurables, on parle d'un caractère quantitatif. **Faux**
- Le recensement est une enquête exhaustive effectué auprès de toute la population. Par contre le sondage est une enquête partielle effectué auprès d'un échantillon représentatif. **Vrai**

- Lorsque les valeurs prises par un individu statistique appartiennent à un intervalle de valeurs, on parle de variable statistique continue. **Vrai**
- A chaque individu de la population on peut associer un ou plusieurs caractères. **Vrai**

Exercice 2 :

Classement des caractères suivant selon leur nature :

Caractère qualitatif	**Caractère quantitatif**
La couleur des cheveux	Le bénéfice net d'une entreprise
Le lieu de naissance	Le nombre de véhicules possédé par les ménages
La langue maternelle	Le nombre de pièces d'un appartement
Les catégories socioprofessionnelles	L'âge des habitants d'une commune

Exercice 3 :

Détermination de la population statistique, de l'unité statistique, du caractère étudié et de la nature du caractère.

Population Statistique	**Unité statistique**	**Caractère**	**Nature du caractère**
les élèves d'une école	Un élève d'une école	Le sport pratiqué	Qualitatif
Les ménages d'un quartier	Un ménage d'un quartier	Le nombre d'enfant	Quantitatif Discret
les arbustes d'une pépinière	Un arbuste d'une pépinière	La taille	Quantitatif Continu

Exercice 4 :

6. La population statistique : **160 jeunes**
7. Le caractère étudié : **Sport préféré**. Sa nature : **Qualitative**
8. Les différentes modalités du caractère : **le football, la natation, l'athlétisme, le tennis et le handball**

Sport préféré	Nombre de jeunes (n_i)	Fréquences relatives (f_i)
Football	60	37.50
Natation	25	15.62
Athlétisme	45	28.13
Tennis	25	15.62
Handball	5	3.13
Total	160	100

CHAPITRE II

TABLEAUX ET GRAPHES DES DISTRIBUTIONS STATISTIQUES A UN CARACTERE

Dans le chapitre précédant, nous nous somme intéresser aux méthodes de collecte de données statistiques. Cependant, ces données brutes, pour être facilement interprétables, doivent êtres rangées et regroupées dans des tableaux statistiques. De plus, pour permettre une lecture rapide des données statistiques il est judicieux de les représenter graphiquement.

La présentation des données sous forme de tableau statistique vise à présenter de manière simplifiée et claire les données pour le lecteur; tandis que les graphiques permettent un aperçu encore plus simple et surtout plus rapide du phénomène étudié (Py, 1996).

2.1. Généralités sur le tableau statistique

La première chose à faire pour synthétiser des données numériques en grand nombre est de compter combien de fois chaque modalité de la série statistique apparaît. Pour cela, on range ces modalités dans l'ordre croissant

On appelle alors effectif de la modalité, et on ne le note ni, le nombre d'occurrences de cette modalité dans la série statistique.

On obtient alors un tableau statistiques constitué des modalités et des effectifs du caractère. Par ailleurs, la somme des effectifs de chaque modalité est l'effectif total de la population, que l'on notera N.

Cela étant quel que soit le type de caractère et la nature de la variable, le tableau statistique se présente toujours de la même manière et suivant le même principe. La structure d'un tableau statistique de base se présente généralement comme suit :

Tableau n°… : « Intitulé du tableau et date des données…………… »

Caractère	Effectifs (n_i)
Modalité 1	n_1
Modalité 2	n_2
.	.
.	.
.	.
Modalité k	n_k
Total	N

On appelle cette opération du passage des données brutes au tableau statistique le tri à plat (Anderson & al., 2006).

l'*effectif* ou la fréquence absolu d'une valeur est le **nombre d'individus** associé à chaque modalité.

Par contre **la *fréquence relative*** notée f_i associée à une valeur est la **proportion d'individus** associés à cette valeur. C'est un nombre entre 0 et 1, ou exprimé en pourcentage.

Si n_i est l'effectif d'une valeur et N l'effectif total de la population, la fréquence relative associée à cette valeur sera $f_i=n_i/N$ (ou $f_i=n_i/N\times100$ si elle est exprimée en pourcentages).

- **Effectifs cumulés, Fréquences cumulées**

En plus des modalités et des effectifs ou des fréquences relatives du caractère, un tableau statistique peut aussi contenir, notamment dans le cas du caractère quantitatif des effectifs et des fréquences cumulées croissante et/ décroissante.

L'effectif cumulé croissant noté $n_i\nearrow$ (ou décroissant, noté $n_i\searrow$) de xi est la somme des effectifs des valeurs inférieures (ou supérieures) ou égales à xi.

Ainsi les effectifs cumulés croissants se déterminent comme suit

- $n_1\nearrow = n_1$
- $n_2\nearrow = n_1\nearrow + n_2$
- $n_3\nearrow = n_2\nearrow + n_3$
- . . .
- . . .
- . . .
- $n_k\nearrow = n_{k-1}\nearrow + n_k = N$

Par contre les effectifs cumulés décroissants se déterminent comme suit :

- $n_1\searrow = N$
- $n_2\searrow = n_1\searrow - n_1$
- $n_3\searrow = n_2\searrow - n_2$
- . . .
- . . .
- . . .
- $n_k\searrow = n_{k-1}\searrow - n_{k-1} = n_k$

La fréquence cumulée croissante notée $f_i\nearrow$ ou F_i (décroissante notée $f_i\searrow$) de xi est la somme des fréquences des valeurs inférieures (ou supérieures) ou égales à xi. La détermination des fréquences cumulée croissantes ou décroissantes se fait de la même manière que pour le calcul des effectifs cumulée croissants ou décroissants.

La détermination des effectifs ou des fréquences cumulés croissante et décroissante permet de répondre aux questions : combien en chiffre ou en proportion ont plus de ….ou moins de…un ensemble de modalités ? directement du tableau statistique sans avoir à faire les calculs régulièrement.

Tout tableau statistique établi dans le cadre d'une recherche scientifique doit obligatoirement être présenté selon la structure définie ci-dessus et doit contenir les indications suivantes (Hamdani, 2006) :

• Intitulé ou titre du tableau Il a pour rôle d'indiquer ce qui est représenté globalement par le tableau.

• Le numéro du tableau Dans tout travail de recherche, la présentation des tableaux doit être numérotée, ce qui facilite les renvois et facilite la lecture du manuscrit.

• Intitulés des colonnes Ils renseignent sur le contenu de chaque colonne, de telle sorte à clarifier encore davantage l'intitulé du tableau lui-même.

• L'unité de mesure Elle offre un renseignement supplémentaire en indiquant au lecteur dans quelle unité sont mesurées les modalités étudiées (kilogramme, milligramme, gramme, etc.).

• La date Elle renseigne le lecteur sur la période ou la date à laquelle sont collectées les données présentées dans le tableau. L'obligation de préciser la date des données, répond au souci que certaines données sont très variables dans le temps, la date permet d'apporter une certaine consistance et une certaine pertinence aux données présentées 14 par le chercheur. En méthodologie, cela reflète l'honnêteté intellectuelle du chercheur (ex ; on ne peut pas parler de la situation de la population algérienne en 2013, en s'appuyant sur des données de 1965 !).

• La source Elle indique d'où proviennent les données présentées. Cette indication obligatoire permet, selon le cas, de confirmer ou d'infirmer la pertinence de ces données. Elle reflète aussi l'honnêteté intellectuelle du chercheur. Par exemple, certaines sources sont réputées pour la non fiabilité de leurs données, des données constituées par le chercheur lui-même ne sont pas forcément acceptables par la communauté scientifique,…

Exercice d'application : Une enquête auprès d'un groupe d'étudiants relative au nombre de livres que chacun d'entre eux a lu au cours des vacances d'été a donné les résultats suivants : Karim (1) ; Bilal (02) ; Brahim (03) ; Souad (05) ; Dounia (05) ; Nawal (00) ; Malik (02) ; Reda (01) ; Hamid (03) ; Omar (04) ; Islam (01) ; Hanane (02) ; Farid (02) ; Rachid (03) ; Zohra (01) ; Ines (05) ; Mohamed (00) ; Khaled (1) ; Juba (02) ; Rafik (01) ; Dihia (03) ; Akli (01) ; Kamel (02) ; Toufik (04) ; Djamel (03) ; Mahdi (01) ; Nesma (02) ; Samir (01) ; Amar (00) ; Fatiha (02).

Questions :

1. Quel est le caractère étudié ? préciser sa nature.
2. Quelles sont les différentes modalités du caractère ?
3. Construire un tableau statistique.
4. Calculer les fréquences relatives.
5. Calculer les fréquences relatives cumulées.
6. Quelle est la proportion des étudiants ayant lu moins de trois livres ?
7. Quelle est la proportion des étudiants ayant lu plus de deux livres ?
8. Quelle est la proportion des étudiants ayant lu, au moins, trois livres ?

9. Quelle est la proportion des étudiants ayant lu, au plus, quatre livres ?
10. Quelle est le nombre d'étudiants ayant lu moins de trois livres ?

Réponses :

1. Le caractère étudié : **Les livres lus au cours des vacances d'été ;** sa nature : **Quantitatif Discret.**
2. Les différentes modalités du caractère : **00, 01, 02, 03, 04 et 05**
3. Construction du tableau statistique et élaboration des différents calculs.

Nombre de livres lus (x_i)	Nombre d'étudiants (n_i)	f_i	f_i (%)	$n_i \nearrow$	$n_i \searrow$	$f_i \nearrow$	$f_i \searrow$
00	03	0.10	10	3	30	10	100
01	09	0.30	30	12	27	40	90
02	08	0.27	27	20	18	67	60
03	05	0.17	17	25	10	84	33
04	02	0.06	6	27	5	90	16
05	03	0.10	10	30	3	100	10
Total	30	1	100	///	///	/////////	////////

6. La proportion des étudiants ayant lu moins de trois livres : **67%** ce sont ceux qui ont lus 02 livres +01 livre +aucun livre (10+30+27). La lecture du résultat se fait directement dans la colonne des f_i croissants.
7. La proportion des étudiants ayant lu plus de deux livres : 33% ce sont les étudiants ayants lus 03, 04 et 05 livres (10+06+17=33). En fait, la lecture se fait directement dans la colonne des fréquences relatives cumulées décroissantes.
8. La proportion des étudiants ayant lu au moins trois livres (**donc plus de deux livres**) : **33%.**
9. La proportion des étudiants ayant lu au plus quatre livres **(moins de 05 livres) 90%.**
10. Le nombre d'étudiants ayant lu moins de trois livres : 20 étudiants. La lecture se fait directement à l'aide de la colonne des effectifs décroissant.

Section 2 : Représentation graphique

Il y a lieu de préciser, de prime abord que pour chaque type de caractère et de variable il existe un ou plusieurs modes de représentation graphique spécifique. De plus il existe plusieurs types de représentation graphique que nous ne pouvons pas voir dans le cadre de ce cours de première année.

2.1. Le caractère qualitatif

Deux types de représentation graphique sont souvent utiliser pour les caractères qualitatifs ; il s'agit du diagramme à secteur circulaire et du diagramme à barre

Le diagramme à secteur circulaire est à privilégier pour représenter des séries dont le caractère est qualitatif. Les parts du diagramme ont des aires qui sont proportionnelles aux effectifs de chaque modalité.

Exemple : soit le tableau qui représente la répartition des entités économique en Algérie, en 2011, selon le secteur d'activité

Secteur d'activité	Nombre d'entités	%
Industrie	95 445	10
Construction	9 117	1
Commerce	511 700	55
Services	317 988	34
Total	934 250	100

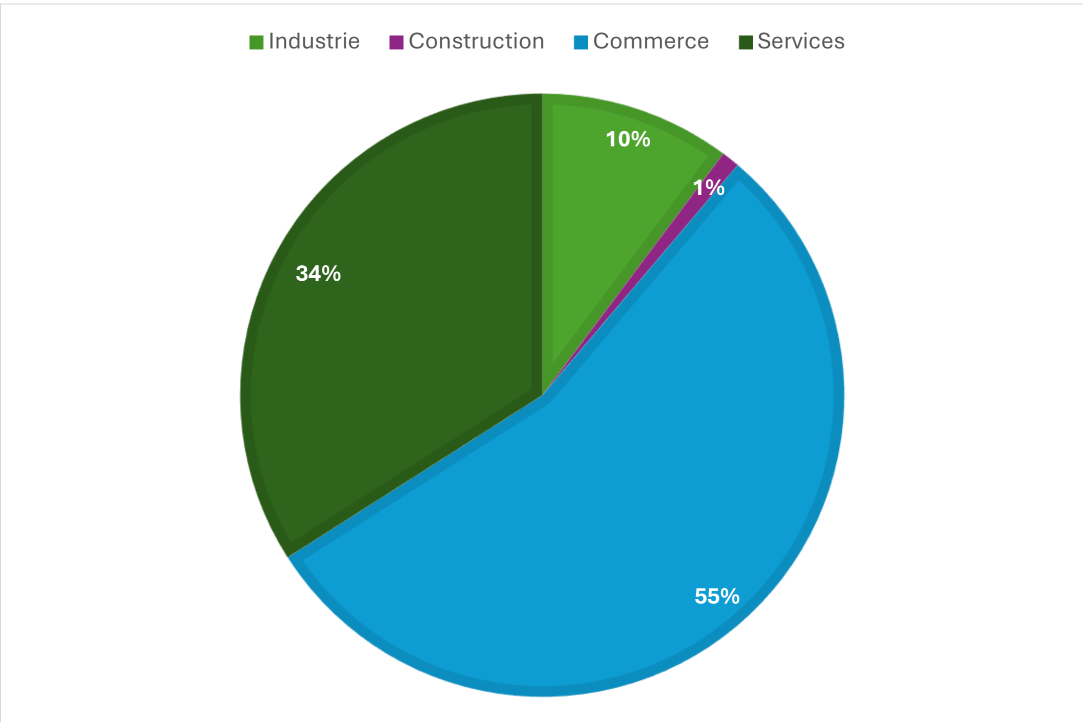
Question : représenter ces données par un diagramme à secteur circulaire.

Réponse : A l'aide du logiciel Exel dans lequel on introduit juste les deux premières colonnes nous obtenons la représentation ci-après.

On utilise sinon la règle de trois pour transformer les effectifs ou plus simplement les fréquences relatives en degrés

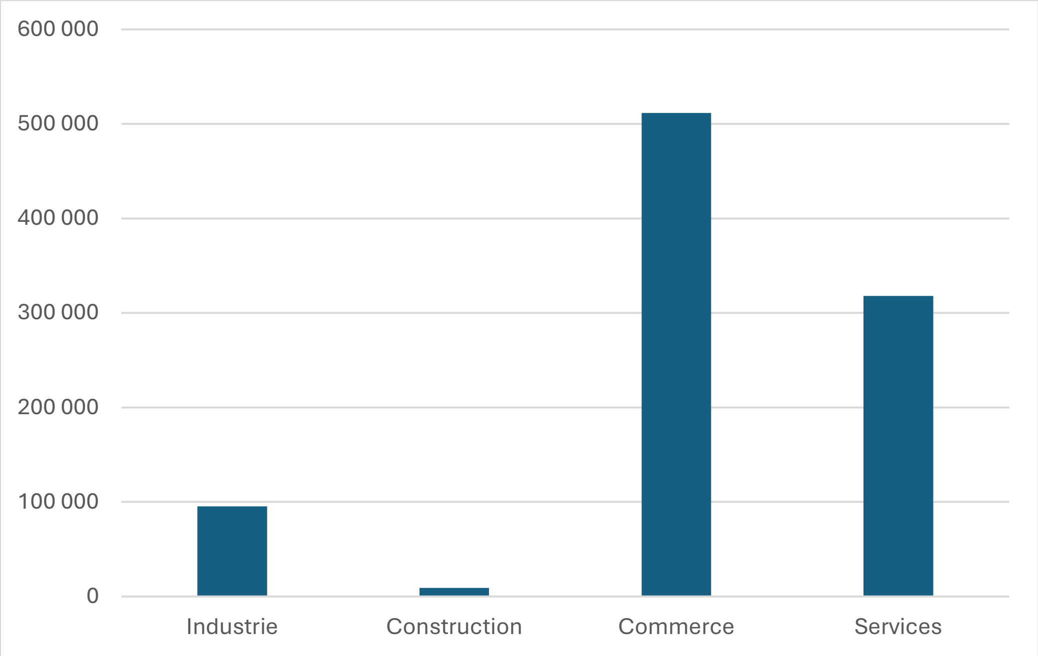
Ex : 100% $\longrightarrow$ 360°

10% $\longrightarrow$ X   X=10x360/100   X=36°



Le diagramme à barres (en tuyaux d'orgue) est un autre type de graphique utilisé pour les caractères qualitatifs. C'est la longueur de chaque barre qui est proportionnelle aux effectifs ou aux fréquences relatives.

Exemple : reprenons l'exemple précèdent et construisons un diagramme en tuyaux d'orgue



Exercice supplémentaire

Le tableau suivant donne l'évolution de la production de céréales en Algérie (en milliers de quintaux) entre 1962 et 1992.

Années / Céréales	1962	1963	1972	1973	1992
Blé dur	23600	23200	26805	10345	38083
Blé tendre	12200	12700	11810	4154	12918
Orge	2880	3280	6668	1991	5776
Avoine	8200	6900	7427	3897	18100

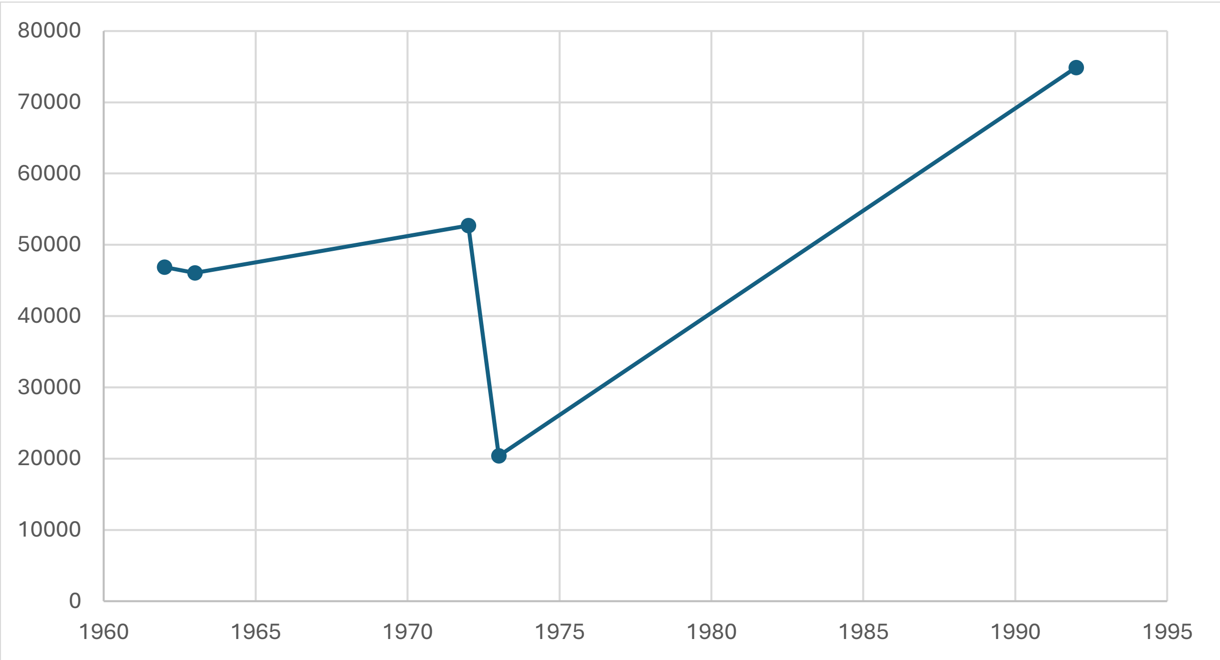
1. Faire une représentation graphique pour l'ensemble des céréales par une droite.
2. Représenter graphiquement chacune des céréales (digramme à barre).

3. Représenter les quatre types de céréales ensemble pour chaque compagne ou année. Et comparer cette nouvelle représentation à la précédente.

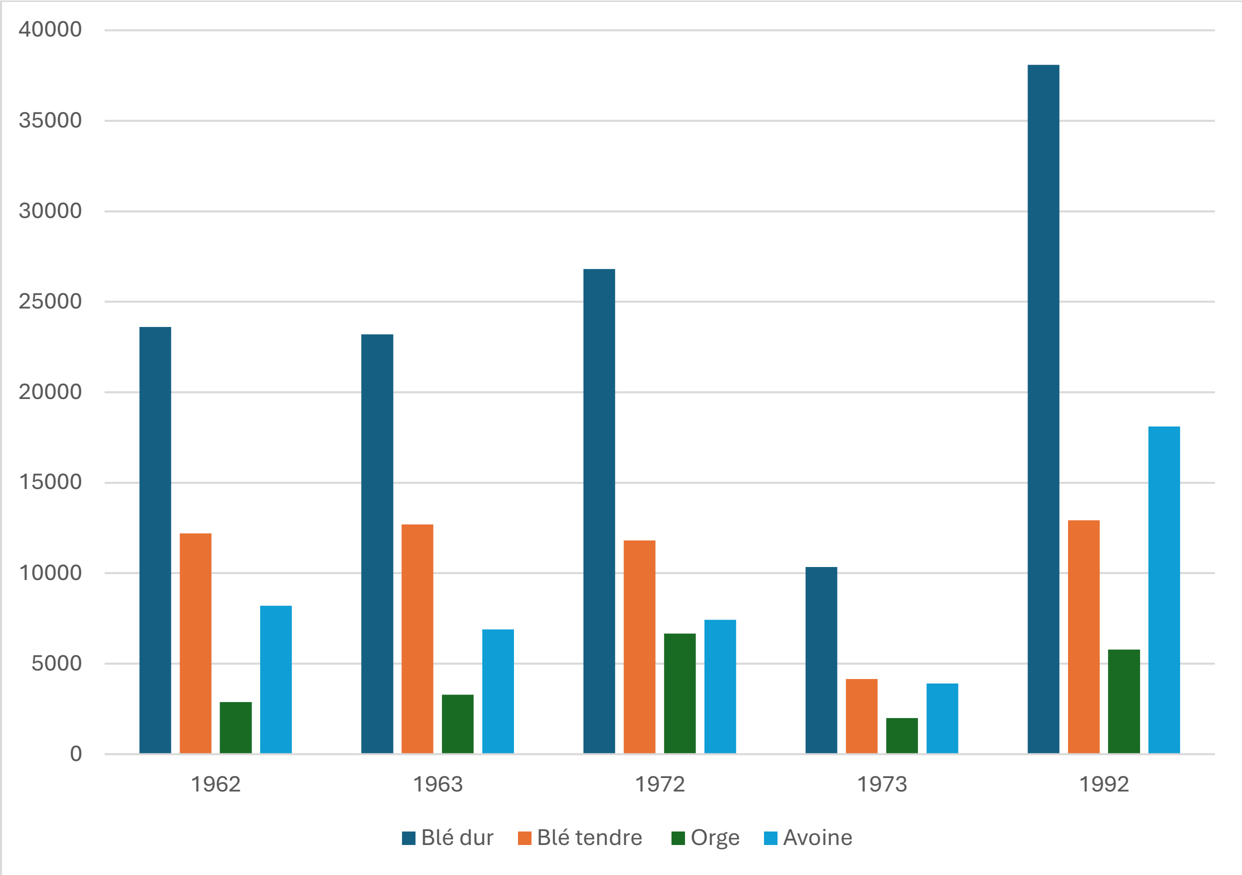
Réponses

1. Il faut d'abord calculer le total pour chaque année et ensuite faire une droite.

	1962	1963	1972	1973	1992
Blé dur	23600	23200	26805	10345	38083
Blé tendre	12200	12700	11810	4154	12918
Orge	2880	3280	6668	1991	5776
Avoine	8200	6900	7427	3897	18100
Total	46880	46080	52710	20387	74877



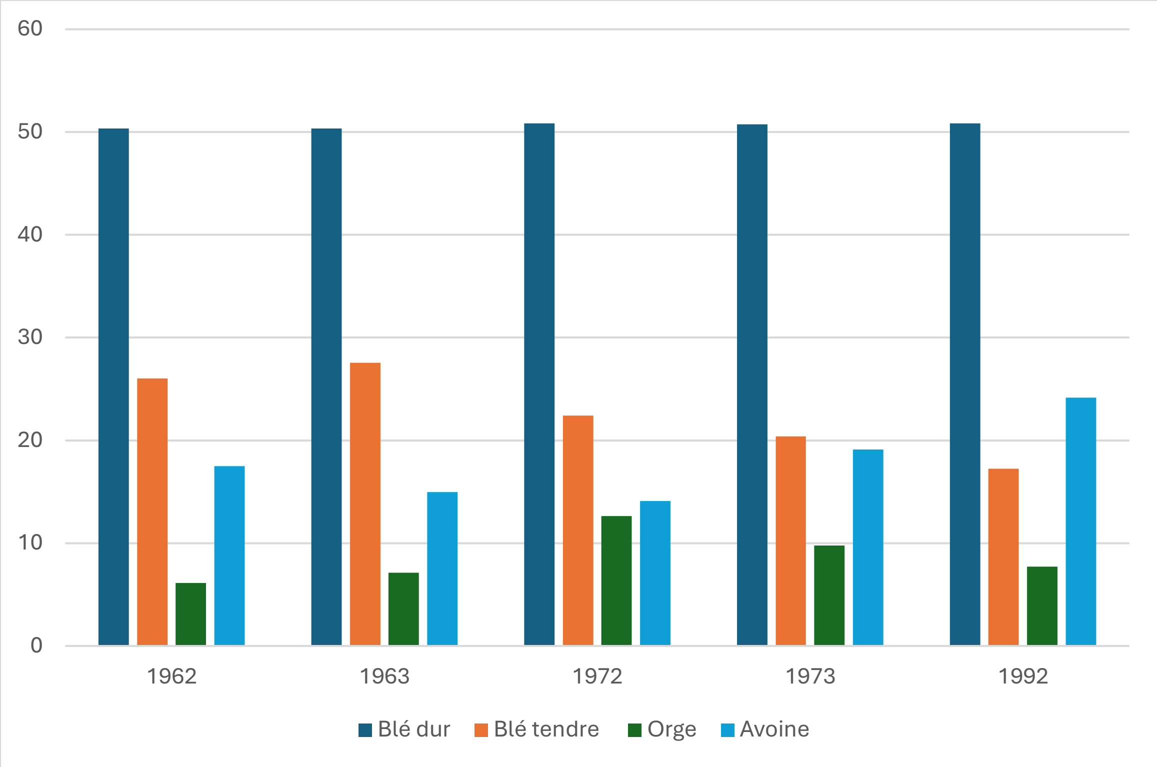
2. **Dans ce cas il faut simplement faire la représentation suivante**



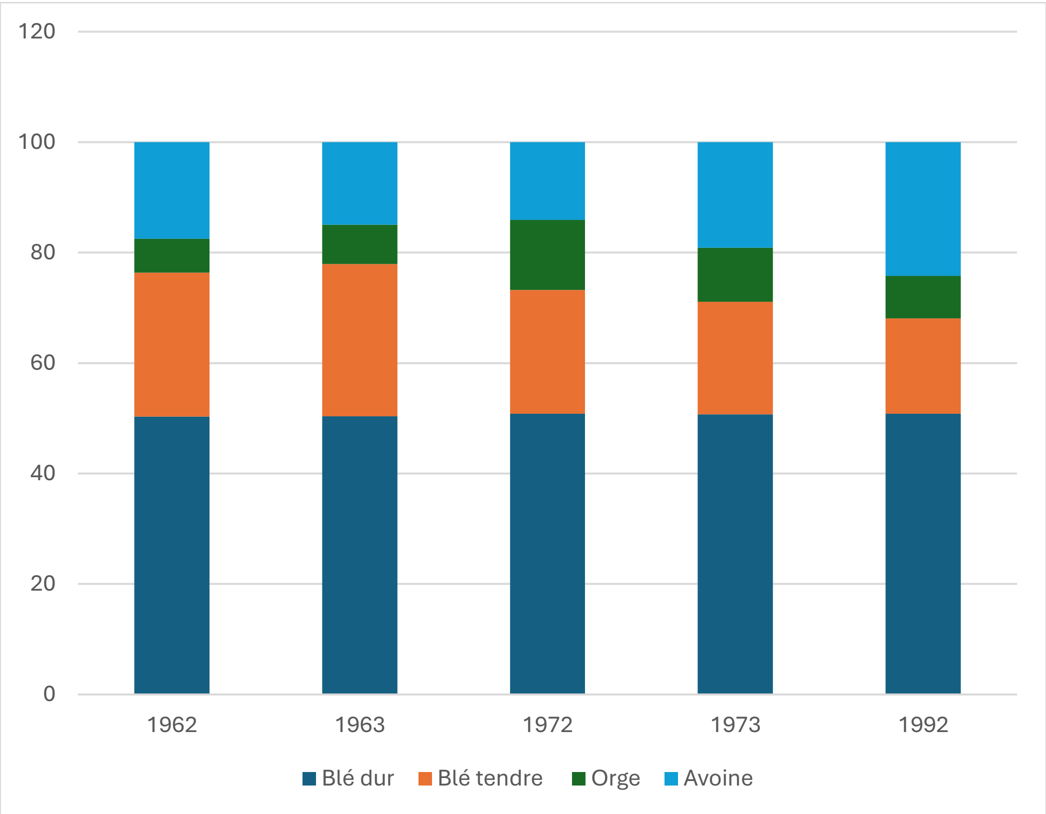
3. Il faut d'abord construire le tableau suivant et ensuite faire des tuyaux d'orgue accolées ou superposées. La deuxième représentation est meilleure car elle permet de mieux visualiser l'évolution de la part de chaque céréale annuellement.

	1962	1963	1972	1973	1992
Blé dur	50,34	50,35	50,85	50,74	50,86
Blé tendre	26,02	27,56	22,41	20,38	17,25
Orge	6,14	7,12	12,65	9,77	7,71
Avoine	17,5	14,97	14,09	19,11	24,17
Total	100	100	100	100	100

a. Tuyaux d'orgues accolés



b. Tuyaux d'orgues superposés.



2.2. Le caractère quantitatif

a. Variable statistique discrète

Deux types de représentations graphiques sont souvent utilisées dans le cas des variables statistiques discrète. Il s'agit ***du diagramme en bâton*** qui s'emploi pour la représentation des effectifs et des fréquences et de la ***courbe cumulative dite en escalier*** qui s'emploie pour la représentation des effectifs ou fréquences cumulés.

Diagramme à bâtons

La représentation graphique des effectifs d'une variable (ou caractère) discret s'effectuera sous la forme d'un graphique en bâtons.

La valeur observée (recueillie ou donnée) du caractère (ou « variable ») est portée sur l'axe des « x » (abscisses) et l'effectif (valeur recensée) correspondant sera portée sur l'axe des « y » appelé « axe des ordonnées » ou « axe des n_i »

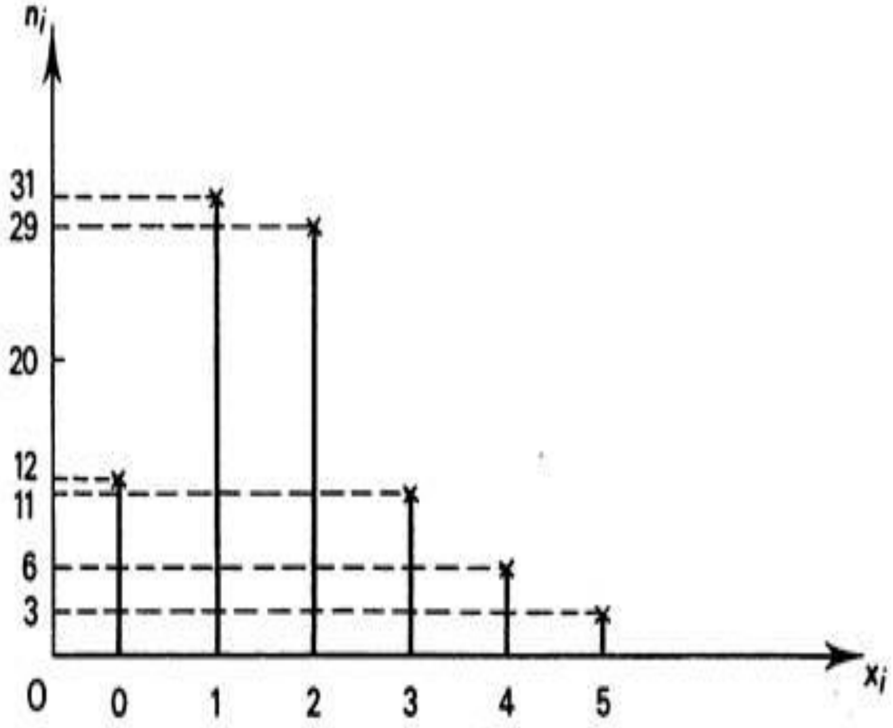
L'axe des abscisses sera l'axe des « x_i » et l'axe des ordonnées sera l'axe des « f_i ».

Le « bâton » est le segment dont la hauteur (ou sa longueur) est proportionnelle à l'effectif correspondant.

Exemple : la série statistique ci-dessous représenté par le diagramme en bâtons, informe sur la façon dont se distribue le personnel (foyer) d'une entreprise en fonction du nombre d'enfants.

Nombre d'enfants par foyer « x_i »	Nombre de foyer concernés n_i
0	**12**
1	**31**
2	**29**
3	**11**
4	**6**
5	**3**

Question : Représenter graphiquement les données de ce tableau.



Les courbes cumulatives

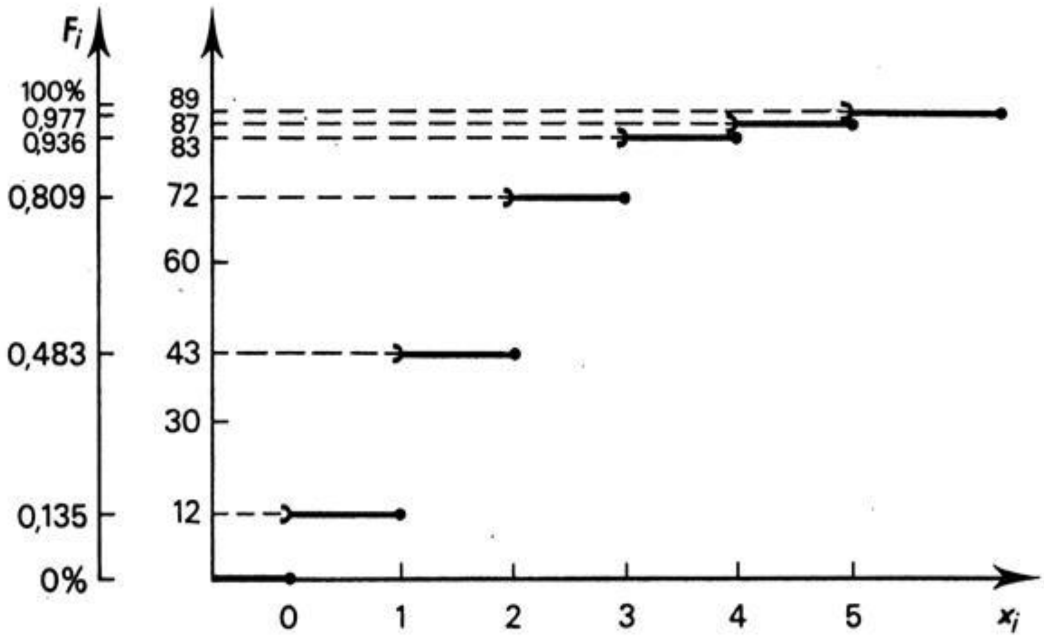
La représentation graphique des effectifs ou des fréquences cumulées croissante et décroissante d'une variable discrète s'effectue sous la forme d'un graphique en escalier.

Les « sauts » correspondent aux valeurs possibles de la variable et sont égaux aux « fréquences cumulées croissantes ou décroissantes.

Exemple : En 1997 : On a fait un sondage auprès des 89 foyers d'une citée, on veut étudier la répartition des foyers en fonction du nombre d'écran (téléviseur ou ordinateur) par foyer.

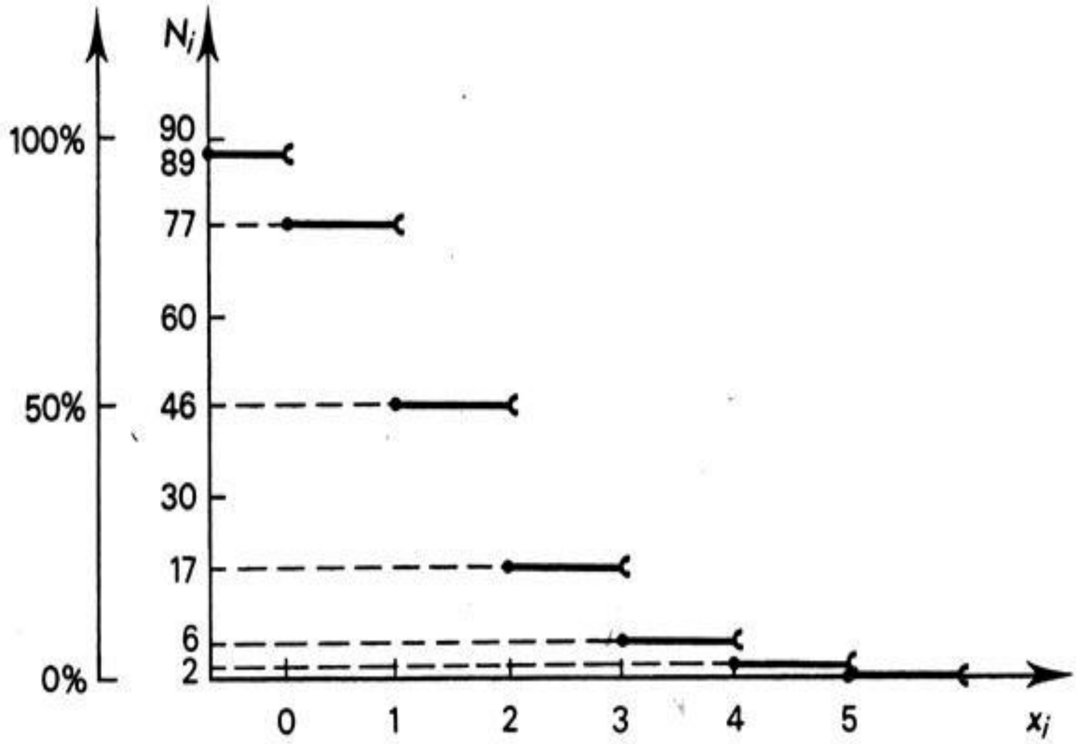
x_i	**Effectifs**			**Fréquences**		
	n_i	**$n_i \nearrow$**	**$n_i \searrow$**	**f_i**	**F_i ($f_i \nearrow$)**	**$f_i \searrow$**
0	12	12	89	0,135	0,135	1
1	31	43	77	0,348	0,483	0,865
2	29	72	46	0,326	0,809	0,517
3	11	83	17	0,124	0,936	0,191
4	4	87	6	0,045	0,977	0,067
5	2	89	2	0,022	1	0,022
Total	**89**			**1**		

Les fréquences ou les effectifs cumulé croissant se présentent comme suit :



On peut observer par exemple sur le graphe que 72 foyers (ou 80,9 %) possèdent 2 écrans ou moins.

Par contre les effectifs ou les fréquences cumulées décroissantes se représentent comme suit :



On remarquera, aussi, par exemple, que 46 foyers (soit 51,7%) possèdent 2 écrans ou plus.

b. Variable statistique continue

Comme pour les variables discrètes, on resonne en termes d'effectifs ou de fréquences simples ou en termes d'effectifs ou de fréquences cumulées. Ainsi, pour la représentation des effectifs ou fréquences simples deux représentations sont privilégiées, à savoir ; ***l'histogramme*** ou le **polygone des fréquences**, tandis que pour la représentation des effectifs ou fréquences cumulées on fait appel aux courbes cumulatives.

- **L'Histogramme**

La représentation graphique des effectifs ou des fréquences d'une variable continue est appelée un histogramme.

On portera en abscisses les valeurs des classes des caractères (variables) et on portera en ordonnées les effectifs ou les fréquences relatives correspondants.

Principe de construction de l'histogramme : Pour chaque classe, on élève un rectangle ayant une base proportionnelle à l'intervalle de classe et une hauteur proportionnelle à l'effectif ou fréquence simple. Dans ce cas, ce sont les « surfaces », et non les hauteurs, qui « **sont proportionnelles à l'effectif** ».

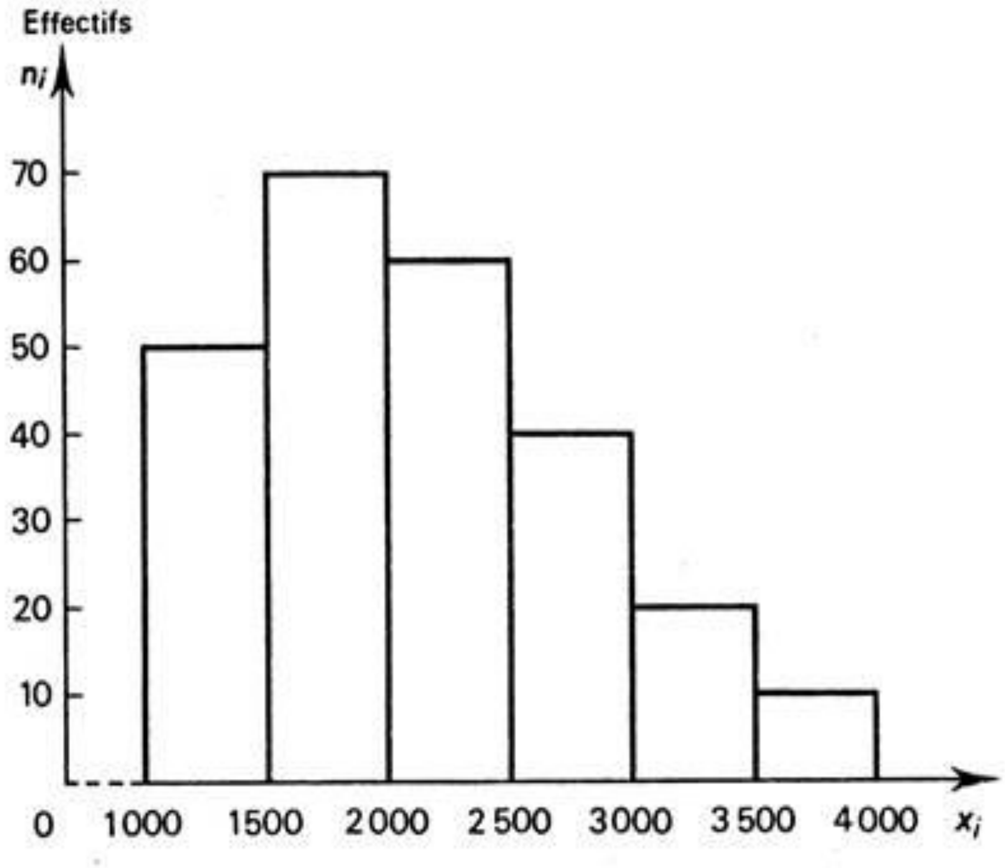
Cependant, avant de construire l'histogramme il faut vérifier si les classe sont d'amplitudes égales ou inégales.

Cas 1. Les amplitudes sont égales

Dans ce cas l'Histogramme se construit facilement comme dans l'exemple suivant :

Exemples : construire l'histogramme pour le tableau suivant qui représente la distribution des salaires mensuels, en $, des salariés d'une société.

Salaire	**Effectifs ni**
1000 à 1500	50
1500 à 2000	70
2000 à 2500	60
2500 à 3000	40
3000 à 3500	20
3500 à 4000	10
Total	250



Cas 2. Les amplitudes de classes sont inégales

Supposons que l'on nous demande d'observer pour la distribution précédente la nouvelle répartition des salaires de l'année suivante. L'entreprise étant expansion, nous remarquons une amélioration des salaires mais aussi que les amplitudes de classes sont inégales.

Salaire en $	**Effectifs ni**	**Amplitude a_i**
1000 à 1500	20	500
1500 à 2000	40	500
2000 à 2500	60	500
2500 à 3500	90	1000
3500 à 4000	30	500
4000 à 4500	10	500
Total	250	

Dans le cas des amplitudes, pour respecter la proportionnalité des surfaces, il va falloir, dans ce cas, « rectifier », en conséquence, les hauteurs. Autrement la représentation graphique n'est pas correcte.

Plusieurs méthodes sont possibles : corriger les effectifs (n_ic) ou les fréquences (f_ic) ou calculer la densité soit avec les effectifs ($D_i=n_i/a_i$) soit avec les fréquences ($D_i=f_i/a_i$)

Avant de corriger les effectifs ou les fréquences on doit calculer les différentes amplitudes pour pouvoir repérer "***l'amplitude de base***" qui correspond à la plus petite amplitude. Une fois cette amplitude repérée on corrige les effectifs ou les fréquences dont l'amplitude est différente de l'amplitude de base.

Dans notre exemple nous considérons que l'amplitude de base (a_b) est égale à 500, tandis l'amplitude de la quatrième classe a_4= 1000. Ainsi, nous déterminons y tel que :

$Y=a_4/a_b$= 1000/500= 2

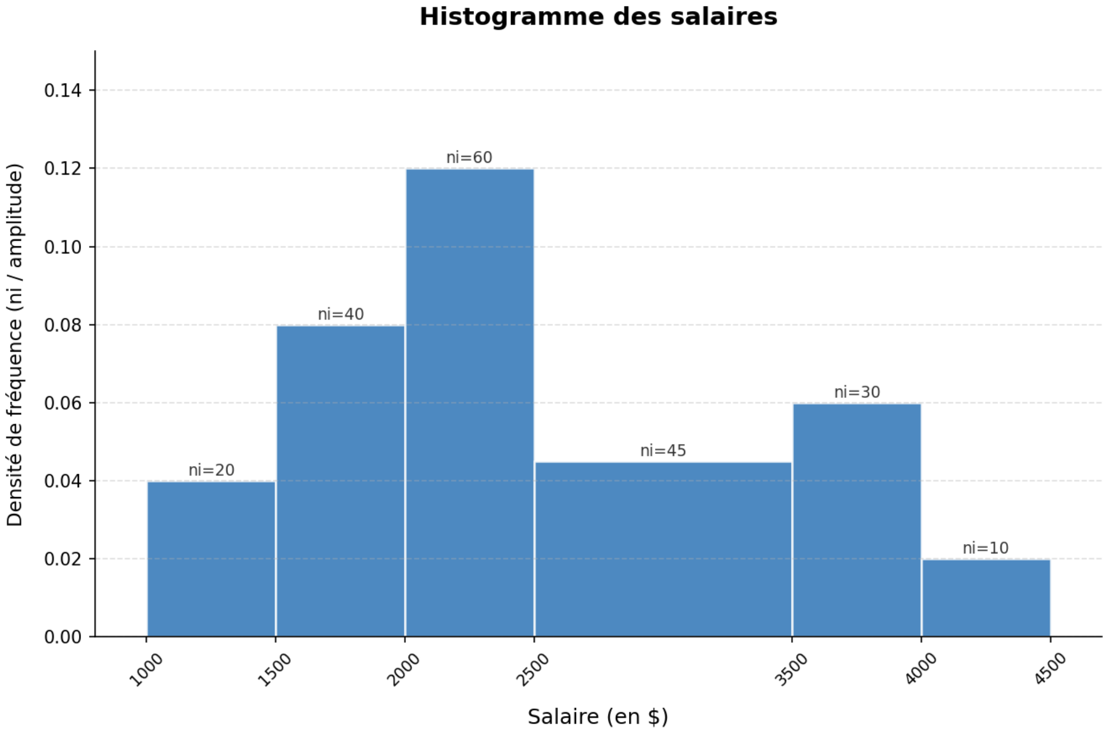
Ensuite nous calculons n_4c tel que $n_4= n_4c/y$ = 90/2= 45.

En fait, c'est comme si la classe [2500-3500[est devisée en deux classes [2500-3000[et [3000-3500[contenant chacune 45 individus.

Nous raisonnons de la même manière pour corriger les fréquences relatives.

Remarque : dans le cas ou le tableau contient plusieurs classes d'amplitudes différentes les unes des autres, ou dans le cas ou la détermination de y donne lieu à un nombre décimale (chiffre avec virgule) il est préférable d'utiliser les densités au lieu de corriger les effectifs ou les fréquences.

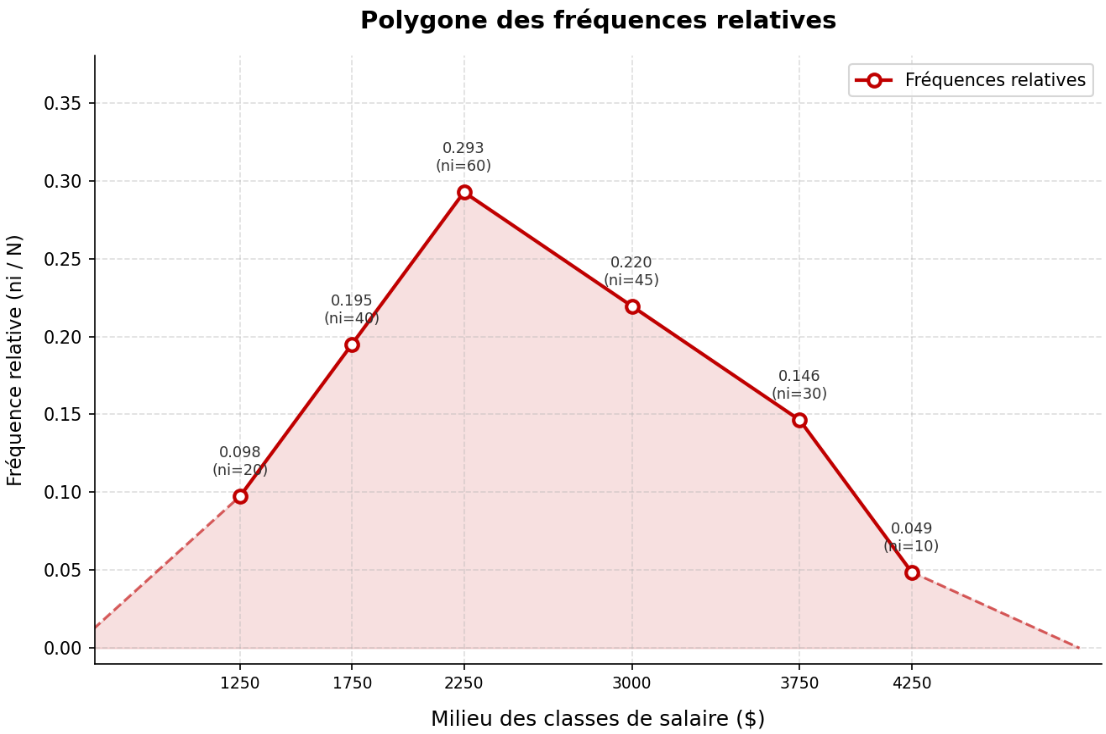
Salaire en $	**ni**	**a_i**	n_ic	f_i	f_ic	$D_i=n_i/a_i$	$D_i=f_i/a_i$
1000 à 1500	20	500	20	8	8	0.04	0.016
1500 à 2000	40	500	40	16	16	0.08	0.032
2000 à 2500	60	500	60	24	24	0.12	0.048
2500 à 3500	90	1000	**45**	36	18	0.09	0.036
3500 à 4000	30	500	30	12	12	0.06	0.024
4000 à 4500	10	500	10	4	4	0.02	0.008
Total	250	/	/	100	/	/	/



- **Le polygone des fréquences**

Le principe de la représentation par polygone des fréquences repose sur la même logique que celle de l'histogramme, c'est à dire qu'on raisonne en termes de surface. En fait, la construction du polygone se fait après la représentation de l'histogramme : on joint les milieux des sommets des rectangles, de classes égales, après avoir ajouter deux fausses classes aux extrémités au milieux desquelles va démarrer et s'arrêter le polygone.

Exemple : tracer le polygone des fréquences pour l'exemple précédant.



- **Les courbes cumulatives**

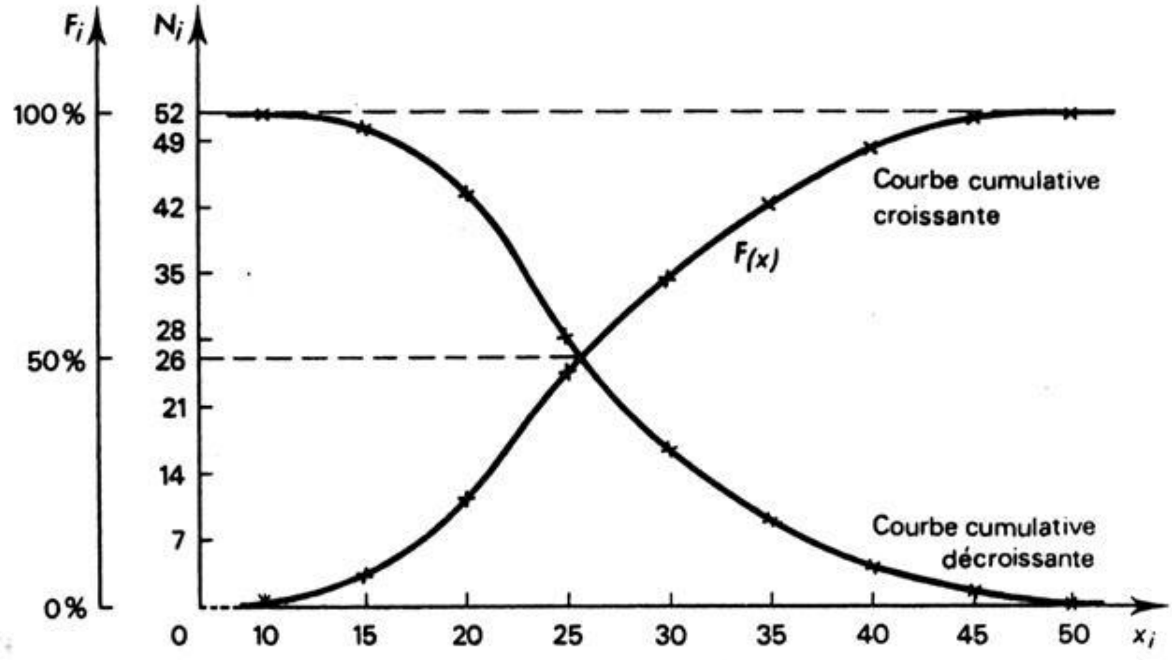
Sur un repère orthonormé on peut tracer soit une courbe des effectifs ou fréquences cumulées croissantes soit une courbe des effectifs ou fréquences cumulées décroissantes soit les deux à la fois.

La courbe des effectifs ou fréquences cumulées croissantes se construit en portant les points correspondant à chaque classe à la limite « supérieure » de l'intervalle de classe.

La courbe des effectifs ou fréquences cumulées décroissantes se construit en portant les points correspondant à chaque classe à la limite « inférieure » de l'intervalle de classe

Remarque : dans la représentation des courbes cumulatives on utilise les effectifs ou les fréquences de l'exercice sans aucune correction même si les classes sont d'amplitudes inégale. Il faut juste faire attention à l'échelle.

Classes	n_i	$n_i \nearrow$	$n_i \searrow$	f_i	$f_i \nearrow$	$f_i \searrow$
10 ; 15	3	3	52	5.77	5.77	100
15 ; 20	9	12	49	17.31	23.08	94.23
20 ; 25	12	24	40	23.07	46.15	76.92
25 ; 35	18	42	28	34.62	80.77	53.85
35 ; 40	6	48	10	11.54	92.31	19.23
40 ; 45	3	51	4	5.77	98.08	7.69
45 ; 50	1	52	1	1.92	100	1.92
Total	52	/	/	/	/	



Exercices pour révision

Exercice 1 : La répartition des employés d'une entreprise selon la catégorie socio-professionnelle (CSP) se présente comme suit :

C.S.P	Ouvriers	employés	Cadres moyens	Cadres supérieurs	$\sum$
n_i	500	250	200	50	.....

Question : Donner deux représentations graphiques à cette distribution

Exercice 2 : Le tableau suivant donne l'évolution de la production de céréales en Algérie (en milliers de quintaux) entre 1962 et 1992.

Années / Céréales	1962	1972	1992
Blé dur	23600	26805	38083
Blé tendre	12200	11810	12918
Orge	2880	6668	5776
Avoine	8200	7427	18100

Question : Représenter les quatre types de céréales ensemble pour chaque compagne ou année.

Exercice 3 : Soit la répartition du nombre de terminaux de connexion internet par bureau dans une administration publique.

Nombre de terminaux	0	1	2	3	4	∑
Nombre de bureaux	10	8	6	4	2	....

Questions : 1- Représenter graphiquement cette distribution.

2- Calculer les effectifs et les fréquences cumulés croissants et faire la représentation graphique adéquate.

Exercice 4 : Soit la distribution des salaires horaires (en Euro) dans une entreprise :

Classes	[5-10[	[10-15[	[15-20[	[20-25[	[25-30[	∑
n_i	4	6	12	5	3	....

Questions : 1- Donner deux représentations graphiques à cette distribution

2- Calculer les effectifs et les fréquences cumulés croissants et décroissants et faire la représentation graphique adéquate.

Exercice 5 : le regroupement des deux dernières classes de l'exercice précédent nous donne la distribution suivante :

Classes	[5-10[	[10-15[	[15-20[	[20-30[	∑
n_i	4	6	12	8	....

Questions : 1- Donner deux représentations graphiques à cette distribution

2- Calculer les effectifs et les fréquences cumulés croissants et décroissants et faire la représentation graphique adéquate.

CHAPITRE III.

LES PARAMETRES DE TENDANCE CENTRALE.

Pour faciliter l'interprétation des données statistiques on résume ces dernières par des valeurs représentatives qu'on appelle « paramètres ou caractéristiques ». Trois types de paramètres peuvent exister ; il s'agit des paramètres de tendance centrale, de paramètres de dispersion et de paramètres de formes et de concentration.

Dans ce présent chapitre nous nous intéressons aux caractéristiques de tendance centrale qui représentent des valeurs d'une distribution d'un caractère statistique choisie pour sa représentativité d'une tendance de cette distribution, soit son centre (la moyenne arithmétique), soit sa fréquence dominante (le mode) ou sa position médiane dans l'ensemble des valeurs (la médiane).

Ainsi, trois paramètres de position peuvent être étudiés :

- Le Mode
- La Médiane
- La Moyenne arithmétique

Avant d'aborder ces paramètres, nous devons souligner, comme la précise le statisticien Britannique Yule (1945, cité par Grais, 2000) qu'une bonne caractéristique de tendance centrale ou de dispersion doit répondre aux propriétés suivantes :

- Être définit de façon objective
- Dépendre de toutes les observations de façon à caractériser réellement toute la serie
- Avoir une signification concrète facile à concevoir
- Être simple à calculer
- Être peu sensible aux fluctuations d'échantillonnage
- Se prêter aisément aux calculs algébriques

Section 1 : le Mode

1.1.Définition

Le *mode d'une série statistique* est une valeur de la série pour laquelle l'effectif associé est le plus grand. Autrement dit, c'est la valeur de la variable qui se répète le plus (la plus fréquente)

Une série statistique peut comporter deux modes (si deux valeurs sont maximales).

Une série statistique peut comporter aucun mode (si aucune valeur ne se répète ou plusieurs valeurs sont maximales).

Exemples

1- Soit la distribution des données suivantes : 2, 2, 5, 8, 10, 10, 10, 15, 16, 22.

Le mode de cette distribution est 10. On écrit alors : $M_o = 10$.

2- Soit encore la distribution des données suivantes : 2, 2, 2, 5, 8, 10, 10, 10, 15, 16, 22.

Dans ce cas la distribution possède deux modes et on dit qu'elle est bi-modale M_o= [2 ; 10]. Le calcul s'arrête ici car on ne fait pas le moyenne des deux valeurs.

3- Si encore nous avons les deux séries suivantes

Serie 1 : 2, 2, 2, 5, 5, 5, 8, 10, 10, 10, 15, 16, 22.

Serie 2 : 2, 5, 8, 10, 15, 16, 22.

Dans ces deux cas la série n'a pas de Mode

1.2. Cas des variables discrète

Dans le cas d'une variable discrète, le Mode correspond à la valeur du caractère x_i correspondant au plus grand effectif (n_i) ou à la plus grande fréquence relative (f_i).

Exemple : Soit le tableau suivant qui donne le nombre d'enfants par ménage.

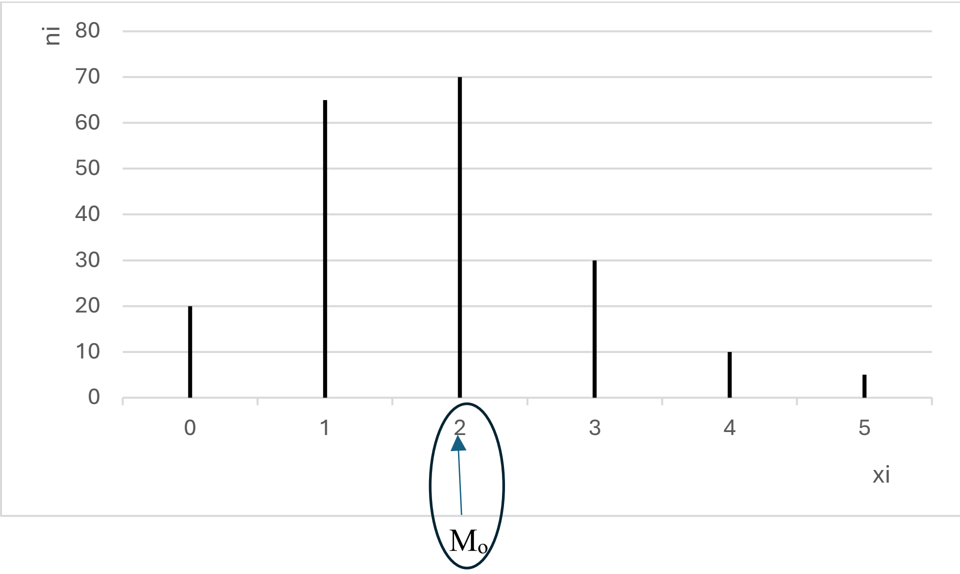
Nombre d'enfants (x_i)	Nombre de ménages (n_i)	Fréquences relatives (f_i)
0	20	0.1
1	65	0.325
2	70	0.35
3	30	0.15
4	10	0.05
5	5	0.025

Question : déterminer le Mode de cette distribution

Réponse : Nous observons sur la colonne des n_i que le plus grand effectif est 70 qui correspond à la variable x_i= 2. Ainsi le Mode M_o est égal à 2.

On peut aussi trouver cette valeur à l'aide des fréquences relatives fi max = 0,35 $\Longrightarrow$ M_o= 2

Remarque : nous pouvons aussi déterminer le Mode d'une variable statistique discrète graphiquement à l'aide de du diagramme en bâton. Dans ce cas le mode correspond au bâton le plus haut comme c'est illustré dans le graphique suivant.



1.3.Cas des variables continues

Pour déterminer le Mode d'une variable statistique continue il nous faut d'abord déterminer la classe modale. Si toutes les valeurs sont regroupées en classes de même dimension, la classe modale est celle dont l'effectif est le plus élevé dans la distribution statistique.

Une fois la classe modale située on calcule le Mode par la formule suivante :

$$M_O = x_0 + d\frac{(n_{mo} - n_{mo-1})}{(n_{mo} - n_{mo-1}) + (n_{mo} - n_{mo+1})}$$

x_0: est la limite inférieure de la classe modale

d_i: est l'amplitude de la classe modale

n_{mo}: est l'effectif de la classe modale

n_{mo-1} : est l'effectif de la classe précédent la classe modale

n_{mo+1} : est l'effectif de la classe après la classe modale

Exemple : déterminer le mode pour la distribution statistique suivante

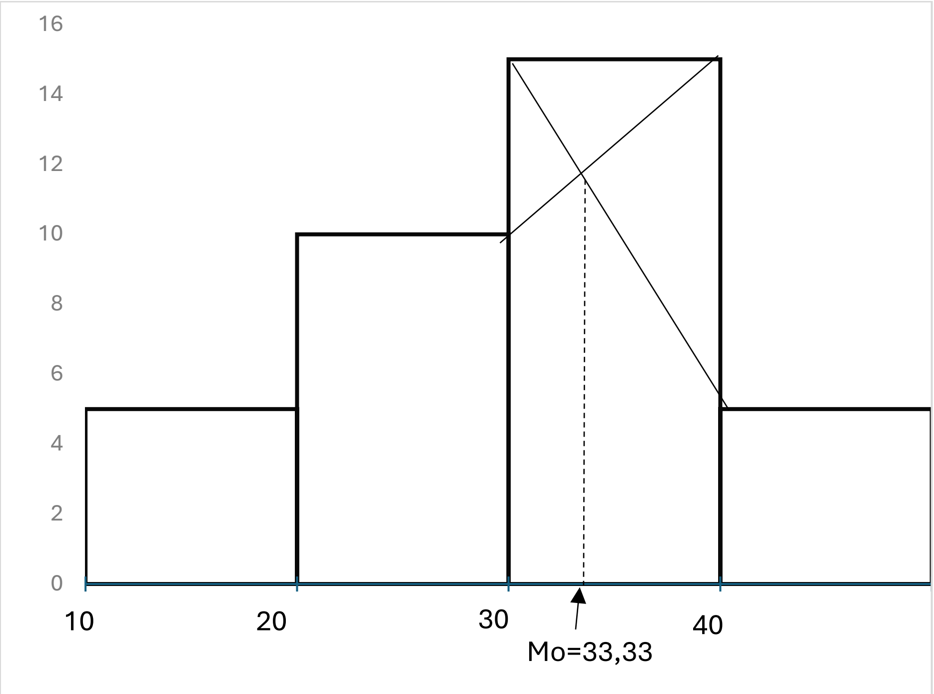
Classes	**n_i**	**f_i**
[10-20[	5	14.29
[20-30[	10	28.57
[30-40[	15	42.85
[40-50[	5	14.29
Total	35	100

La classe modale est [30-40[, car c'est celle qui à l'effectif et/ou la fréquence maximum.

Sachant que : $M_O = x_0 + d\frac{(n_{mo} - n_{mo-1})}{(n_{mo} - n_{mo-1}) + (n_{mo} - n_{mo+1})}$

Alors : $M_O = 30 + 10 \frac{(15-10)}{(15-10)+(15-5)} = 33{,}33$ Mo=33,33

Nous pouvons, par ailleurs, déterminer le mode graphiquement comme c'est illustrer dans le graphique suivant :



Remarque : dans le cas où les classes sont d'amplitudes inégales il nous faut d'abord corriger les effectifs ou les fréquences relatives avant de calculer le mode ou le représenter graphiquement (voir chapitre 2)

Section 2 : La Médiane

1. Définition

La médiane notée « Me » est la valeur de la variable qui partage la série statistique en deux sous-ensembles égaux.

Avant de la déterminer, il faut d'abord ranger les valeurs de la variable par ordre croissant ou décroissant.

Exemple : soit la série statistique suivante : 10- 25- 6- 30- 3- 16- 20.

Question : déterminer la Médiane de cette série.

Réponse : série ordonnée : 3- 6- 10- 16- 20- 25- 30

M_e

2. Calcul de la Médiane

a- Cas d'une variable discrète

- Si le nombre d'observations est *impair*, la médiane se calcul facilement, comme dans l'exemple précèdent.

- Si le nombre d'observations est *pair*, il n'y a pas de Médiane mais un « intervalle médian ».
  Exemple : soit la série suivante : 5- 6- 8- 9- 10- 11- 14- 15.
  $\sum n_i = 8$ (pair) $\longrightarrow M_e = [9- 10]$
- Sur le tableau statistique ; la médiane est déterminée à partir de la colonne des effectifs ou fréquences cumulés, où on repère la valeur $(\frac{N}{2})$ ou 0.5 ou 50%.

Exemple : Soit le tableau suivant qui donne le nombre d'enfants par ménage.

x_i	n_i	f_i	$n_i \nearrow$	$f_i \nearrow$
0	20	0.1	20	0.1
1	65	0.325	85	0.425
2	70	0.35	155	0.775
3	30	0.15	185	0.925
4	10	0.05	195	0.975
5	5	0.025	200	1

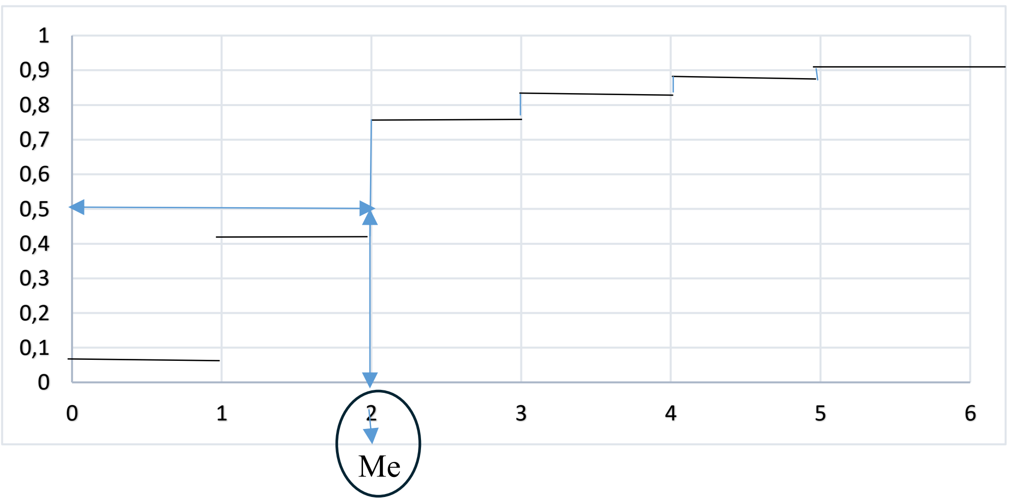
Question : Déterminer la Médiane

$TH_2 = (\frac{N}{2}) = \frac{200}{2} = 100$. On cherche la valeur dans ($n_i \nearrow$) ; elle se trouve dans 155 $\Rightarrow M_e = 2$

Ou encore ; 0.5 se trouve dans 0.775 dans les ($f_i \nearrow$) $\Rightarrow M_e = 2$

- On peut aussi déterminer la Médiane graphiquement à l'aide de la courbe en escalier.

Exemple : représenter le tableau précédant à l'aide d'une courbe en escalier et repérer la Médiane.



b- **Variable statistique continue**

Dans le cas d'une variable continue, il faut d'abord repérer la classe Médiane. La classe Médiane se détermine à l'aide des effectifs ou fréquences cumulés, où on repère la valeur $(\frac{N}{2})$ ou 0.5 ou 50%.

Une fois cette classe Médiane déterminée, on calcule la Médiane à l'aide de la formule suivante.

$$M_e = x_0 + d\,\frac{TH_2 - (n_{me-1}\nearrow)}{n_{me}}$$

x_0 : limite inférieure de la classe Médiane.

d : amplitude de la classe Médiane.

$TH_2 = (\frac{N}{2})$ ou 0.5 ou 50%.

$\overset{\nearrow}{n}_{me-1}$: effectif cumulé croissant avant la classe Médiane.

n_{me} : effectif de la classe Médiane.

Exemple : Calculer la Médiane pour le tableau statistique suivant.

Classes	n_i	f_i	$f_i \nearrow$	$n_i \nearrow$	$n_i \searrow$
10-20	5	14.29	14.29	05	35
20-30	10	28.57	42.86	15	30
30-40	15	42.85	85.75	30	20
40-50	5	14.29	100	35	5
Total	35	100			

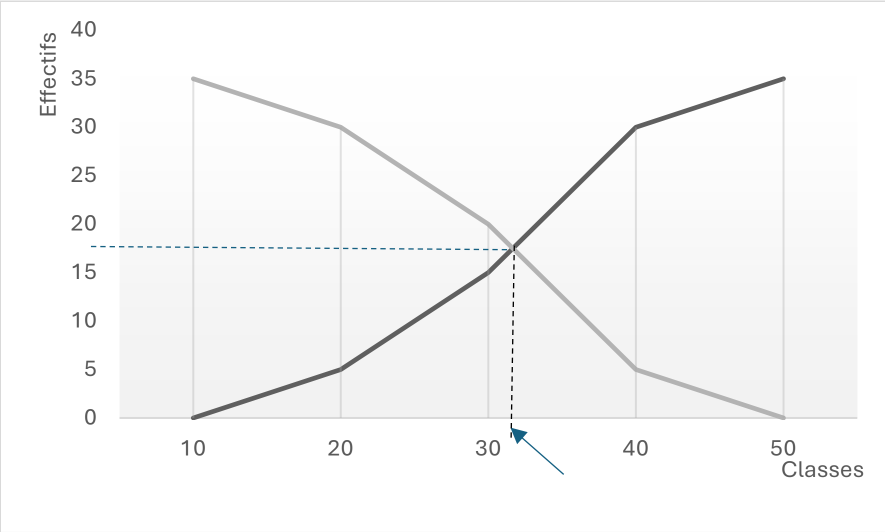
$TH_2 = (\frac{N}{2}) = (\frac{35}{2}) = 17.5$. Cette valeur se trouve dans 30 dans les $\overset{\nearrow}{n_i}$.

Ainsi, la classe Médiane est [30-40[

$$M_e = 30 + 10\,\frac{17.5-15}{15} = 31.66$$

$$\text{Ou } M_e = 30 + 10\,\frac{50-42.86}{42.85} = 31.66$$

- La Médiane peut, aussi, être déterminer graphiquement à l'aides des courbes cumulatives, comme suit.



La Médiane correspond à la valeur de x_i correspondant à l'intersection des deux courbes.

3- *Généralisation de la Médiane :(les quantiles)*

La logique est la même que la Médiane. On cherche les valeurs de la variable qui divisent la série statistique, non plus, en deux, mais en quatre, en dix ou en cent sous-ensembles égaux.

a. Les Quartiles (Q_i) ce sont des valeurs qui partagent la série statistique en quatre sous ensemble égaux qui contiennent chacun 25% des observations ou un quart ($\frac{N}{4}$) du total des effectifs.

Le calcul se fait comme pour la Médiane.

$$Q_i = x_0 + d\,\frac{TH_i - (n_{Qi-1})\nearrow}{n_{QI}} \qquad TH_i = \left(\frac{iN}{4}\right) \qquad i = 1, 2, 3$$

b. Les Déciles : ce sont des valeurs qui partagent la série stat en dix sous-ensembles, qui contiennent chacun 10% ou ($\frac{N}{10}$) des observations. Le calcul se fait comme suit.

$$D_i = x_0 + d\,\frac{TH_i - (n_{Di-1})\nearrow}{n_{DI}} \qquad TH_i = \left(\frac{iN}{10}\right) \qquad i = 1, 2, 3 \ldots\ldots 9.$$

c. Les Centiles : ce sont des valeurs qui partagent la série stat en cent sous-ensembles, qui contiennent chacun 1% ou ($\frac{N}{100}$) des observations. Le calcul se fait comme suit.

$$C_i = x_0 + d\,\frac{TH_i - (n_{Ci-1})\nearrow}{n_{CI}} \qquad TH_i = \left(\frac{iN}{100}\right) \qquad i = 1, 2, 3 \ldots\ldots 99$$

Remarque: $M_e = Q_2 = D_5 = C_{50}$

Section 3 : La moyenne arithmétique ($\overline{X}$)

1. Définition

La moyenne arithmétique ($\overline{X}$) d'une série statistique est égale à la somme des valeurs observées par le nombre des observations.

- On dit qu'une moyenne arithmétique est simple lorsqu'à chaque valeur de la variable ne correspond qu'une seule observation.

$$\overline{X}= \frac{\sum x_i}{N}$$

Exemple : calculer la moyenne des notes d'un étudiant, dont les notes sont.

13-8- 7- 12- 14- 5

$\overline{X}= \frac{\sum x_i}{N} = \frac{13+8+7+12+14+5}{6} =9.83$

- Par ailleurs, on dit qu'une moyenne arithmétique est pondérée lorsqu'à chaque valeur de la variable correspond plusieurs observations.

$\overline{X}= \frac{\sum x_i n_i}{N}$ $\overline{X}= \frac{\sum x_i n f_i}{\sum fi}$

Remarque : la moyenne arithmétique pondérée s'emploie quand les données sont regroupés en classes ou quand les données discrètes se répètent.

2. **Procédés de calculs**.

a. Variable statistiques discrètes.

Il faut, généralement, ajouter la colonne ($x_i.n_i$) ou la colonne ($x_i.f_i$) avant de faire le calcul.

Exemple : reprenez l'exemple précédant et calculer la moyenne arithmétique.

x_i	n_i	f_i	($x_i.n_i$)	($x_i.f_i$)
0	20	10	0	0
1	65	32.5	65	32.5
2	70	35	140	70
3	30	15	90	45
4	10	05	40	20
5	5	02.5	25	12.5
Total	200	100	360	180

$\overline{X}= \frac{\sum x_i n_i}{N} = \frac{360}{200} = 1.8$ $\overline{X}= \frac{\sum x_i n f_i}{\sum fi} = \frac{180}{100} = 1.8$

b. Variable statistique continue

Par convention, on choisit, pour les calculs, comme valeur de la variable (x_i), le centre de chaque classe. Une fois tous les centre de classe déterminée, on peut ajouter la colonne ($x_i.n_i$) ou la colonne ($x_i.f_i$)

Exemple : Calculer la moyenne arithmétique des données suivantes.

Classe	n_i	f_i	x_i	$x_i.n_i$	$x_i.f_i$
10-20	5	14.29	15	75	214.35
20-30	10	28.57	25	250	714.25
30-40	15	42.85	25	525	1499.75
40-50	5	14.29	45	225	643.05
Total	35	100		1075	3071.75

$\overline{X}= \frac{\sum x_i n_i}{N} = \frac{1075}{35} = 30.71$ $\overline{X}= \frac{\sum x_i . f_i}{\sum fi} = \frac{3071.75}{100} = 30.71.$

c. Calcul de la moyenne arithmétique par la Méthode de changement de variable

D'une façon générale, pour éviter de trop volumineux calculs, surtout quand les valeurs (x_i) correspondent à des nombres importants, on choisit une nouvelle origine (x_0) et une unité de mesure (a). Et, on déduit une variable auxiliaire **(x'_i)** tel que : $x'_i = \frac{x_i - x_0}{a}$

$(\overline{x'}) = \frac{\sum x'_i n_i}{N}$ $\overline{X}= x_0 + a(\overline{x'})$

Exemple : reprenez l'exemple précédant et calculer $\overline{X}$ par la méthode de changement de variable, avec x_0= 35 et a=10.

Classe	n_i	x_i	x'_i	$x'_i n_i$
10-20	5	15	-2	-10
20-30	10	25	-1	-10
30-40	15	25	0	0
40-50	5	45	1	5
Total	35			-15

$(\overline{x'}) = \frac{\sum x'_i n_i}{N} = \frac{-15}{35} = - 0.428$

$\overline{X}= x_0 + a(\overline{x'})$ = 35- 10(-0.428) = 30.71.

3. Propriétés de la moyenne arithmétique.

- La moyenne arithmétique est influencée par les valeur extrêmes alors que la Médiane ne l'est pas.
- La somme des écarts à la moyenne est nulle ($\sum x_i - \overline{X}$)= 0
- La somme des carrées des écarts à la moyenne est minimale pour a=$\overline{X}$, c'est-à-dire $(\sum x_i - a)^2$ est minimale pour a=$\overline{X}$; a étant une valeur quelconque.
- Si n_1 nombres ont une moyenne m_1 ; n_2 nombres ont une moyenne m_2 ;...n_k nombres ont une moyenne m_k. Alors la moyenne de tous les nombres vaut
  $\overline{X}= \frac{n_1 m_1 + n_2 m_2 + \cdots n_k m_k}{n_1 + n_2 + n_3}$

Exemple : dans une usine employant 80 employés, 60 employés gagnent un salaire moyen de 300DA/Heure et les 20 autres gagnent un salaire horaire moyen de 200 DA/H. Quel est le salaire moyen de tous les employés.

Réponse : $\overline{X}= \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2} = \frac{(60.300)+(20.200)}{60+20}$ =275

4- généralisation de la notion de moyenne

La moyenne arithmétique n'est un cas particulier de la moyenne. En utilisant la même logique de construction on peut définir d'autres moyennes à savoir la moyenne géométrique et la moyenne harmonique.

a. **La moyenne géométrique (G).**

C'est la racine $n^{éme}$ du produit des n valeur positives du caractère x. On l'emploie dans le calcul du taux d'accroissement moyen et dans le calcul des moyennes de coefficient multiplicateurs.

- La moyenne géométrique simple :

$G=\sqrt[n]{x_1 x_2 \ \ldots . x_k}$ ou $G= \sqrt[n]{\pi x_i}$ π : Produit.

Le calcul peut se faire également par le logarithme

$\text{Log } G=\frac{1}{n} \sum log\ x_i$

Exemple : calculer la moyenne géométrique, pour deux valeur 8 et 12.

$G=\sqrt[2]{8.12}$ G= 9,8.

- La moyenne géométrique pondérée.

Dans ce cas, à chaque valeur x_i correspondent plusieurs observations (n_i). Ainsi ;

$G= \sqrt[n]{x_1^{n1} x_2^{n2} \ \ldots \ldots x_k^{nk}}$ ou $G=\sqrt{\prod x_i^{ni}}$ ou $\log = \frac{1}{n} \sum_{x=1}^{k} n_i\ log x_i$

Exemple :

La chiffre d'affaires d'une entreprise augmente de 5% les 2 premières années, de 9% les quatre années suivantes et de 12% la dernière année. Quelle est l'augmentation annuelle moyenne du chiffre d'affaires ?

$G=\sqrt[7]{(1.05)^2(1.09)^4(1.12)^1}$= 1.083.

T_m= G-1 ou G-100

T_m= 1.083-1 =0.083 ou 8.3%.

La chiffre d'affaires a connu un taux d'accroissement annuel moyen de 8.3% durant les (07) ans.

b. **La moyenne Harmonique**.

C'est la valeur de la variable pour laquelle son inverse est la moyenne arithmétique de l'inverse des valeurs de la variable. On l'emploie dans le calcul des moyennes de pourcentage et de rapport et surtout dans celui des durées ou des vitesses moyennes.

- Moyenne Harmonique simple. $H=\frac{N}{\sum \frac{1}{x_i}}$

- Moyenne Harmonique pondérée $H=\frac{N}{\sum \frac{n_i}{x_i}}$

Exemple : sur un trajet, un automobiliste fait 60Km/H à l'aller et 30Km/H au retour. Quelle est la vitesse moyenne sur l'ensemble du trajet.

Vitesse moyenne=Distance totale/durée totale

$Vm=\frac{D1+D2}{T1+T2}=\frac{2D}{\frac{D1}{T_1}+\frac{D1}{T_1}}=\frac{2}{\frac{1}{60}+\frac{1}{30}}=$ $V_{m}=40Km/H$

Exercices pour révision

Exercice 1 : les deux séries ci-dessous représentent les prix en dinars de deux produits alimentaires dans cinq points de vente :
Série 1 : 14,2 ; 13,8 ; 14,2 ; 13,9 ; 14.
Série 2 : 14,1 ; 13,8 ; 14,3 ; 15,2 ; 13,5 ; 14.
Calculer le prix modal, médian et moyen pour chaque série.

Exercice 2 : le tableau suivant indique la répartition des journées de travail des éléments d'une brigade de sapeurs- pompiers, selon le nombre d'interventions quotidiennes, au cours d'une année donnée.

Nombre d'interventions	0	1	2	3	4	5
Nombre de jours	100	130	70	35	20	10

Calculer les trois paramètres de tendance centrale.

Exercice 3 : Les chiffres d'affaires annuels en 10^5 DA de 65 petites entreprises sont représentés par le tableau suivant :

Chiffres d'affaires	[8-10[	[10-14[	[14-16[	[16-18[	[18-24[	[24-30[
Effectifs	5	20	14	8	9	9

1-Calculer le mode et la médiane. Interpréter.
2-Calculer le premier quartile, le neuvième décile et le dixième centile.
3-Calculer la moyenne arithmétique par la méthode de changement de variable (avec : $X_0 = 15$, $a = 2$).
4-Représenter graphiquement le mode et la médiane de cette série.

Exercice 4 : Dans une entreprise, le salaire moyen est de 67 570 DA, le salaire moyen des hommes est de 77 500 DA et le salaire moyen des femmes est de 52 310 DA. Sachant que le nombre d'employés de cette entreprise est de 330, déterminer l'effectif des hommes et des femmes de cette entreprise.

Exercice 5: une entreprise dépense un budget « B » chaque trimestre pour l'achat d'affiches publicitaires.
-Au premier trimestre, le prix de l'affiche était de 350 DA.
-Au deuxième trimestre, le prix de l'affiche était de 380 DA.
-Au troisième trimestre, le prix de l'affiche était de 400 DA.
-Au troisième trimestre, le prix de l'affiche était de 450 DA.

Calculer le prix moyen de l'affiche sur les quatre trimestres.

Exercice 6 : au cours de la période 1994-1999, les exportations du pétrole brut ont évolué de la façon suivante :

Années	1994	1995	1996	1997	1998	1999
Pourcentage de variation par rapport à l'année précédente	-3,8	-18,3	+14 ,2	-3,1	-1,5	+8,9

1-Calculer le taux annuel moyen de variation au cours de la période 1993-1999.
2-Sachant que le taux moyen d'accroissement des exportations était de 9, 2% par an pendant la période 1969-1982, de 12,4% pendant la période 1982-1993, calculer le taux de variation annuel moyen pour la période 1969-1999.

Exercice 7 : un touriste américain en voyage en France, échange des dollars dans les conditions suivantes :
1000$ contre des euros au taux de 1,30$/€.
2000$ contre des euros au taux de 1,32$ /€.
1500$ contre des euros au taux de 1,35$/€.
-Quel est le taux de change moyen ? Quel type de moyenne ce calcul fait –il intervenir ?

De retour aux USA, il lui reste 400€ qu'il échange dans les conditions suivantes :
300€ contre des dollars au taux de 1,25$/€.
100€ contre des dollars au taux 1,20$/€.
-Quel est le taux de change moyen ? Quel type de moyenne ce calcul fait –il intervenir ?

CHAPITRE IV

LES PARAMETRES DE DISPERSION

Après les paramètres de tendance centrale étudiés dans le chapitre précédent, il s’agit dans le présent chapitre d’étudier d’autres paramètres qui consistent à évaluer ou à calculer l’éloignement des valeurs par rapport à leur valeur centrale, le plus souvent leur moyenne arithmétique. Ce sont les paramètres de « dispersion » qu’on appelle aussi les « écarts ».

Si l’on considère l'exemple suivant de B. Py (2007), de deux séries des notes des étudiants dans deux groupes A et B :

Groupe A : 2- 2- 2- 2- 10- 18- 18- 18- 18.

Groupe B : 9- 9- 9- 9- 10- 11- 11- 11- 11.

Par un calcul simple on peut déduire que les deux séries ont la même moyenne arithmétique (et la même médiane). Pourtant, elles reflètent deux réalités différentes. En effet, alors que dans le groupe B les étudiants présentent un niveau général moyen, voire bon, pour tous les étudiants, dans le groupe A nous avons près de la moitié des étudiants qui sont très loin en dessous de la moyenne et une autre moitié qui présente un niveau très loin au-dessus de la moyenne. Autrement dit, le même paramètre X ne reflète pas la même réalité dans les deux séries. Par conséquent, il faut se méfier des raisonnements ayant pour seul support les paramètres de tendance centrale. Ces derniers sont intéressants mais sont insuffisants car, le plus souvent, il est nécessaire d’avoir des renseignements sur la répartition (l’éloignement) des valeurs entre elles et autour de leur valeur centrale (la moyenne le plus souvent), c'est-à-dire sur leur « dispersion ».

Aussi, en statistique, on dépasse le raisonnement par les seuls paramètres de tendance centrale en analysant la dispersion et/ou la concentration des valeurs de la série.

On appelle **dispersion statistique** ; la tendance qu'ont les valeurs de la distribution d'un caractère à s'étaler de part et d'autre d'une valeur centrale et/ou à s'éloigner les unes des autres. Ainsi, la dispersion, objet du présent chapitre, analyse la fluctuation ou l’éloignement des valeurs par rapport à une valeur centrale (généralement la moyenne arithmétique) ou dans un intervalle. Cette dispersion est appréhendée par la notion d’« écarts ».

Dans le présent chapitre nous étudions successivement quatre paramètres de dispersion ; les trois premiers mesurent la dispersion absolue, tandis que le dernier paramètre mesure la dispersion relative

- Les écarts simples
- L’écart absolu moyen
- L’écart type et la variance
- Le coefficient de variation

Section 1. Les écarts simples

Il s'agit essentiellement de l'intervalle de variation, appelé aussi « étendue » (e), des intervalles inter-quantiles ou des intervalles interdéciles ou inter-centiles

1.1. L'étendue

L'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :

Etendue de X = X_{max}-X_{min}

Exemple : soit les données suivantes sur la taille en cm d'un lot de vis produites par une entreprise : 0,90, 0,95 ; 1 ; 1,01 ; 1,05

Etendue **e**=1,05- 0,90= 0,15 cm.

Comme nous le remarquons à travers cet exemple le calcul de l'étendue est des plus simple à réaliser. Cependant, ces résultats sont souvent influencés par les valeurs extrêmes et aberrantes. Ainsi, on recourt aux autres paramètres d'écarts simples.

1.2. L'intervalle interquartiles : Il est égal à Q_3-Q_1

Sachant que Q_1 et Q_3 désignent le premier et le troisième quartile.

Cet indice fournit un renseignement sur l'étalement des valeurs de part et d'autre de la médiane. Evidemment, il faut d'abord calculer le premier et le troisième quartile avant de calculer l'intervalle.

1.3. L'intervalle interdéciles : est égal à D_9-D_1

D_1 et D_9 désignent le premier et le neuvième décile.

Cet indice fournit, également, un renseignement sur l'étalement des valeurs de part et d'autre de la médiane

Section 2. L'Ecart absolu moyen

L'écart absolu moyen est la moyenne de la valeur absolue des écarts à la moyenne. Autrement dit, c'est la distance moyenne à la moyenne. Bien qu'il soit moins utilisé, on peut calculer de la même manière l'écart absolu médian qui est la moyenne des écarts à la médiane**.**

- $E(\bar{x})=\frac{\sum|x_i-\bar{x}|}{N}$ (formule simple)
- $E(\bar{x})=\frac{\sum|x_i-\bar{x}|\, n_i}{N}$ (formules pondérées)

Section 3. La variance et l'écart type

La variance d'une série statistique est la moyenne arithmétique des carrés des écarts à la moyenne arithmétique.

On désigne la variance d'une série statistique par la lettre « V ».

- $V(x)=\frac{\sum(x_i-\overline{x})^2}{N}$ (formule simple)
- $V(x)=\frac{\sum n_i(x_i-\overline{x})^2}{N}$ formule de définition
- $V(x)=\frac{\sum n_i x_i^2}{N}-\bar{x}^2$ formule développée.

L'écart type noté $\sigma(x)$est quant à lui définit, à partir de la variance car il est égal à la racine carrée de la variance

- $\sigma(x)=\sqrt{V(x)}$

Section 4. Le coefficient de variation (Cv):

Lorsque l'on veut comparer la dispersion de deux caractères, le meilleur des indicateurs est le coefficient de variation qui mesure la dispersion relative.

- $Cv=(\sigma(x)/\overline{x}).100$

Exemple : Une machine remplit automatiquement des paquets de farine (marqués 1 kg). Un échantillon de 100 paquets fournit les renseignements suivants :

Poids (en Kg)	**Effectifs n_i**	**x_i**	**$x_i.\ n_i$**	**$\lvert x_i-\overline{x}\rvert\ n_i$**	**$(x_i-\overline{x})^2 n_i$**
0,992-0,996	3	0.994	2,982	0,036	0.004
0,996-1,000	5	0.998	4,99	0,04	0.008
1,000-1,004	24	1.002	24,048	0,096	0,221
1,004-1,008	35	1,006	35,21	0	0
1,008-1,012	21	1,010	21,21	0.084	0.148
1,012-1,016	12	1,014	12,168	0.096	0.110
Total	100		100,608	0.352	0.491

Questions : 1- Calculer les paramètres de dispersion

2- Dire si la machine est bien réglée ? Il est nécessaire pour cela que la moyenne soit comprise entre 1,004 et 1,012 et l'écart type soit inférieur à 0,002

- $e=1,016-0,992= 0,024$
- $\overline{x}=\frac{\sum x_i n_i}{N}=\frac{100,608}{100}$ $\overline{x}=1,006$
- $E(\bar{x})=\frac{\sum\lvert x_i-\bar{x}\rvert\ n_i}{N}$ $E(\bar{x})=\frac{0,352}{100}= 0,0035$
- $V(x)=\frac{\sum n_i(x_i-\overline{x})^2}{N}$ $V(x)=\frac{0,491}{100}= 0,005$

- $\sigma(x)=\sqrt{V(x)}$ $\sigma(x)=\sqrt{0{,}005}=0{,}07$
- Cv=$\sigma(x)/\overline{x}$ =(0,07/1,005). 100= 0,069 ou 6,9%

Les résultats montrent clairement que la machine est mal réglée, car l'écart type est très grand par rapport à la valeur souhaitée.

Exercice supplémentaire : Soit le tableau suivant qui donne la répartition des salaires mensuels d'une entreprise

Salaires 10^3	Effectif	fi=	F(x)	Yi=
[24-27[	4			
[27-30[	21			
[30-33[	104			
[33-36[	163			
[36-39[	121			
[39-42[	57			
[42-45[	22			
[45-48[	10			
Σ	502			

<u>TRAVAIL A FAIRE :</u>

1. Calculer le salaire moyen.

2. Calculer la variance.

3. Calculer l'écart-type.

4. Calculer le coefficient de variation.

Corrigé : Construction du tableau

Salaires 10^3	ni	fi=$\frac{n_i}{N}$ (%)	F(x)	xi	$x_i.\ n_i$	$\boldsymbol{\lvert x_i-\overline{x}\rvert\ n_i}$	$n_i(x_i-\overline{x})^2$	$\boldsymbol{x_i^2.n_i}$
[24-27[	4	0.8	0.8	25.5	102	40.36	407.23	2601
[27-30[	21	4.18	4.98	28.5	598.5	148.89	1055.63	17057.25
[30-33[	104	20.72	25.7	31.5	3276	425.36	1739.72	103194
[33-36[	163	32.47	58.17	34.5	5623.5	177.67	193.66	194010.75
[36-39[	121	24.10	82.27	37.5	4537.5	231.11	441.42	170156.25
[39-42[	57	11.35	73.62	40.5	2308.5	279.87	1374.16	93494.25
[42-45[	22	4.38	97.80	43.5	957	174.02	1376.49	41629.5

[45-48[	10	2	100	46.5	465	109.1	1190.28	21622.5
Σ	502				17868	1586.38	8078.59	643764.75

1. Le salaire moyen.

E($\overline{x}$)= $\frac{\sum|x_i - \overline{x}|\, n_i}{N}$ Et $\overrightarrow{X} = \frac{\sum xi.ni}{N} = \frac{17868}{502} = 35.59$

E($\overline{x}$)= $\frac{1586.38}{502}$ = 3.16

2. La variance.

V(x)= $\frac{\sum n_i (x_i - \overline{x})^2}{N}$ = 8078.59/502= 16.09

Ou V(x)=$\frac{\sum n_i x_i^2}{N} - \bar{x}^2 = \frac{643764.75}{502} - \overline{35.59}^2$= 1282.39- 1266.64= 15.75

3. L'écart-type.

$\sigma(x)$=$\sqrt{\text{V(x)}}$ = $\sqrt{16.09}$ = **4.01**

4. Le coefficient de variation.

Cv=($\sigma(x)$/$\overline{X}$).100= (4.01/35.59). 100= 11.26 %

Exercices pour la révision

Exercice1.

10 étudiants ont obtenu les notes suivantes à deux examens :

a/0-7-7-7-7-7-7-7-10-11
b/3-4-5-6-7-7-8-9-10-11

Question : Peut-on affirmer que ces deux séries sont semblables, en se basant sur leurs paramètres de tendance centrale ($\bar{X}, M_e, M_O$)..

Exercice 2. Calculer l'étendue, l'écart absolu moyen, la variance et l'écart type de chacune des distributions suivantes :

1. 150-250-300-350-200.
2. 9-3-7-6-7-8-2-4-6-5-7-2-8-4-9-6-4-5-5-4-10-9-7-4-2-8-7-7-6-5-3-5-10-9-7-6-3-8-7-6.

Laquelle de ces deux distributions présente la plus grande dispersion ?

Exercice 2

Le tableau suivant indique la distribution du poids de 100 étudiants

Poids	60-63	63-66	66-69	69-72	72-75	Total
Effectifs	5	18	42	27	8	100

- ✓ Calculer l'étendue de la distribution.
- ✓ Calculer l'écart absolu moyen
- ✓ Calculer l'écart type par :
  - La formule de définition,
  - La formule développée.
  - La méthode de changement de variable.

CHAPITRE V
LES PARAMETRES DE CONCENTRATION ET DE FORME

En statistiques descriptives, après avoir calculé les paramètres de position (moyenne, médiane, mode) et de dispersion (variance, écart-type, coefficient de variation), il est indispensable d'étudier la forme de la distribution. Ces paramètres de forme permettent de caractériser plus précisément la distribution d'une variable et de la comparer à la loi normale.

La concentration et sa mesure occupe une place importante en économie. Au départ, la mesure de la concentration ne concernait que les revenus mais par la suite elle a été élargie à d'autres variables qui doivent être toutefois positives, continues et dont l'addition a un sens. A titre illustratif, le salaire est une variable positive, continue et dont l'addition a un sens, alors que l'âge malgré qu'il soit une variable positive, continue mais son addition n'a aucun sens. Autrement dit, le salaire est un caractère sommable, par contre le poids et l'âge ne sont pas sommables.

La concentration peut être définie comme la présence dans une partie de la distribution d'une masse $(x_i\, n_i)$ plus importante qu'ailleurs.

Les caractéristiques de forme sont des paramètres calculés permettant de renseigner sur l'allure, la forme et la configuration de la courbe des fréquences ; c'est-à-dire la relation: $x_i \to f_i$; sans nécessairement tracer celle-ci.

Généralement, on utilise deux grandes mesures de la forme :

- **L'asymétrie :** elle a pour but de renseigner sur la façon régulière ou non dont les observations se répartissent de part et d'autre d'une valeur centrale, souvent$\overline{\mathrm{X}}$.
- **L'aplatissement :** elle sert à déterminer l'effet des variations de la variable x_i sur la variation des fréquences f_i.

Ce chapitre est subdivisé en deux sections. La première traitera les paramètres de concentration, alors que la deuxième sera consacrée à l'étude des caractéristiques de forme.

Section I : La concentration et sa mesure

Une autre manière d'analyser la dispersion pourrait être aussi l'analyse de sa conséquence. En effet, s'il y a dispersion, il y a forcément concentration des modalités. Si celles-ci se dispersent de part et d'autre de leur valeur centrale, c'est pour se concentrer de part et d'autre des deux côtés de la série (Py, 1996). Dispersion et concentration sont donc deux notions interdépendantes.

I.1-Mesure de la concentration par le calcul

La mesure de la concentration se fait à travers la comparaison de "$\boldsymbol{\Delta}$**M**" à l'intervalle de variation de la variable (l'étendue). "$\boldsymbol{\Delta}$**M**" Étant la différence entre la médiane et la médiale.

$$\frac{\boldsymbol{\Delta M}}{\boldsymbol{e}} = \frac{\boldsymbol{M_l - M_e}}{\boldsymbol{x_{max-}x_{min}}}$$

Lorsque "$\boldsymbol{\Delta}$**M**" est important par rapport à l'étendue, la concentration est forte. Lorsque la distribution est égalitaire, la concentration est faible et ΔM est faible par rapport à l'étendue.

- **La médiane** $(\boldsymbol{M_e})$: nous avons déjà vu que c'est la valeur de la variable qui divise les effectifs en deux parties égales.
- **La médiale** $(\boldsymbol{M_l})$: **c'est un indicateur qui s'apparente à la médiane. La médiale est la valeur de la variable qui divise la masse**$(x_i\, n_i)$ en deux parties égales.
- $\boldsymbol{M_l = x_0 + ai\,\frac{\sum x_i n_i/2 - x_i n_{i\,ml-1}^{\nearrow}}{x_i n_{i\,ml}}}$
- $\boldsymbol{M_l = x_0 + ai\,\frac{0.5 - F`_{ml-1}}{f`_{ml}}}$
- $\boldsymbol{f`} = \frac{x_i n_i}{\sum x_i n_i}$
- F`i=$\frac{x_i n_i^{\nearrow}}{\sum x_i n_i}$

x_0: *limite inféreiure de la classe médiale.*

a_i: *amplitude de la classemédiale .*

$\sum x_i n_i/2$: rang de la classe médiale

$x_i\, n_{i\,ml-1}^{\nearrow}$: $x_i\, n_i$cumulé croissant de la classe qui se situe avnt la classe médiale.

$x_i\, n_{i\,ml}$: effectif de la classe médiale.

$F`_{ml-1}$: masse relative cumulée croissante de la classe qui se situe avant la classe médiale.

$f`_{ml}$ = masse relative de la classe médiale.

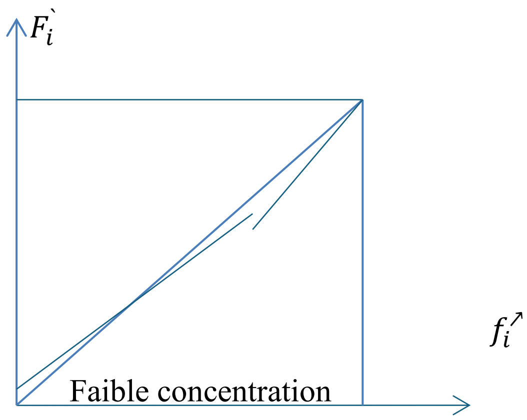
1.2-La courbe de Lorenz et l'indice de Gini

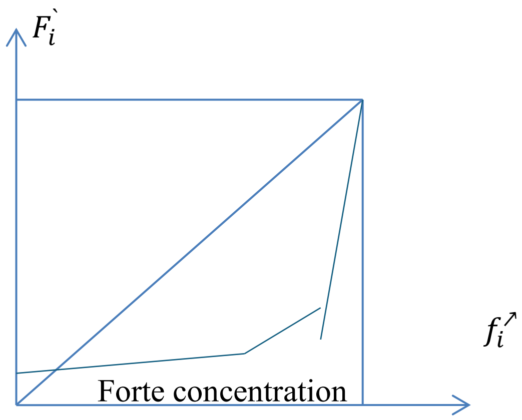
1.2.1-La courbe de Lorenz

La courbe de Loren, du nom de son inventeur Max Lorenz, se trace dans un carré de coté 1. En abscisses, figurent les fréquences relatives cumulées et en ordonnées figurent les $x_i\, n_i$ cumulés rapportés à la somme des $x_i\, n_i$.

On a deux cas :

- Plus la courbe est éloignée de la bissectrice, plus la distribution est concentrée, plus elle se rapproche de la bissectrice, moins la distribution est concentrée.
- Si la courbe se confond avec la bissectrice, la distribution est parfaitement égalitaire.





1.2.2-L'indice de Gini

L'indice de Gini est un ratio qui permet de faire des comparaisons.

- **La méthode Géométrique:** l'indice de Gini est le rapport entre la surface de concentration et la moitié de la surface du carré.

$$\mathbf{IG}= \frac{\boldsymbol{Aire\ de\ concentration}}{\frac{1}{2}} = \mathbf{2} \times \boldsymbol{aire\ de\ concentration} = \mathbf{2\ S}$$

IG varie de 0 à 1 (d'une concentration nulle à une concentration maximale). Pour calculer l'indice de Gini, on peut se servir, entre autres de méthodes d'approximation, de celle donnée par « la méthode des trapèzes » comme le montre le schéma ci-dessous :

La surface d'un trapèze = $\frac{b+b^{`}}{2} \times h$

IG = 2 S

S= (l'aire du triangle ABC)- ($\sum$des aires des trapèzes).

S= $\frac{1}{2} \sum \left(\frac{b+b^{`}}{2}\right) \times h$

S= $\frac{1}{2} - \frac{1}{2}\sum [(F^{`}_{i-1} + F^{`}_{i})\,(F_i - F_{i-1})]$

S= $\frac{1}{2} - \frac{1}{2}\sum [F^{`}_{i-1} + F^{`}_{i})f_i]$

- **La methode analytique:**

IG= $2[\frac{1}{2} - \frac{1}{2}\sum [F^{`}_{i-1} + F^{`}_{i})f_i]$

IG= 1-[$\sum (\boldsymbol{F^{`}_{i-1} + F^{`}_{i})f_i}$]

L'indice de Gini varie de 0 à 1. On a les cas suivants :

- IG= 1 : concentration maximale.
- IG=0 : absence de concentration.
- IG proche de 1 : forte concentration.
- IG proche de 0 : faible concentration.

Section II : Les caractéristiques de forme

Les caractéristiques de forme permettent de préciser l'allure de la courbe des fréquences sans avoir besoin de la tracer. On distingue deux mesures de la forme d'une distribution : l'asymétrie et l'aplatissement.

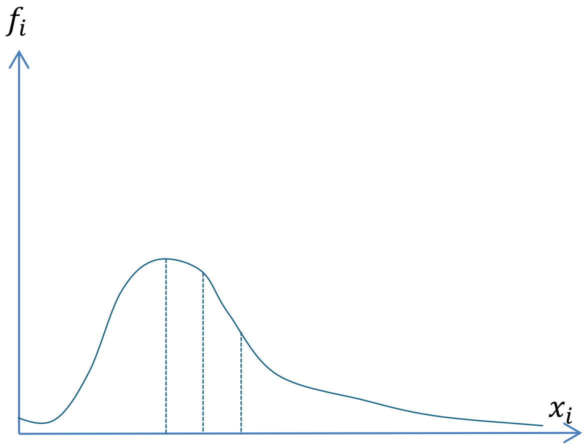
II.1-La mesure de l'asymétrie

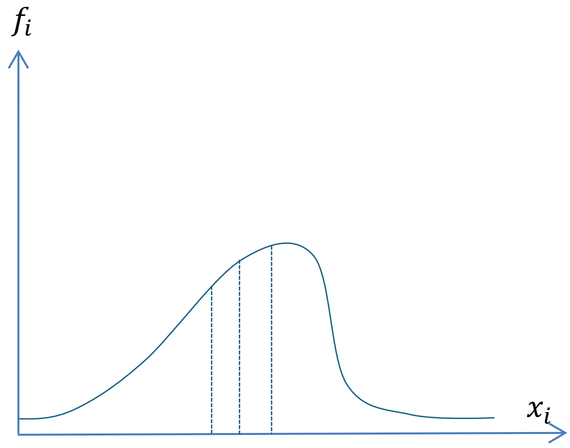
1.1-Définition :

Une distribution est dite symétrique si les observations repérées par leurs fréquences sont également dispersés de part et d'autre d'une valeur centrale : le mode, la médiane, et la moyenne arithmétique.

Dans une distribution symétrique, les trois caractéristiques de tendance centrale : le mode, la médiane, et la moyenne arithmétique sont confondues.

Par contre, une courbe non symétrique est dite oblique. Il existe deux cas de figure : oblique à gauche ou étalée vers la droite, oblique à droite ou étalée vers la gauche. Ces trois cas sont représentés dans les figures ci-après :





1.2-Les coefficients d'asymétrie :

Un coefficient est un nombre sans dimension, sans unité de mesure, permettant les comparaisons. Ces coefficients ne sont généralement valables que si la distribution contient un nombre assez élevé d'observations et qu'elle ne présente pas plusieurs modes. En d'autres termes, la série doit être uni-modale.

On distingue trois coefficients connus par les noms de leurs auteurs : Yule, Pearson, et Fisher.

1.2.1-Le coefficient de Yule :

1.2.2-Le coefficient de Pearson :

Une distribution est dite symétrique si les observations repérées par leurs fréquences sont également dispersés de part et d'autre d'une valeur centrale : le mode, la médiane, et la moyenne arithmétique.

Il se calcule à l'aide de la formule suivante :

$$\boxed{S = \frac{\bar{X} - M_O}{\sigma}}$$

Trois cas de figure peuvent se présenter :

- Si S=0, la courbe est symétrique.
- Si S>0, la courbe est oblique à gauche.
- Si S<0, la courbe est oblique à droite.

1.2.3-Le coefficient de Fisher :

Fisher a utilisé le coefficient de Pearson et l'a comparé à la courbe normale ; c'est-à-dire de part et d'autre de P = 3.

$$\boxed{\gamma = \rho - 3 = \frac{\mu 4}{\sigma^4}}$$

Trois cas de figure sont à distinguer :

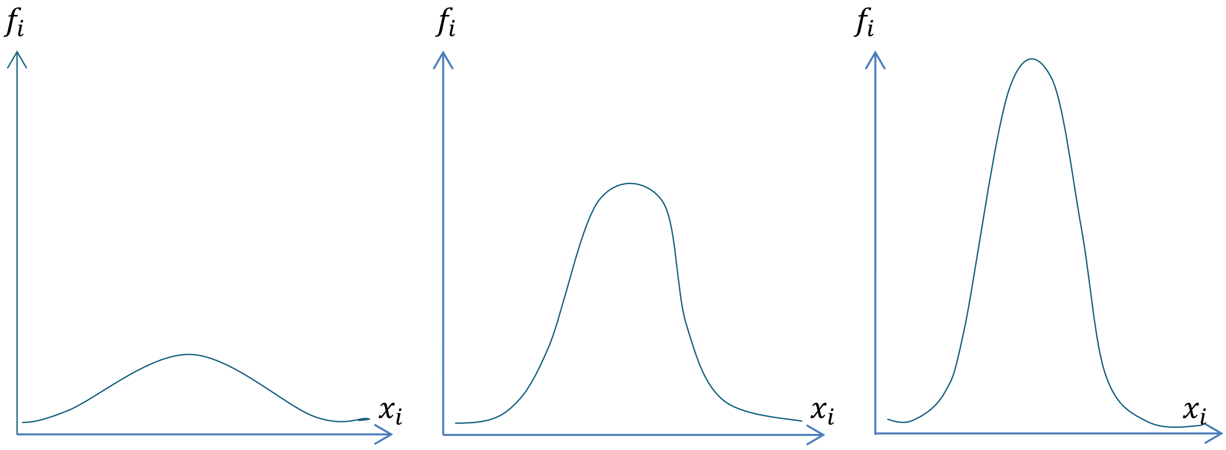
- $\gamma > 0 \rightarrow \rho > 3$ →distribution leptocurtique ou non aplatie.
- $\gamma < 0 \rightarrow \rho < 3$ →distribution platicurtique ou aplatie.
- $\gamma = 0 \rightarrow \rho = 3$ →distribution normale ou mesocurtique.

II.2-La mesure de l'aplatissement :

2.1-Définition :

Une distribution est dite aplatie si une forte variation de la variable entraine une faible variation de la fréquence relative et inversement. On considère qu'une courbe des fréquences est plus ou moins aplatie, par référence à la courbe des fréquences de la loi « Normale », loi de Gauss-Laplace.

On distingue les cas élucidés dans les figures ci-dessous :



Représentation schématique des trois types d'asymétrie

Distribution symétrique ($\gamma_1 = 0$) :

La courbe est en forme de cloche parfaite. Mo = Me = $\bar{x}$. La distribution normale (loi de Gauss) en est l'exemple type.

Asymétrie positive ($\gamma_1 > 0$) :

La queue de la distribution s'étend vers la droite. Le pic (mode) est à gauche de la médiane, elle-même à gauche de la moyenne. Ex : distribution des revenus, durée de vie de produits.

Asymétrie négative ($\gamma_1 < 0$) :

La queue de la distribution s'étend vers la gauche. Le pic est à droite de la médiane, elle-même à droite de la moyenne. Ex : âge à la retraite, notes à un examen facile.

2.1.1-Les coefficients d'aplatissement :

On trouve le coefficient de Pearson et le coefficient de Fisher. Ici, nous allons retenir le coefficient de Pearson (β_2) qui se calcule par la formule ci-après :

$$\beta_2 = \frac{\mu 4}{\sigma^4}$$

Avec :

- $\mu 4$: moment centré d'ordre 4. $\mu 4 = \frac{1}{N} \sum n_i (x_i - \bar{X})^4$

μ_2 = U (x), ∂_r^4 = [U (x)]^2

Trois cas peuvent se présenter :

- Si β_2 = 3, la courbe est normale.
- Si $\beta_2 > 3$ 3, la courbe est leptocurtique.
- Si β_2 = 3, la courbe est platicurtique.

Exercice d'application : Le tableau suivant présente la répartition du chiffre d'affaires d'un certain nombre d'entreprises

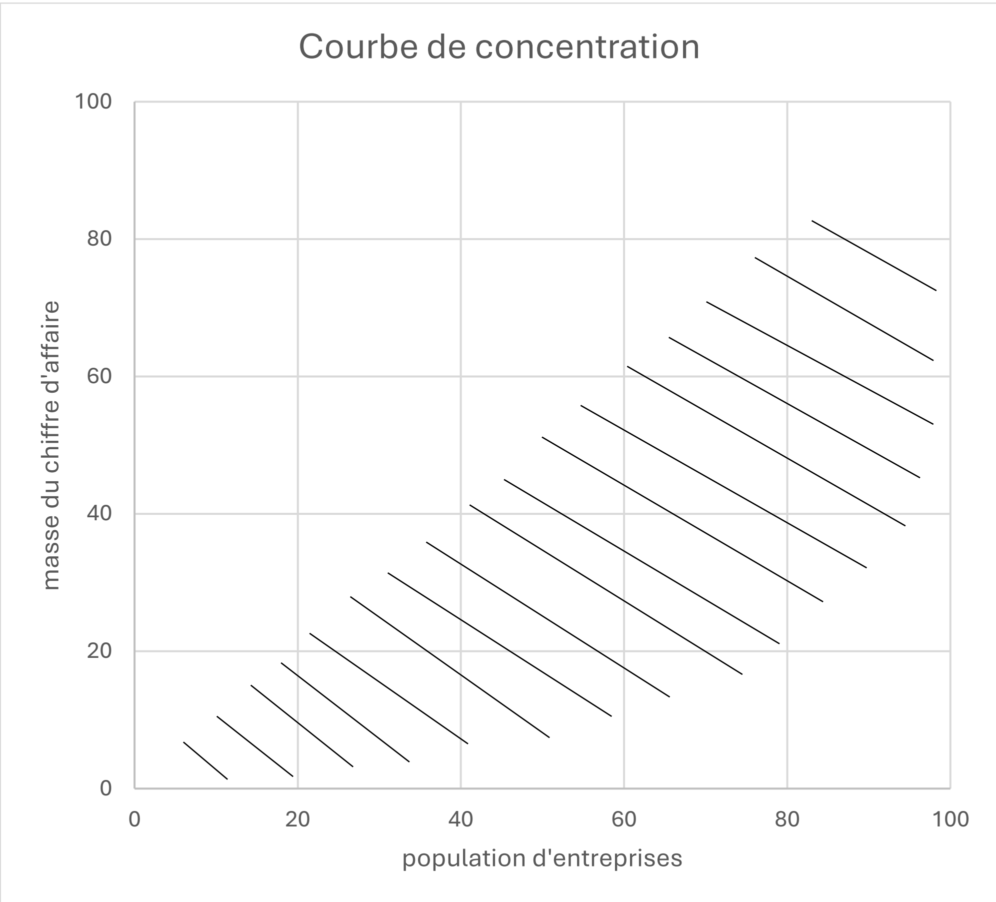
CA (10^3DA)	Nombre d'entreprises (n_i)	%
100 -500	51	23,83
500-1000	60	28,04
1000-1500	38	17,76
1500-2000	15	7,01
2000-5000	32	14,95
5000-10000	11	5,14
10000-100000	7	3,27

Total	214	100

Question : Analyser la concentration à l'aide de la courbe.

Réponse : il faut d'abord construire le tableau ci-après avant de tracer la courbe de concentration

CA (10^3DA)	Nombre d'entreprises (n_i)	%	% cumulé	Centre de classe (x_i)	$x_i * n_i$	%	% cumulé
100 -500	51	23,83	23,83	300	15300	2,14	2,14
500-1000	60	28,04	51,87	750	45000	6,31	8,45
1000-1500	38	17,76	69,63	1250	47500	6,65	15,10
1500-2000	15	7,01	76,64	1750	26250	3,68	18,78
2000-5000	32	14,95	91,59	3500	112000	15,70	34,48
5000-10000	11	5,14	96,73	7500	82500	11,56	46,04
10000-100000	7	3,27	100	55000	385000	53,96	100
Total	214	100	/		713550		



Exercice 1: Le tableau suivant donne le chiffre d'affaires (CA) de 80 entreprises ayant participé à une enquête de publicité :

CA en 10^6 DA	[0-5[	[5-20[	[20-100[	[100-500[	[500-1000[
Effectifs	48	16	08	04	04

1) Calculer $\Delta M/e$. Que pensez-vous de la concentration du chiffre d'affaires ?
2) Tracer la courbe de concentration.
3) Calculer l'indice de GINI.

CHAPITRE VI
LES INDICES DE LA VIE ECONOMIQUE

I- LA NOTION D'INDICE

L'indice est un nombre (supérieur à 0) qui a pour but de traduire l'évolution dans le temps et dans l'espace d'une grandeur mesurable :" évolution des prix, des cours de la bourse, d'une production, l'évolution en termes d'emplois (indice du chômage), etc.

La notion d'indices statistiques apparaît chaque fois que l'on doit comparer deux grandeurs dans l'espace ou dans le temps à travers leur rapport.

Exemple : une entreprise a payé à l'époque t_0 une matière première "x", 30 DA le kg; trois ans plus tard (en t_3) elle la paye 36 DA le kg.

Ces deux prix peuvent être comparés selon trois procédés :

a) soit en faisant leur différence. En trois ans, le prix de la matière première "x" a augmenté en **valeur absolue** de 36 - 30 = 6 DA le kg. On notera que ce procédé ne donne qu'une information imparfaite sur cette augmentation. En effet, elle ne tient pas compte de l'importance du prix à l'instant t_o. En d'autres termes, nous aurions eu le même résultat avec des prix différents, par exemple 23 DA et 29DA le kg.

b) soit en estimant l'accroissement de prix en **valeur relative** :

(36-30) /30= 0,2 ou 0,2 x 100 = 20 %.

c) soit en faisant le rapport suivant : donnant un " Indice" :

(36/30)100 = 120

Nous avons pris comme base de calcul le nombre "100" ; par convention, nous prendrons une base "100" pour les calculs d'indices.

Ce rapport « $\frac{36}{30}$x100 » soit « 120 » est appelé **" indice simple ", ou "indice élémentaire".**

II- LES INDICES ELEMNETAIRES.

1) Définition et calcul

L'indice élémentaire est un nombre sans unité (sans dimension), qui traduit l'évolution dans le temps d'une grandeur mesurable.

Si l'on appelle « V_0 » la valeur d'une grandeur à l'époque 0 (date de référence ou date de base)

Si l'on appelle « V_t » la valeur d'une grandeur à l'époque t (date courante ou date finale)

L'indice de la grandeur à l'époque t par rapport à l'époque 0 est un nombre, sans unité, tel que : $I_{t/0} = \frac{Vt}{V0}$ x100

- Si $I_{t/0} > 100$, il y a une augmentation entre les époques 0 et t.
- Si $I_{t/0} < 100$, il y a une baisse entre les époques 0 et t

Exemple 1 : Le prix du café était de 21,95 DA le Kg en 1970 année de référence, et de 135 DA le Kg en 1995, année de comparaison.

$I_{1995/1970} = \frac{V_{1995}}{V_{1970}}$ x100 = $\frac{135}{21,95}$ x100 ≈ 615 ou 6,15.

On dit que l'indice simple du café en 1995 est de 615 en base 100 en 1970 (ou 6,15).

Remarques :

1°) la variable est ici "chronologique" (le temps). Elle peut, dans d'autres cas, être spatiale et prendre en compte le changement de lieu.

2°) On retiendra que le taux d'accroissement (taux d'évolution) est égal à l'indice moins la base (100) et qu'il s'exprime en pourcentage. Soit 615-100= 515 (le prix du café a crû de 515% entre 1970 et 1995.

3°) Le prix du café est multiplié par 6,15 entre 1970 et 1995.

Exemple d'établissement d'un indice :

Enoncé :

A une date donnée (t_1), un objet était vendu 3020 DA.

Un objet de même type vaut 2490 DA, 8 ans plus tard (t_2). Donner l'indice d'évolution du prix (diminution ou augmentation)

Réponse :

Si l'on fait la division du prix payé 2490 DA en t_2 par 3020 DA. En t_1 on trouve un nombre : (2490/3020) = 0,8245

Ce nombre de « 0,8245 » s'appelle « coefficient multiplicateur » associé à la « diminution ».

La diminution est de 0,8245 - 1 = - 0,1755

Cette diminution peut s'exprimer en pourcentage.

Pour ce faire on multipliera le nombre par 100; ainsi le taux de diminution est de (0,1755 x 100)= 17,55 ; soit 17,55 %

2- Propriétés des indices élémentaires

- ***Identité* :** $I_{t/t=} I_{0/0}$= 1 ou 100
- ***Réversibilité*** (entre les dates notées 0 et 1) : $I_{1/0} = 1 / I_{0/1}$

Exemple *:*

Si un prix augmente de 20% de 2010 à 2012, que dire de son évolution de 2012 à 2010 ?

Réponse :

On a $I_{12/10}$ = 120 et on cherche $I_{10/12}$. Par la propriété de réversibilité.

$I_{10/12}$ = 1/ $I_{12/10}$ = 1/ 1,20= 0.83. Cela signifie que le prix de 2010 est inférieur à celui de 2012 de 17% [(0,83-1) ×100]

- ***Circularité*** (ou transitivité) : $I_{t/0} = I_{t/t-1}$ x $I_{t-1/t-2}$ x …x $I_{1/0}$

On parle dans ce cas du principe de l'enchaînement multiplicateur des indices.

Exemple :

Le chiffre d'affaires (CA) d'une entreprise a augmenté de 30% de 2010 à 2011 et diminué de 15% de 2011 à 2012. Le CA a-t-il diminué ou augmenté de 2010 à 2012 ?

Réponse :

On a $I_{11/10}$ = 130, $I_{12/11}$ = 85 et on cherche $I_{12/10}$ par la propriété de circularité

$I_{12/10} = I_{12/11} \times I_{11/10}$= 0.85x 1.30= 1,105

Autrement dit, le CA a augmenté de [(1,105-1) x100] = 10,5 % de 2010 à 2012.

- **Si valeur** = quantité x prix, alors Iv = Iq x Ip

Exemple :

Soit P et Q les prix et quantités d'un produit vendu par une entreprise. Si le prix de ce produit ont augmenté de 60% de 2000 à 2010 et si les quantités vendues ont diminué de 50% de 2000 à 2010, quelle est l'´évolution des recettes de 2000 à 2010 ?

Réponse :

$I^{v}_{10/00} = I^{P}_{10/00} \times I^{Q}_{10/00}$ =1,60x0, 50= 0.8 soit une baisse des recettes (valeur) de 20%

III-LES INDICES COMPLEXES, SYNTHETIQUES OU COMPOSES.

1) Notion d'indice synthétique :

Les indices simples, que nous avons définis, ne prennent en compte que la variation d'une seule composante d'un phénomène, (exemple le prix de la pomme de terre). Or il est très rare qu'une composante explique, à elle seule, l'évolution d'un phénomène économique complexe : (par exemple l'évolution du prix d'un sandwich n'est pas seulement la conséquence de la variation du prix de la pomme de terre, mais aussi celle des variations de prix d'autres produits et services) Il faut donc disposer d'outils statistiques permettant de tenir compte de plusieurs composantes.

On fera appel aux indices synthétiques, ces indices permettent de tenir compte de plusieurs composantes, en combinant l'évolution simultanée de plusieurs phénomènes, chacun d'eux étant mesuré par un indice simple.

Pour pouvoir déterminer l'influence des prix ou des quantités. On recourt aux indices de Laspeyres ou de Paasche : l'indice de Laspeyres des prix, par exemple, mesure l´évolution des prix de 0 à t en supposant que les quantités sont fixées à la date 0. L'indice de Paasche des quantités, par exemple, mesure l´évolution des quantités de 0 à t en supposant que les prix sont fixés à la date t.

2) Coefficients de pondération ou coefficient budgétaires :

On appelle coefficient budgétaire la part de la valeur du produit à une époque donnée par rapport à la valeur de l'ensemble des produits considérés à la même époque.

✓ **Coefficient budgétaire à l'époque initiale (C^j_0 ou α^j_0)**

$$\alpha^j_0=\frac{P_{j.0}\ .\ Q_{j.0}}{\Sigma\, P_{j.0}\ .\ Q_{j.0}} \quad \text{ou} \quad \alpha^j_0=\frac{P_0\ .\ Q_0}{\Sigma\, P_0\ .\ Q_0}$$

Avec : P_{j0} : Prix du produit j à l'époque 0 (base ou de référence)

Q_{j0} : Quantité du produit j à l'époque 0.

✓ **Coefficient budgétaire à la période courante (C^j_t ou α^j_t)**

$$\alpha^j_t=\frac{P_{j.t}\ .\ Q_{j.t}}{\Sigma\, P_{j.t}\ .\ Q_{j.t}} \quad \text{ou} \quad \alpha^j_t=\frac{P_t\ .\ Q_t}{\Sigma\, P_t\ .\ Q_t}$$

3) <u>La méthode (Indice) de Laspeyres</u> :

L'indice de Laspeyres des prix (ou des quantités)est égal à la moyenne arithmétique des indices élémentaires des prix (ou des quantités) pondérés par les coefficients budgétaires de la période de base (voir : formule de définition). L'indice de Laspeyres peut aussi se calculer par la formule simplifiée en considérant la période initiale (0) comme date de référence.

	Indice des Prix	Indice des Quantités
Formule de Définition	$L^P_{t/0}=\Sigma\alpha_{j.0}\ .\ I^P_{t/0}$	$L^Q_{t/0}=\Sigma\alpha_{j.0}\ .\ I^Q_{t/0}$
Formule Simplifiée	$L^P_{t/0}=\frac{\Sigma P_t.Q_0}{\Sigma P_0.Q_0}$	$L^Q_{t/0}=\frac{\Sigma P_0.Q_t}{\Sigma P_0.Q_0}$

Exemple :

Etudions le panier de trois produits d'une entreprise de transport de voyageurs vendus à une population donnée, en 2010 et 2012.

	2010		2012	
Produit	Prix	Quantités	Prix	Quantités
Ticket simple	1.5	100	2	100
Abonnement hebdomadaire	35	40	50	50
Abonnement mensuel	130	35	175	40

Question :

Calculer les indices de Laspeyres des prix et des quantités par la formule simplifiée.

$L^{P}_{12/10} = \frac{(2\times 100)+(50\times 40)+(175\times 35)}{1.5\times 100+35\times 40+130\times 35}$x100= 136,48

Les prix ont augmenté de 36.48% de 2010 à 2012.

$L^{Q}_{12/10} = \frac{(1.5\times 100)+(35\times 50)+(130\times 40)}{(1.5\times 100)+(35\times 40)+(130\times 35)}\times 100$ =116.39.

Les quantités vendues ont augmenté de 16.39% de 2010 à 2012.

<u>4) La méthode (Indice) de Paasche</u> :

L'indice de Paasche des prix (ou des quantités) est égal à la moyenne harmonique des indices élémentaires des prix (ou des quantités) pondérés par les coefficients budgétaires de la période courante (voir : formule de définition). L'indice de Paasche peut aussi se calculer par la formule simplifiée en considérant la période courante (t) comme date de référence.

	Indice des Prix	Indice des Quantités
Formule de Définition	$P^{P}_{t/0}=\frac{1}{\Sigma\alpha_{j.t}\cdot\frac{1}{IPt/0}}$	$P^{Q}_{t/0}=\frac{1}{\Sigma\alpha_{j.t}\cdot\frac{1}{IQt/0}}$
Formule Simplifiée	$P^{P}_{t/0}=\frac{\Sigma P_t.Q_t}{\Sigma P_0.Q_t}$	$P^{Q}_{t/0}=\frac{\Sigma P_t.Q_t}{\Sigma P_t.Q_0}$

Exemple :

Calculer les indices de Paasche des prix et des quantités sur l'exemple précédent par la formule simplifiée.

$P^{P}_{12/10} = \frac{2\times 100+50\times 50+175\times 40}{1.5\times 100+35\times 50+130\times 40}\times 100 = 136.62.$ Soit une hausse des prix de 36.62% entre 2010 et 2012.

$P^{Q}_{12/10} = \frac{2\times 100+50\times 50+175\times 40}{2\times 100+50\times 40+175\times 35}\times 100 = 116.52.$ (hausse des quantités de 16.52%)

5) L'indice de Fisher :

L'indice de Fisher est la moyenne géométrique simple des indices de Laspeyres et de et de Paasche

Ainsi, Fisher des prix est égale au produit de la racine carrée de Laspeyres et de Paasche des prix **$F^P = \sqrt{L^P . P^P}$**

De même,Fisher des quantités est égale au produit de la racine carrée de Laspeyres et de Paasche des quantités **$F^Q = \sqrt{L^Q . P^Q}$**

***6) L'indice de valeur* globale ($I^{VG}_{t/0}$)**

Nous avons calculé, précédemment, les indices de prix et de quantités de Laspeyres, de Paasche et de Fisher, avec des formules différentes. Par contre, les indices de valeur globales de Laspeyres, de Paasche et de Fisher se calculent tous de la même manière. Autrement dit

$$I^{VG}_{t/0} = \frac{\Sigma P_t . Q_t}{\Sigma P_0 . Q_0} \times 100$$

Remarque :

On peut aussi calculer l'indice des valeurs globales comme suit :

$I^{VG}_{t/0} = L^P_{t/0} . P^Q_{t/0} = L^Q_{t/0} . P^P_{t/0}$

Exercice d'application :

Un responsable d'approvisionnement de rayon a relevé, au cours de deux années, les quantités et les prix de trois produits « A », « B » et « C » et a établi le tableau suivant :

Année	**Articles**	**Quantités (kg)**	**Prix (DA)**	**Valeurs. (Quantité x Prix)**
2000	A	15	10	150
	B	20	5	100
	C	25	8	200
2005	A	25	12	300
	B	25	6	150
	C	35	15	525

Questions :

1. Calculer les indices des prix de Laspeyres et de Paasche par les formules de définition et par les formules simplifiées.
2. Calculer les indices des quantités de Laspeyres et de Paasche par les formules de définition et par les formules simplifiées.
3. Calculer les indices de Fisher des prix et quantités

4. Calculer l'indice de valeur globales et vérifier que $I^{VG}_{t/0}=L^{P}_{t/0} . P^{Q}_{t/0} = L^{Q}_{t/0} . P^{P}_{t/0}$

I) CALCUL D'INDICES DES PRIX

	P_0	Q_0	P_t	Q_t	$P_0.Q_0$	C^j_0	$P_0.Q_t$	$P_t.Q_0$	$P_t.Q_t$	α^j_t	$I^P_{t/0}$	$I^P_{0/t}$	$I^Q_{t/0}$	$I^Q_{0/t}$
A	10	15	12	25	150	0,33	250	180	300	0,31	1,2	0,83	1,66	0,60
B	5	20	6	25	100	0,22	125	120	150	0,15	1,2	0,83	1,25	0,80
C	8	25	15	35	200	0,45	280	375	525	0,54	1,87	0,53	1,4	0,71
∑	/	/	/	/	450	1	655	675	975	1	/	/	/	/

1) ***Formule de définition de Laspeyres des prix***

$L^{P}_{t/0}=\Sigma\alpha_{j.0} . I^{P}_{t/0} \Rightarrow L^{P}_{2005/2000} = (0{,}33.1, 2) +(0{,}22.1{,}2) +(0{,}45.1{,}87) \approx 1.5$ ou 150

2. ***Formule simplifiée***

$$L^{P}_{t/0}=\frac{\Sigma P_t.Q_0}{\Sigma P_0.Q_0} \Rightarrow L^{P}_{2005/2000} = \frac{675}{450} = 1{,}5 \text{ ou } 150$$

Soit, selon la formule de Laspeyres, une augmentation des prix des trois (3) produits de 50%, entre 2000 et 2005, toute chose étant égale par ailleurs.

3) ***Indice de Paasche des prix (formule de définition)***

$$P^{P}_{t/0}=\frac{1}{\Sigma\alpha_{j.t} . \frac{1}{IPt/0}} \Rightarrow P^{P}_{2005/2000} = \frac{1}{(0{,}31.0{,}83)+(0.15.0{,}83)+(0{,}54.0{,}53)} \approx 1.49 \text{ ou } 149$$

4) ***Formule simplifiée de Paasche des prix***

$$P^{P}_{t/0}=\frac{\Sigma P_t.Q_t}{\Sigma P_0.Q_t} \Rightarrow P^{P}_{2005/2000} = \frac{975}{655} \approx 1.49 \text{ ou } 149$$

Soit, selon la formule de Paasche, une augmentation des prix des trois (3) produits, de près de 49%, entre 2000 et 2005, toute chose étant égale par ailleurs.

II) INDICES DES QUANTITES

1) ***Laspeyres des quantités*** (formule de définition)

$L^{Q}_{t/0}=\Sigma\alpha_{j.0} . I^{Q}_{t/0} \Rightarrow L^{Q}_{2005/2000}= (0{,}33.1, 66) + (0{,}22.1, 25) + (0{,}45.1, 4) = 1{,}45$ ou 145

2) ***Laspeyres des quantités*** (formule simplifiée)

$$L^{Q}_{t/0}=\frac{\Sigma P_0.Q_t}{\Sigma P_0.Q_0} \Rightarrow L^{Q}_{2005/2000}=\frac{655}{450} = 1.45 \text{ ou } 145$$

Soit, selon la formule de Laspeyres, une augmentation des quantités demandées des trois (3) produits de 45%, entre 2000 et 2005, toute chose étant égale par ailleurs.

3) ***Paasche des quantités*** (formule de définition)

$$P^{Q}_{t/0}=\frac{1}{\Sigma \alpha_{j.t} \cdot \frac{1}{IQt/0}} \Rightarrow P^{Q}_{2005/2000}=\frac{1}{(0{,}31.0{,}60)+(0{,}15.0{,}80)+(0{,}54.0{,}71)}=1{,}44 \text{ ou } 144$$

4) ***Paasche des quantités*** (formule simplifiée)

$$P^{Q}_{t/0}=\frac{\Sigma P_t.Q_t}{\Sigma P_t.Q_0} \Rightarrow P^{Q}_{2005/2000}=\frac{975}{675}=1{,}44 \text{ ou } 144$$

Soit, selon la formule de Paasche, une augmentation des quantités demandées des trois (3) produitsde 44%, entre 2000 et 2005, toute chose étant égale par ailleurs.

III) INDICE DE FISHER

1) ***Fisher des prix***

$$F^{P}=\sqrt{L^{P}.P^{P}} \Rightarrow F^{P}=\sqrt{1{,}5.1{,}49}=1{,}495 \text{ ou } 149{,}5$$

Soit, selon la formule de Fisher, une augmentation des prix des trois (3) produits, de près de 49%, entre 2000 et 2005, toute chose étant égale par ailleurs.

2) ***Fisher des quantités***

$$F^{Q}=\sqrt{L^{Q}.P^{Q}} \Rightarrow F^{Q}=\sqrt{1{,}45.1.44}=1{,}445 \text{ ou } 144{,}5$$

Soit, selon la formule de Fisher, une augmentation des quantités demandées des trois (3) produits de 44,5%, entre 2000 et 2005, toute chose étant égale par ailleurs.

IV) INDICE DE VALEUR GLOBALES

$$I^{VG}_{t/0}=\frac{\Sigma P_t.Q_t}{\Sigma P_0.Q_0}=\frac{975}{450}=2{,}16$$

On vérifie par ailleurs que :

$L^{P}_{t/0} . P^{Q}_{t/0} = L^{Q}_{t/0} . P^{P}_{t/0} = I^{VG}_{t/0} \Rightarrow (1{,}5.\ 1{,}44)= (1{,}49.\ 1{,}45) = 2{,}16$

Soit, une augmentation de la valeur (dépenses) des trois (3) produits de 116%, entre 2000 et 2005, toute chose étant égale par ailleurs.

<u>Série de T.D 6</u>

Exercice n° 1: Le tableau suivant indique les prix unitaires d'un produit au cours de la période 2010-201 3:

Année	2010	2011	2012	2013
Prix (DA)	120	130	125	135

1- Calculer les indices élémentaires des prix pour l'année t en base 100 l'année t-1. Interpréter les résultats.

2-Calculer l'indice des prix sur l'ensemble de la période (2010-2013).

3- Sur la période (2005 - 2013) les prix ont été multipliés par 2. Calculer l'indice des prix de la période (2005 – 2010). En déduire l'indice des prix « $I_{2005/2010}$ ».

Exercice n° 2 : Dans un magasin spécialisé on a connu les deux situations suivantes :

1- Pendant une saison, les ventes d'un produit ont cru de 6% alors que la quantité vendue a diminué de 4%. Quelle est la variation du prix du produit ?

2- Pendant une autre saison, la quantité vendue d'un produit a diminué de 6% alors que le prix du produit a augmenté de 11%. Quelle est la variation des ventes du produit ?

Exercice n° 3 : Calculer l'indice du pouvoir d'achat pour un individu dans les situations suivantes :

- si les salaires augmentent de 10% ;
- si les prix diminuent de 34% ;
- si le salaire croit de 10% tandis que les prix augmentent de 15%.

Exercice n° 4: Un salarié gagnait 10 300 DA en 1996, 12 500 DA en 2000, 17 000 DA en 2004 et 19 000 DA en 2007.

On vous donne les indices des prix suivants :

$I_{2000/1996}$=1,10 ; $I_{2004/2000}$= 1,26 ; $I_{2007/2004}$= 1,50.

1- Exprimer les différents salaires en dinar constant de 2007.

2-Déterminer l'évolution réelles de ces salaires durant les périodes suivantes : 1996-2000, 2000-2004, 2004-2007, 1996-2007.

Exercice 5 : La structure du budget des ménages à bas revenus et les indices des prix sont donnés par le tableau ci-dessous :

Rubriques	Structure du budget en 2010	Indices des prix		
		$I_{2009/2008}$	$I_{2010/2009}$	$I_{2011/2010}$
Habitation	32,6	106,6	105,5	104,4
Alimentation	27,5	101,4	102,4	103,2
Transport	12,6	101,8	100,3	101,4
Habillement	07,6	105,0	103,2	103,0
Culture-loisirs	07,5	104,1	101,5	101,2
Santé	06,8	102,5	103,6	104,3
Vacances	01,5	103,9	104,4	102,3
Divers	03,9	101,9	102,1	101,0

1- Déterminer l'évolution du coût de la vie pour l'année 2010 en base 100 l'année 2009. Expliquer le choix de la formule utilisée.

2-Calculer l'indice du coût de la vie pour l'année 2011 base 100 l'année 2010. Justifier le choix de la formule.

Sachant que pour l'année 2010 le budget par l'unité de consommation est de 360 000 DA/an, quelles devraient être les dépenses pour l'année 2011 pour que les ménages maintiennent leur pouvoir d'achat.

Exercice 6 : Un institut national de statistique d'un pays imaginaire veut déterminer les indices synthétiques des prix, des quantités et des valeurs globales de quatre (4) groupes de service A,B,C,D. Les enquêteurs fournissent les résultats suivants :

Services	Prix unitaire en 1980	Prix unitaire en 1990	Quantités consommées en unités physiques en 1980	Quantités consommées en unités physiques en 1990
A	180	700	220	176
B	100	365	240	120
C	2020	3040	50	52
D	30	60	500	500

1 -Calculer l'indice de LASPEYRES des prix, des quantités en utilisant les formules simplifiées.

2-Calculer l'indice de PAASCHE des prix, des quantités en utilisant les formules simplifiées.

3-Calculer l'indice de FISHER des prix et des quantités.

4-Calculer l'indice des valeurs globales.

CORRIGE DE LA SERIE

Exercice 1 :

1/Les indices élémentaires des prix pour l'année t en base 100 l'année t-1 :

On sait que $I^P_{t/0} = \frac{p_t}{p_0}$

$I^P_{2011/2010} = \frac{p_{2011}}{p2010} = \frac{130}{120} = 1{,}083$ ou 108,3 ; $I^P_{2012/2011} = \frac{p_{2012}}{p2011} = \frac{125}{130} = 0{,}961$ ou 96,1 ;

$I^P_{2013/2012} = \frac{p_{2013}}{p2012} = \frac{135}{125} = 1{,}08$ ou 108.

Interprétation :

- *Par les indices* : le prix du produit a été multiplié par 1,083 de 2010 à 2011, par 0,961 de 2011 à 2012 et par 1,08 de 2012 à 2013
- *En termes du taux d'évolution (taux de croissance)* : le prix a augmenté de 8,3% entre 2010 et 2011 ; le prix a baissé de 3,9% entre 2011 et 2012 et a augmenté de 8% entre 2012 et 2013 [(1,08-1) ×100= 8%].

2/ Calcul de l'indice des prix sur l'ensemble de la période (2010-2013) :

$I^P_{2013/2010} = I^P_{2013/2012} \times I^P_{2012/2011} \times I^P_{2011/2010}$ (propriété de la circularité).

$I^P_{2013/2010} = 1{,}08 \times 0{,}961 \times 1{,}083 = 1{,}124$.

Cela veut dire que le prix de ce produit a été multiplié par 1,124 de 2010 à 2013 ou bien le prix a augmenté de 12,4% de 2010 à 2013.

3/ Calcul de l'indice des prix de la période 2005-2010 :
a/ Nous avons $I^P_{2013/2005} = 2$; $I^P_{2010/2005} = ?$
Pour calculer $I^P_{2010/2005}$, nous allons utiliser la propriété de la circularité

$$I^P_{2013/2005} = I^P_{2013/2010} \times I^P_{2010/2005} \Longrightarrow I^P_{2010/2005} = \frac{Ip2013/2005}{Ip2013/2010} = \frac{2}{1{,}124} = 1,7793.$$

Cela veut dire que le prix de ce produit a augmenté de [(1,7793-1) ×100] = 77,93% entre 2005 et 2010.
b/ En déduire $I^P_{2005/2010}$: nous allons utiliser la propriété de la réversibilité

$$I^P_{2005/2010} = \frac{1}{Ip2010/2005} = \frac{1}{1{,}7793} = 0{,}562.$$

Exercice 2 :
1/ Les ventes ont augmenté de 6% (0,06) cela veut dire que $I^{Valeur} = 1+0{,}06 = 1{,}06$ ou $I^V = 100+6 = 106$.
La quantité vendue a diminué de 4% (0,04) $\Rightarrow I^Q = 1- 0{,}04 = 0{,}96$ ou 96.
Nous avons $V=P\times Q \Rightarrow I^V=I^P\times I^Q \Rightarrow I^P = \frac{I^V}{I^Q} = \frac{1{,}06}{0{,}96} = 1{,}1041$.
Cela veut dire que le prix du produit a augmenté de 10,41%.

2/ La quantité a diminué de 6% (-0,06) $\Rightarrow I^Q = 1-0{,}06 = 0{,}94$
Le prix a augmenté de 11% $\Rightarrow I^P = 1+ 0{,}11 = 1{,}11$.
Pour connaitre la variation des ventes de ce produit, on calcule : $I^V=I^P\times I^Q$
$I^V = 0{,}94\times 1{,}11 = 1{,}0434$,
Cela veut dire que les ventes de ce produit ont augmenté de [(1,0434-1) ×100] = 4%.

Exercice 3 : Calcul de l'indice du pouvoir d'achat (I^{PA}) :
-Si les salaires augmentent de 10% $\Rightarrow I^{PA} = 1+0{,}10 = 1{,}10$ ($I^S = I^{PA}$) car les prix n'ont pas changé).
-Si les prix diminuent de 34% $\Rightarrow I^P = 1- 0{,}34 = 0{,}66$; $I^{PA} = \frac{1}{I^p_{t/0}} = \frac{1}{0{,}66} = 1{,}5151$; donc le pouvoir d'achat augmente de 51,51%.

-Si le salaire augmente de 10% et les prix augmentent de 15% $\Rightarrow I^S = 1+0{,}10 = 1{,}10$ et $I^P = 1+0{,}15 = 1{,}15$. $I^{PA} = \frac{I^S}{I^p} = \frac{1{,}10}{1{,}15} = 0{,}9565$ donc le pouvoir d'achat baisse de (0, 9565-1) ×100 = 4, 34%.

Exercice 4: On a: S_{1996}= 10300DA; S_{2000}= 12500DA; S_{2004}= 17000DA; S_{2007}= 19000DA
$I^P_{2000/1996} = 1,10$; $I^P_{2004/2000} = 1,26$; $I^P_{2007/2004} = 1,50$
1/ *Les différents salaires en dinars constant de 2007*:
En raison de l'inflation, les salaires mensuels, exprimés en dinars courants des périodes où ils ont été perçus, ne sont pas directement comparables. On peut les convertir en dinars constant d'une même période, par exemple 2007, c'est-à -dire qu'on détermine les salaires équivalents en DA de 2007 (au sens de l'évolution générale des prix). Cette opération s'appelle « la déflation » de la série des salaires.
Ainsi

-les 19000DA de salaire mensuel de 2007 sont exprimés en DA constant de 2007.

-Les 17000DA de 2004 sont en dinars courants de 2004 et sont équivalents à

17000×$I^p_{2007/2004}$= 17000×1,50 = 25500DA de 2007.

-Les 12500DA de 2000 sont en dinars courants de 2000 et sont équivalents à

12500×$I^p_{2007/2004}\times I^P_{2004/2000}$= 12500×1,5×1,26= 23625DA de 2007.

-Les 10300DA de 1996 sont en dinars courants de 1996 et sont équivalents à

10300× $I^p_{2007/2004}\times I^P_{2004/2000} \times I^P_{2000/1996}$= 10300×1,50×1,26×1,10= 21413,7 DA de 2007.

2/ L'évolution réelle des salaires : il faut calculer l'indice des salaires.

a/Période 1996-2000 : $I^s_{2000/1996} = \frac{Salaire\ en\ 2000}{Salaire\ en\ 1996} = \frac{23625}{21413,7}$ = 1,1032> 1. Il s'agit d'une augmentation du salaire réel (pouvoir d'achat) de 10,32% de 1996 à 2000.

b/ Période 2000-2004 : $I^s_{2004/2000} = \frac{Salaire\ en\ 2004}{Salaire\ en\ 2000} = \frac{25500}{23625}$ = 1,0793. C'est une augmentation du salaire réel de 7,93% entre 2000 et 2004.

c/ Période 2004-2007 : $I^s_{2007/2004} = \frac{Salaire\ en\ 2007}{Salaire\ en\ 2004} = \frac{19000}{25500}$ = 0,7450<1. C'est une baisse du pouvoir d'achat de (0,7450-1) ×100 = -25,50%.

d/ Période 1996-2007 : $I^s_{2007/1996} = \frac{Salaire\ en\ 2007}{Salaire\ en\ 1996} = \frac{19000}{21413,7}$ = 0,8872. C'est une baisse du pouvoir d'achat de (0,8872-1) ×100 = 11,28%.

Exercice 5 :

1/L'évolution du cout de la vie pour l'année 2010 en base 100 l'année 2009 : Nous allons utiliser la formule de Paasche car nous disposons des coefficients budgétaires (structure du budget) de l'année 2010 qui est l'année courante (t).

$$P^p_{t/0} = \frac{1}{\sum \frac{\alpha^j_t}{I^p_{t/0}}}$$

$$P^p_{2010/2009} = \frac{1}{\frac{0,326}{1,055}+\frac{0,275}{1,024}+\frac{0,126}{1,003}+\frac{0,076}{1,032}+\frac{0,075}{1,015}+\frac{0,068}{1,036}+\frac{0,015}{1,044}+\frac{0,039}{1,021}} = \frac{1}{0,9679} = 1,031\ ou\ 103,1;$$

Soit une augmentation des prix de 3,1% de 2009 à 2010.

2/L'indice du cout de la vie pour l'année 2011 base 100 l'année 2010 :

Nous allons utiliser la formule de Laspeyres car nous avons les coefficients budgétaires de l'année 2010 qui représente l'année de base (0).

$L^p_{t/0} = \sum \alpha^j_0 \times I^p_{t/0} \Longrightarrow L^p_{2011/2010}$= (0,326×1,044)+(0,275×1,032)+ (0,126×1,014)+ (0,076×1,03)+ (0,075×1,012)+ (0,068×1,043)+ (0,015×1,023)+ (0,039×1,010).

$L^p_{2011/2010}$= 1,032, soit une augmentation des prix de (1,032-1) ×100 = 3,2%.

3/ Le budget de l'année 2010= 360000DA/an.

Pour que les ménages maintiennent leur pouvoir d'achat, le revenu doit augmenter de 3,2% donc il doit être multiplié par 1,032, ce qui donne 360000×1,032= 371520DA pour l'année 2011.

Exercice 5 :

Services	P_0	Pt	Q_0	Q_t	P_tQ_0	P_0Q_0	P_0Q_t	P_tQ_t
A	180	700	220	176	154000	39600	31680	123200
B	100	365	240	120	87600	24000	12000	43800
C	2020	3040	50	52	152000	101000	105040	158080
D	30	60	500	500	30000	15000	15000	30000
Total					423600	179600	163720	355080

1/L'indice de Laspeyres des prix ($L_{t/0}^{p}$) et des quantités ($L_{t/0}^{Q}$) avec les formules simplifiées.

A/ $L_{t/0}^{p} = \frac{\sum P_t.Q_0}{\sum P_0.Q_0} \times 100 = \frac{423600}{179600} \times 100 = 2{,}358 \times 100 = 235{,}8$.

Cela signifie que les prix ont augmenté de 135, 8%.

B/$L_{t/0}^{p} = \frac{\sum P_0.Q_t}{\sum P_0.Q_0} \times 100 = \frac{163720}{179600} \times 100 = 0{,}9115 \times 100 = 91{,}15$.

Cela signifie que les quantités consommées ont baissé de 8,9% [(0,9115-1) × 100].

2/Les indices de Paasche des prix ($P_{t/0}^{p}$) et des quantités ($P_{t/0}^{Q}$) :

$A\ /P_{t/0}^{p} = \frac{\sum P_t.Q_t}{\sum P_0.Q_t} \times 100 = \frac{355080}{163720} \times 100 = 2{,}1688$ ou 216,88, soit une hausse des prix 116,88%.

B/ $P_{t/0}^{Q} = \frac{\sum P_t.Q_t}{\sum P_t.Q_0} \times 100 = \frac{355080}{423600} \times 100 = 0{,}8382$ ou 83,82%, soit une baisse des quantités demandées de 16,18%.

3/ Les indices de Fisher des prix ($F_{t/0}^{p}$) et des quantités ($F_{t/0}^{Q}$) :

A/ $F_{t/0}^{p} = \sqrt{L_{t/0}^{p} \times P_{t/0}^{p}} = \sqrt{2{,}358 \times 2{,}1688} = 2{,}2614$ ou 226,14 ; soit une hausse des prix de 126,14%.

B/ $F_{t/0}^{Q} = \sqrt{L_{t/0}^{Q} \times P_{t/0}^{Q}} = \sqrt{0{,}9115 \times 0{,}8382} = 0{,}8741$ ou 87,41, soit une baisse des prix de 12,59%.

4/ L'indice des valeurs globales ($I_{t/0}^{VG}$) :

$I_{t/0}^{VG} = \frac{\sum P_t.Q_t}{\sum P_0.Q_0} \times 100 = \frac{355080}{179600} \times 100 = 1{,}9770$ ou 197,7 ; soit une hausse de 97,7% des valeurs globales (dépense totale).

CHAPITRE VII

LES DISTRIBUTIONS STATISTIQUES À DEUX CARACTÈRES : ÉTUDE DE LA RÉGRESSION, L'AJUSTEMENT ET LA CORRÉLATION

INTRODUCTION :

Dans les chapitres précédents, on s'est intéressé aux distributions à un seul caractère (univariées). Cependant, une population statistique peut être décrite selon un, deux ou plusieurs caractères. Dans ce qui va suivre, on se limitera aux distributions à deux caractères $x$et y. On peut, par exemple, étudier une population de salariés selon deux caractères :

- L âge (x) et le salaire (y)
- Le salaire (x) et le nombre d'enfants (y)
- L'âge (x) et la qualification (y) …..

D'une manière générale, on peut avoir à étudier simultanément deux caractères quantitatifs ou deux caractères qualitatifs ou encore, un caractère quantitatif et l'autre qualitatif. Le tableau statistique que l'on obtient est un tableau à deux dimensions (tableau de contingence ou à double entrée) où il s'agit de bien faire correspondre les effectifs des lignes et des colonnes.

Nous allons nous intéresser ensuite aux relations éventuelles pouvant exister entre deux caractères quantitatifs (X et Y), c'est-à-dire, si l'un a une influence sur l'autre ou bien s'ils sont complètement indépendants, ceci peut être traité par l'étude de la corrélation.

I/Les tableaux de contingence :

1/Distributions conjointes, marginales et conditionnelles

A/ Distributions conjointes :

Soit une population composée de N individus sur lesquels on observe les variables x et y. Les k modalités de x sont désignées par x_1,x_2, …, $x_k$et les p modalités de y par y_1, y_2,…, y_p.

- La répartition de N observations appelée distribution conjointe se présente sous la forme d'un tableau (ci-dessous) à double entrée où figure en **ligne** les modalités de x et en **colonne** les modalités de y.
- L'effectif n_{ij} désigne le nombre de fois où les modalités de x et les modalités de y ont été observées simultanément.
- L'effectif $\boldsymbol{n_{i.}}$ désigne le nombre total d'observations des modalités de x quelle que soit la modalité de y et $n_{i.} = \sum n_{ij} \quad j = 1$ jusqu'à p ; « j » varie, donc devient « **.** »

- L'effectif $\boldsymbol{n}_{.j}$ représente le nombre total d'observations des modalités de y quelle que soit la modalité de x**;** et nous avons $\boldsymbol{n}_{.j} = \sum n_{ij} \;\; i = 1$ jusqu'à k ; « i » varie, donc devient « **.** »

Le tableau de contingence peut être présenté comme suit :

$x \setminus y$	y_1	y_2	…	…	…	…	y_p	**total** $(\boldsymbol{n}_{i.})$
x_1	n_{11}	n_{12}	…	…	…	…	n_{1p}	$\boldsymbol{n}_{\mathbf{1}.}$
x_2	n_{21}	n_{22}	…	…	…	…	n_{2p}	$\boldsymbol{n}_{\mathbf{2}.}$
⋮	⋮	⋮					⋮	⋮
⋮	⋮	⋮					⋮	⋮
⋮	⋮	⋮					⋮	⋮
x_k	n_{k1}	n_{k2}	…	…	…	…	n_{kp}	$\boldsymbol{n}_{\boldsymbol{k}.}$
Total $\boldsymbol{n}_{.j}$	$\boldsymbol{n}_{.\mathbf{1}}$	**n.2**	…	…	…	…	$\boldsymbol{n}_{.\boldsymbol{p}}$	$\boldsymbol{n}_{..}$**=N**

B/ Distributions marginales

Disposant d'une distribution conjointe, on peut déduire les distributions marginales qui permettent d'étudier séparément chaque variable, ce qui nous permettra de calculer leurs caractéristiques de tendance centrale ou de dispersion. Ainsi les k couples$(x_i, n_{i.})$ définissent la distribution marginale des observations selon la modalité $x$quelle que soit la modalité de y. Cette distribution est représentée par la dernière **colonne** du tableau. De même les $p$couples$\left(y_j, n_{.j}\right)$définissent la distribution marginale des observations suivant la variable y quelle que soit la modalité de x. On obtient :

Distribution marginale de $\boldsymbol{x}$:

x_i	x_1	x_2	$\cdots$	$\cdots$	x_k	total
$n_{i.}$	$n_{1.}$	$n_{2.}$	$\cdots$	$\cdots$	$n_{k.}$	$n_{..} = N$

Distribution marginale de$\boldsymbol{y}$:

y_j	y_1	y_2	$\cdots$	$\cdots$	y_p	total
$n_{.j}$	$n_{.1}$	$n_{.2}$	$\cdots$	$\cdots$	$n_{.p}$	$n_{..} = N$

C/ Distributions conditionnelles

La distribution conditionnelle correspondant à une modalité x_i de la variable $x$suivant les modalités de y est appelée distribution conditionnelle de y pour $x = x_i$. Cette distribution est donnée par le tableau suivant :

$y\,/\,x = x_i$	y_1	y_2	…	…	y_p	total
n_{ij}	n_{i1}	n_{i2}	…	…	n_{ip}	$n_{i.}$

Symétriquement on peut définir la distribution conditionnelle de $x$pour $y = y_j$comme suit :

$x\,/\,y = y_j$	x_1	x_2	…	…	x_k	total
n_{ij}	n_{1j}	n_{2j}	…	…	n_{kj}	$n_{.j}$

2/Notion de fréquences relatives

Comme pour une distribution à un seul caractère, il est possible de calculer des fréquences relatives pour une distribution bi-variée.

A/ Fréquences relatives partielles sur l'effectif total :

La fréquence relative partielle est définie comme étant le rapport du nombre d'individus possédant simultanément la modalité $x_i$de x et la modalité $y_j$de y sur l'effectif total.

$$f_{ij} = {}^{n_{ij}}/_{n_{..}} \text{ et } \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij} = 1$$

La somme des fréquences relatives partielles de tous les couples de valeurs$\{x_i, y_j\}$ est égale à un.

B/Fréquences relatives marginales

Pour la distribution marginale de x: $f_{i.} = \frac{n_{i.}}{n_{..}}$

Pour la distribution marginale de y : $f_{.j} = \frac{n_{.j}}{n_{..}}$

Et $\sum_{i=1}^{k} f_{i.} = 1 = \sum_{j=1}^{p} f_{.j}$. La somme des fréquences marginales est égale à un.

C/Fréquences relatives conditionnelles

On a p fréquences relatives conditionnelles de $x$selon y puisque j varie de 1 jusqu'à p :

$f$de i si j $f_{i/j} = \frac{n_{ij}}{n_{.j}}$

On a k fréquences relatives conditionnelles de y selon $x$puisque i varie de 1 jusqu'à k :

$f$de j si i $f_{j/i} = \frac{n_{ij}}{n_{i.}}$

3/ les paramètres des lois marginales et conditionnelles :

A/Les paramètres des lois marginales (moyenne et variance)

a/Les paramètres des lois marginales selon x

Si x est un caractère quantitatif, on définit les paramètres marginaux (moyenne et variance) à partir de la **colonne** marginale où se trouvent les effectifs $n_{i.}$correspondant respectivement aux k modalités de x.

a) La moyenne marginale de x est $\bar{x}$. Elle est définie comme suit :

$$\bar{x} = \frac{1}{n..}\sum_{i=1}^{k} n_{i.}\, x_i = \sum_{i=1}^{k} f_{i.} x_i$$

b) La variance marginale de x notée $V(x)$. On peut la calculer de deux manières :

Formule de définition : $V(x) = \frac{1}{n..}\sum_{i=1}^{k} n_{i.}\,(x_i - \overline{x})^2 = \sum_{i=1}^{k} f_{i.}\,(x_i - \overline{x})^2$

Formule développée :$V(x) = \frac{1}{n..}\sum_{i=1}^{k} x_i^2 n_{i.} - \overline{x}^2 = \sum_{i=1}^{k} f_{i.} x_i^2 - \overline{x}^2$

Remarque : Dans certains ouvrages, d'autres notations sont utilisés ($\overline{\overline{X}}$; $\overline{\overline{Y}}$…) qui se lisent (X double barre ou Y double barre). Il y a lieu de noter que les deux types de notations sont corrects, et nous avons opté pour la plus simple.

b/ Les paramètres des lois marginales de y :

Si y est un caractère quantitatif, on définit les paramètres marginaux $\bar{y}$ et $V(y)$ à partir de la ligne marginale où se trouvent les effectifs$n_{.j}$ correspondant chacun respectivement aux p modalités y_i de y

a) La moyenne marginale $\bar{y}$: $\bar{y} = \frac{1}{n..}\sum_{j=1}^{p} n_{.j} y_j = \sum_{j=1}^{p} f_{.j} y_j$

b) La variance $V(y)$:

Par définition : $V(y) = \frac{1}{n..}\sum_{j=1}^{p} n_{.j}\left(y_j - \bar{y}\right)^2 = \sum_{j=1}^{p} f_{.j}\left(y_j - \bar{y}\right)^2$ ou

Formule développée :$V(y) = \frac{1}{n..}\sum_{j=1}^{p}\left(n_{.j} y_j^2\right) - \bar{y}^2 = \sum_{j=1}^{p} f_{.j} y_j^2 - \bar{y}^2$

Pour les données groupées d'une distribution conjointe, on peut définir la covariance $V(xy)$comme suit :

a) Formule de définition

$$Cov(xy) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{p}\left[(x_i - \bar{x})\left(y_j - \bar{y}\right)\right] n_{ij}$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{p}\left[(x_i - \bar{x})(y_j - \bar{y})\right] f_{ij}$$

b) Formule développée :

$$Cov(xy) = \left(\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}x_i y_j\right) - \bar{x}\bar{y}$$
$$= \left(\sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij}x_i y_j\right) - \bar{x}\bar{y}$$

B/les paramètres des lois conditionnelles

a/paramètres des distributions conditionnelles de x selon y

Il y a p distributions conditionnelles de x selon y auxquelles correspondent p paramètres conditionnels (p moyennes et p variances)

a) Les moyennes conditionnelles de x selon y, $y = y_j$ (y_j fixe)

$$\bar{x}_j = \frac{1}{n_{.j}}\sum_{j=1}^{k} n_{ij}x_i = \sum_{i=1}^{k} f_{i/j}\, x_i$$

$\bar{x}_j$ est la moyenne conditionnelle de x sachant que y_j fixe ($y = y_j$)

b) Les variances conditionnelles de x selon y ($y = y_j$)

Par définition : $V_j(x) = \frac{1}{n_{.j}}\sum_{i=1}^{k}\left[\left(x_i - \bar{x}_j\right)^2 n_{ij}\right]$

Ou$= \sum_{i=1}^{k}\left(x_i - \bar{x}_j\right)^2 f_{i/j}$

Formule développée:$V_j(x) = \frac{1}{n_{.j}}\sum_{i=1}^{k} n_{ij}x_i^2 - \bar{x}_j^2 = \sum_{i=1}^{k} x_i^2 f_{i/j} - \bar{x}_j^2$

b/paramètres des distributions conditionnelles de y selon x

Ici on fixe $x = x_i$

De façon analogue, on a k distributions conditionnelles de $\boldsymbol{x}$ selon **y** auxquelles correspondent k paramètres conditionnels (moyennes, variances)

a. Les moyennes conditionnelles de $\boldsymbol{y}$ selon $\boldsymbol{x}$

$$\bar{y}_i = \frac{1}{n_{i.}}\left[\sum_{j=1}^{P} n_{ij}y_j\right] = \sum_{j=1}^{P} f_{j/i}y_j$$

b. Les variances de y selon x

Par définition

$$V_i(y) = \frac{1}{n_{i.}}\sum_{j=1}^{P}\left[\left(y_j - \bar{y}_j\right)^2 n_{ij}\right]$$

ou $$= \sum_{j=1}^{P}\left(y_j - \bar{y}_j\right)^2 f_{j/i}$$

Formule développée :

$$V_i(y) = \frac{1}{n_{i.}}\sum_{j=1}^{P}\left(n_{ij}y_j^2\right) - \bar{y}_j^2$$

Ou $= \sum_{j=1}^{P}\left(f_{j/i}y_j^2\right) - \bar{y}_j^2$

Application 1:

Le tableau suivant donne la répartition de 1000 familles selon l'âge du père (X_i) et le nombre d'enfants (Yj)

Yj / Xi	Moins de 2 enfants	[2-5[	5 et plus	**Totaux ($n_{i.}$)**
Moins de 25 ans	100	20	5	125
[25 -30[	50	25	15	90
[30-40[	30	100	100	230
40 et plus	20	200	335	555
Totaux ($n_{.j}$)	200	345	455	1000

Questions :

1/ Quel est le nombre de familles ayant de 2 à 5 enfants et dont l'âge du père est compris entre 30 et 40 ans ?

2/ Quel est le nombre de familles ayant moins de 2 enfants ?

3/ Quel est le nombre de familles dont l'âge du père est égal à 40 ans et plus ?

4/ Que signifie le nombre 5 de la première ligne et de la troisième colonne ?

5/ Donner les valeurs de $\mathbf{n_{11}}$; $\mathbf{n_{23}}$; $\mathbf{n_{2.}}$; $\mathbf{n_{.3}}$

6/Déterminer la distribution marginale du caractère X.

7/Déterminer la distribution marginale du caractère Y.

8/Dégager la distribution conditionnelle de Y selon X$\in [30-40[$.

9/Dégager la distribution conditionnelle de X selon Y∈ [2-5[.

10/Calculer **f_{33} ; $f_{2.}$; $f_{.3}$; $f_{1/2}$ avec i fixé et $f_{3/2}$ avec j fixé.**

Réponses :

1/ Le nombre de familles ayant de 2 à 5 enfants et dont l'âge du père est compris entre 30 et 40 ans est égal à 100 familles (l'effectif partiel **n_{32}).**

2/ Le nombre de famille ayant moins de 2 enfants est égal à 200 (l'effectif marginal de Y : **$n_{.1}$**).

3/ Le nombre de familles dont l'âge du père est égal à 40 ans et plus est 555 (L'effectif marginal de X : **$n_{4.}$**)

4/ Le nombre 5 de la première ligne et de la troisième colonne (**n_{13}**) représente le nombre de famille ayant 5 enfants ou plus et dont l'âge du père est inférieur à 25 ans.

5/ **n_{11}**= 100 ; **n_{23}**= 15 ; **$n_{2.}$** = 90 ; **$n_{.3}$** = 455

6/ Distribution marginale de X :

X âge du père	Moins de 25 ans	[25-30[	[30-40[	40 et plus	Total
$n_{i.}$	125	90	230	555	1000

7/ Distribution marginale de Y :

Y nombre d'enfants	Moins de 2 enfants	[2-5[	5 et plus	Total
$n_{.j}$	200	345	455	1000

8/ Distribution conditionnelle de Y selon X$\in [30-40[$:

Y/X∈ [30-40[	Moins de 2 enfants	[2-5[	5 et plus	Total
n_{3j}	30	100	100	230 (**$n_{3.}$**)

9/ Distribution conditionnelle de X selon Y∈ [2-5[:

X/Y∈ [2-5[	Moins de 25 ans	[25-30[	[30-40[	[40 et plus	Total
n_{i2}	20	25	100	200	345 (**$n_{.2}$**)

10/ Les fréquences relatives :

f_{33} $= \frac{n_{33}}{n_{..}} = \frac{100}{1000}$ = 0, 10 ou 10% (Fréquence partielle sur l'effectif total). Cela signifie qu'il y a 10% de familles ayant 5 enfants et plus et dont l'âge du père est compris entre 30 et 40.

$f_{2.}$ $= \frac{n_{2.}}{n_{..}} = \frac{90}{1000}$ = 0,09 ou 9% (Fréquence marginale de X). Cela signifie qu'il y a 9% de familles dont l'âge du père est compris entre 25 et 30 ans quel que soit le nombre d'enfants.

$f_{.3}$ $= \frac{n_{.3}}{n_{..}} = \frac{455}{1000}$ = 0,455 ou 45,5% (Fréquence marginale de Y). Cela veut dire qu'il y a 45,5% de familles ayant 5 enfants et plus quel que soit l'âge du père.

$f_{1/2}$ avec i fixé $= \frac{\boldsymbol{n_{21}}}{\boldsymbol{n_{2.}}} = \frac{50}{90}$ = 0,5555 ou 55,55% (la fréquence conditionnelle de Y avec j = 1 si i = 2. Cela signifie que parmi les familles dont l'âge du père est compris entre 25 et 30 ans, 55,55% ont moins de deux enfants.

$f_{3/2}$ avec j fixé $= \frac{\boldsymbol{n_{32}}}{\boldsymbol{n_{.2}}} = \frac{100}{345}$ = 0,2898 ou 28,98% (la fréquence conditionnelle de X avec i = 3 si j = 2). Cela veut dire que parmi les familles ayant de 2 à 5 enfants, 28,98% des pères ont l'âge compris entre 30 et 40 ans.

Application 2 :

On a interrogé 200 élèves d'un lycée (garçons et filles) sur le type d'études supérieures qu'ils envisagent de poursuivre (littérature, scientifique ou technique). Les 2 variables étudiées sont donc « Etude » et « Sexe ». Les données sont les suivantes :

y / x	$littérature$	$scientifique$	$technique$	$total$
$garçons$	60	42	18	120
$filles$	60	18	02	80
$total$	120	60	20	200

- n_{11} = 60 n_{12} = 42 n_{23} = 02
- Avec le tableau précédent il est difficile de tirer des conclusions car le nombre de filles est différent du nombre de garçons. C'est pour cela qu'on remplace, généralement, les effectifs par les fréquences partielles $\mathbf{f_{ij}}$
- Dans le tableau suivant, les fréquences f_{ij} correspondent aux rapports entre les n_{ij} et l'effectif total $\mathbf{n_{..}}$ ici N= $n_{..}$ =200.

Tableau des fréquences partielles

y / x	$littérature$	$scientifique$	$technique$	$total$
garçons	0.3	0.21	0.09	0.6
filles	0.3	0.09	0.01	0.04
$total$	0.6	0.3	0.1	1.00

Tableau des effectifs et des fréquences marginales

Les effectifs **lignes** et **colonnes** sont notés respectivement $n_{.j}$ et $n_{i.}$et correspondent aux effectifs pour chaque modalité de la variable.

Tableaux des effectifs marginaux

X (Sexe)	Garçons	filles	Total
$\mathbf{n_{i.}}$	120	80	200

Y (Etude)	Littérature	scientifique	technique	Total
$\mathbf{n}_{.j}$	120	60	20	200

Tableaux des fréquences marginales

X Sexe	Garçons	Filles	Total
$\mathbf{f}_{i.}$	0,6	0,4	1

D'après le tableau, 60% des élèves sont des garçons et 40% sont des filles.

Y Etude	Littérature	Scientifique	Technique	Total
$\mathbf{f}_{.j}$	0,6	0,3	0,1	1

D'après le tableau, 60% des élèves sont des littéraires. 30% sont des scientifiques et 10% sont dans la filière technique.

Tableau des effectifs conditionnels :

y / x	***littérature***	***scientifique***	***technique***	$\mathbf{n}_{i.}$
garçons	60	42	18	120
filles	60	18	02	80
$\mathbf{n}_{.j}$	120	60	20	200

IL est alors possible de calculer les fréquences relativement aux effectifs marginaux et non plus à l'effectif total.

- Fréquences **des garçons et des filles** pour chaque modalité de la variable étudiée (étude). Les totaux considérés sont : 120 pour littérature, 60 pour scientifique et 20 pour technique. On obtient les fréquences conditionnelles de x résumées dans le tableau suivant.

Tableau des fréquences conditionnelles de x

y / x	$littérature$	$scientifique$	$technique$
$garçons$	60/120=0.5	42/60=0.7	18/20=0.9
$filles$	60/120=0.5	18/60=0.3	02/20=0.1
total	1	1	1

On peut voir, par exemple que 50% des littéraires sont des garçons, que 30% des scientifiques sont des filles et que 90% des techniques sont des garçons.

- Fréquence de chaque type d'études pour les garçons et les filles séparément. Le tableau suivant présente les fréquences des effectifs partiels selon la distribution de **y**. Les totaux considérés sont ceux de la dernière colonne à savoir 120 pour les garçons et 80 pour les filles. Ici on divise par les $\mathbf{n_{i.}}$

Tableau des fréquences conditionnelles de y

y / x	$littérature$	$scientifique$	$technique$	$total$
garçons	60/120=0.5	42/120=0.35	18/120=0.15	1
filles	60/80=0.75	18/80=0.225	02/80=0.025	1

On peut ainsi voir que pour les :

Les garçons : 50% sont des littéraires, 35% sont des scientifiques et 15% sont des techniques.

Les filles : 75% des filles sont des littéraires ,22.5% sont des scientifiques et seulement 2,5% sont techniques.

Section 2 : Ajustement, regression et correlation

Nous avons vues dans la section precedente comment construire un tableau statistique avec deux (ou plusieurs) caractéres et comment calculer un certain nombre de parametttres. Dans la suite de ce chapitre nous allons nous interesser à la liaison qui peut exister entre les deux variables (x et y). En fait, il est frequent de s'interroger sur la relation qui peut exister entre deux grandeurs, en particulier dans les problemes de prévision et d'estimation.

Trois types de problemes peuvent apparaitre :

- On dispose d'un certain nombre de points (x_i ; y_j) tel que x_i et y_j sont les valeurs prises par les grandeurs (x et y). On essaye de déterminer une relation fonctionnelle entre x et y. Cette relation s'écrira y= f (x, a, b…) et le problème sera d'ajuster au mieux les paramètres a, b…pour que la courbe représentative de la fonction (f) passe au plus près des points (x_i ; y_j). Il s'agit ici d'un problème d'ajustement analytique.
- On essaye de déterminer la relation statistique qui existe entre les deux grandeurs x et y. on considère que la variation de la variable x explique celle de y, c'est ainsi qu'on considère x comme une variable explicative et y comme une variable expliquée. On a alors d'un point de vue statistique une relation de cause à effet. Ce type d'analyse s'appelle analyse de la régression.
- Les deux grandeurs sont aléatoires et on cherche à savoir si les variations sont liées. On cherche alors à mesurer l'intensité et le degré d'association entre les variables. On est en présence d'un problème de corrélation.

Par ailleurs, on distingue trois types de liaisons entre les caractères x et y.

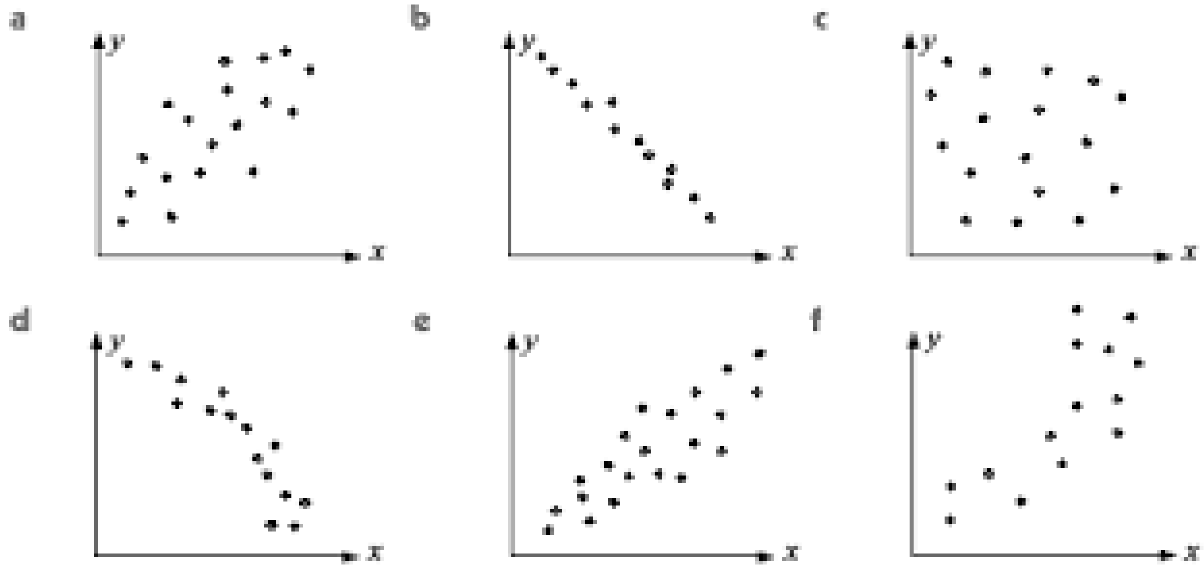
1. ***Indépendance totale (absence de liaison)*** : Dans ce cas les deux caractères sont complétement indépendants car leurs influences réciproques sont nulles.
2. ***Liaison fonctionnelle ou dépendance totale :*** Dans ce cas les deux caractères sont intimement liés de tel sorte que la variation de l'un implique, certainement, celle de l'autre. Ceci est l'objectif de de l'étude de l'ajustement et plus particulièrement de l'analyse de la régression.
3. ***La liaison relative*** : Dans ce cas la liaison entre les deux caractères n'est ni nulle ni totale mais relative (faible, moyenne ou forte).

Remarques :

- Dans ce cours destiné aux étudiants de première année universitaire nous n'allons pas distingués entre l'ajustement et la régression. Nous allons volontairement confondre les deux notions quoi que nous avons souligner dans l'introduction de cette section que chacune des notions répond à un type d'analyse particulier.
- Nous allons traiter uniquement des cas d'études à deux variables et contrairement à la première section nous allons traiter des séries simples c'est-à-dire ne contenant pas les n_{ij}

I- ***L'AJUSTEMENT (REGRESSION SIMPLE)*** :

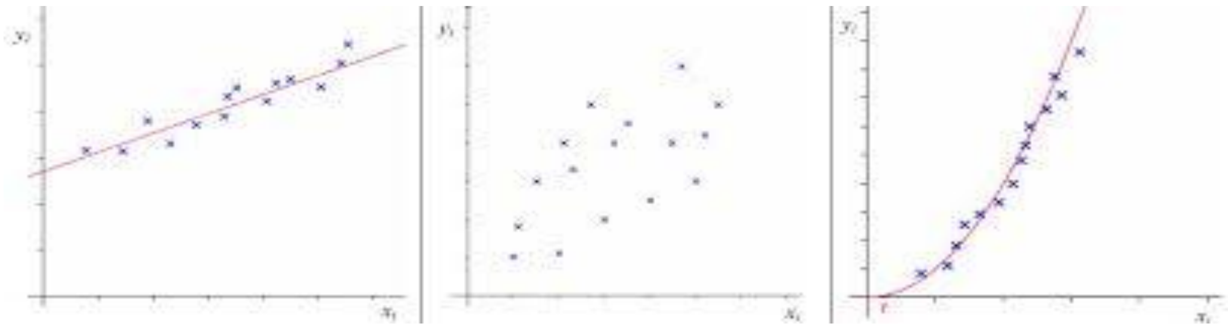
Le premier réflexe qui nous vient à l'esprit quand nous sommes devant un tableau statistique est de vouloir visualiser le phénomène en le représentant graphiquement. Il s'ensuit une représentation sous forme d'un nuage de points qui peut prendre plusieurs formes comme nous le voyons dans ce qui suis :



Pour mieux apprécier la relation entre les variables x et y trois types d'ajustement sont possibles.

a- **L'ajustement graphique** : il y a lieu de rappeler que la représentation graphique précédente (nuage de points) s'obtient comme suit :
   - On porte sur l'axe des abscisses, les valeurs de la variable x.
   - On porte sur l'axe des ordonnées, les valeurs de la variable y.

Une fois le nuage de points représenté nous cherchons à réaliser, à main levée, une courbe qui passe au plus près de l'ensemble des points.



Si le nuage de points forme une droite comme dans le premier et dernier graphe on parle d'une liaison linéaire entre les deux variables.

Evidement la liaison peut ne pas être linéaire, comme il peut ne pas y avoir de liaison entre les variable (indépendance entre x et y).

b- Ajustement mécanique : Dans ce cas deux méthodes peuvent être utilisées.

- ✓ **Méthode des moyennes échelonnées :** elle consiste à diviser la série statistique en plusieurs groupes, pour chaque groupe on calcule la Médiane (Me) pour les valeurs de la variable x et la Moyenne arithmétique ($\bar{X}$) pour les valeurs de la variable y.

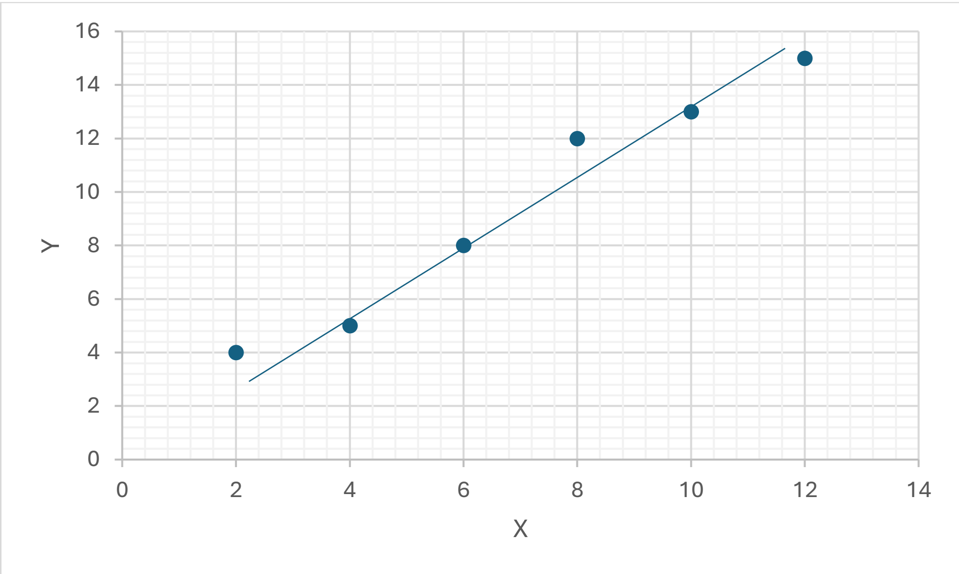
Exemple : soit la série bi-variée suivante

X	2	4	6	8	10	12
Y	4	5	8	12	13	15

Question : Déterminer l'ensemble des points correspondant aux couples (x_i ; y_j) par la méthode de moyennes échelonnées (ordre 3).

Solution :

- On forme des sous-ensembles composés de 3 valeurs chacun.
- On calcul les Médianes pour les sous-ensembles de la variable x_i. On aura donc
  - ✓ Pour le premier groupe de valeurs x_i ; c'est-à-dire (2 ;4 ;6) : M_e= 4.
  - ✓ Pour le deuxième groupe de valeurs x_i ; c'est-à-dire (8 ;10 ;12) : M_e= 10.
- On calcul les moyennes arithmétiques pour les sous-ensembles de la variable y_j. On aura donc :
  - ✓ Pour le premier groupe de valeurs y_j ; c'est-à-dire (4 ;5 ; 8) : $\bar{X}$= 5,66.
  - ✓ Pour le deuxième groupe de valeurs y_j ; c'est-à-dire (12 ;13,15): $\bar{X}$= 13,33.
- On déduit alors les coordonnés des deux points déjà calculés : P_1 (4 ;5,66) et P_2(10 ;13,33).
- On réalise, enfin, la représentation graphique qui reprend les données du tableau (nuage des points) sur lequel nous traçons une droite qui passe par les deux points, précédemment, calculés.



La méthode des moyennes mobiles : le principe de calcul ressemble à celui des moyennes échelonnées (Médiane pour les x_i et moyenne arithmétique pour les y_j). La différence se situe dans la formation des sous-ensembles qui ne sont pas strictement distincts les uns des autres. Autrement dit, les valeurs se répètent dans plusieurs sous-ensembles.

Exemple : Soit la série bi-variée suivante

X_i	2	4	6	8	10	12	14	16	18
Y_j	4	5	8	12	13	15	18	21	24

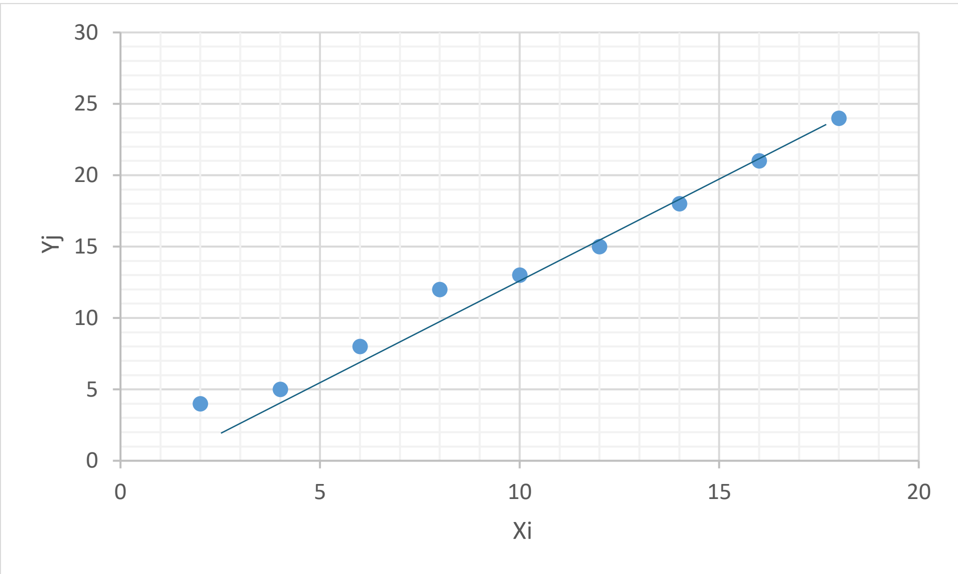
Question : Déterminer l'ensemble des points correspondant aux couples (x_i ; y_j) par la méthode de moyennes mobiles (ordre 3).

Solution :

- On forme des sous-ensembles composés de 3 valeurs chacun. On calcul, alors les médianes pour les sous-ensembles de la variable x_i et les moyennes arithmétiques pour les sous-ensembles de la variable y_j. Cela nous permet de déduire les coordonnés des points correspondant aux couples (x_i ; y_j).

Sous-ensembles x_i	Sous-ensembles y_j	Coordonnées (x_i ; y_j)
- 2 ;4 ;6 M_e=4	- 4 ;5 ;8 $\bar{X}$ =4,66	(4 ;4,66)
- 4 ;6 ;8 M_e=6	- 5 ;8 ;12 $\bar{X}$ = 8,33	(6 ; 8,33)
- 6 ;8 ;10 M_e=8	- 8 ;12 ;13 $\bar{X}$ = 11	(8 ;11)
- 8 ;10 ;12 M_e=10	- 12 ;13 ;15 $\bar{X}$ = 13,33	(10 ;13,33)
- 10 ;12 ;14 M_e=12	- 13 ;15 ;18 $\bar{X}$ = 15,33	(12 ; 15,33)
- 12 ;14 ;16 M_e=14	- 15 ;18 ;21 $\bar{X}$ = 18	(14 ; 18)
- 14 ;16 ;18 M_e=16	- 18 ;21 ;24 $\bar{X}$ = 21	(16 ;21)

- On réalise, ensuite, la représentation graphique qui reprend les données du tableau (nuage des points) sur lequel nous traçons une droite qui passe par les points moyens précédemment calculés.



c- Ajustement linéaire ou affiné

On désir ici faire passer, dans le nuage de points représentatif de la série statistique étudiée, une droite qui représente au mieux la relation de dépendance de Y par rapport à X.

L'équation de cette droite est du type Y= a X + b.

Avec - "a" comme coefficient directeur de la droite.

- "b" ordonnée à l'origine.

Plusieurs méthodes de détermination sont possibles, mais, la plus utilisée est la méthode des moindres carrées.

Comme nous l'avons constaté plus haut, il ne faut pas s'attendre à trouver les points correspondants aux observations tous alignés. En effet, certains points se situent au-dessus et d'autres au-dessous de la courbe. Le principe de la droite des moindres carrées est de minimiser les écarts entre les observations et les coordonnées de la droite.

La droite Y= a X + b telle que la somme des carrées des écarts soit minimale est la droite d'ajustement de Y en fonction de X. On peut également chercher à exprimer X en fonction de Y. On cherche alors la droite d'ajustement de X en Y d'équation X= $\acute{a}$ Y+$\acute{b}$.

- **Droite d'ajustement de Y en fonction de X**

Nous avons vu plus haut que la méthode des moindres carrées repose sur le principe de la minimisation des écarts entre les points observés et les points de la droite. Pour que ces écarts soient tous positifs on élève leur somme au carré. C'est ainsi, qu'on parle de la méthode des moindres carrées.

Cela dit, concrètement, il nous faut minimiser $S = \sum_{i=1}^{n}(y_i - ax_i - b)^2$

Ceci revient à trouver a et b qui rendent minimale la somme S.

Nous n'allons pas revenir, en détail, sur la méthode de détermination des coefficients d'ajustement a et b. Nous notons alors :

$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ ou encore $a = \frac{\sum_{i=1}^{n} x_i\, y_i - n\bar{x}\,\bar{y}}{\sum_{i=1}^{n}(x_i)^2 - n\bar{x}^2}$

Par ailleurs $b = \bar{y} - a\bar{x}$

- **Droite d'ajustement de X en fonction de Y**

On exprime X en fonction de Y. Ainsi, l'équation s'écrit X= $\acute{a}$ Y+$\acute{b}$

$\acute{a} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$ $\acute{b} = \bar{x} - \acute{a}\bar{y}$

II- <u>LA CORRELATION</u>

Nous avons souligné plus haut que pour mesurer l'intensité de la relation entre deux variables x et y nous utilisons un indicateur appelé coefficient de corrélation.

Le coefficient de corrélation linéaire r entre les deux variables X et Y se définit soit par la formule

$r = \sqrt{a.\acute{a}}$

Ou encore par la formule

$$r = \frac{Cov\,(x,y)}{\sqrt{V(x).V(y)}} \quad ou \qquad r = \frac{Cov\,(x,y)}{\sigma(X).\sigma(Y)}$$

Autrement dit, $$r = \frac{\sum_{i=1}^{n}(x_{i-}\ \bar{x})(y_{i-}\ \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_{i-}\ \bar{x})^2 .\ \sum_{i=1}^{n}(y_{i-}\ \bar{y})^2}}$$

Ou plus précisément $$r = \frac{\sum_{i=1}^{n}(x_i.y_{i-}\ n\,\overline{xy}}{\sqrt{\sum_{i=1}^{n} X_i^2 .\ n\,\overline{X}^2 .\ \sum_{i=1}^{n} Y_i^2 .n\ \overline{Y}\ ^2}}$$

Le coefficient de corrélation varie entre -1 et +1.

-Si r =0, il y a absence de corrélation entre x et y.

-Si r = +1 ou -1, il y a une corrélation maximale entre x et y, c'est-à-dire que tous les points sont alignés. On parle alors d'une liaison fonctionnelle.

-Si r est proche de +1 ou de -1, cela indique une très forte corrélation linéaire entre les deux variables.

-Si r est proche de zéro, alors il s'agit d'une faible corrélation linéaire entre les deux variables.

Le signe positif (+) signifie que les deux variables varient dans le même sens.

Le signe négatif (-) signifie que les deux variables varient en sens inverse.

Nous pouvons aussi calculer le coefficient de détermination r^2, il exprime le pourcentage de variation de la variable y expliquée par la variable x.

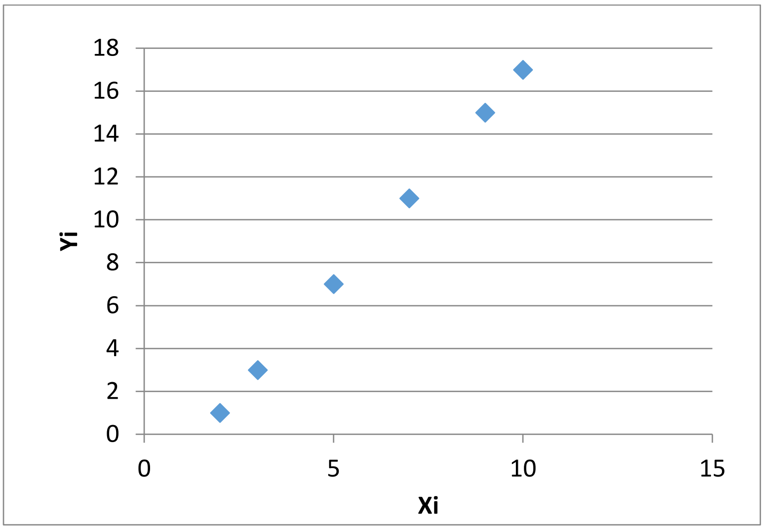
De plus nous pouvons vérifier que $r^2 = a \times \acute{a}$

Application : Soit la série bi-variée suivante où X représente les résultats au test (noté sur 10) de six (6) employés et Y les rendements (en douzaine d'unités).

Application : Soit la série bi-variée suivante où X représente les résultats au test (noté sur 10) de six (6) employés et Y les rendements (en douzaine d'unités).

X_i	2	3	5	7	9	10
Y_i	1	3	7	11	15	17

1/ Représenter le nuage de points.
2/ Trouver l'équation de la droite de régression de Yen X par la méthode des moindres carrés.
3/ Trouver l'équation de la droite de régression de X en Y.
4/ Calculer les coefficients de corrélation et de détermination.
5/ Estimer le rendement d'un employé ayant obtenu un résultat de 4 sur 10.
Solution :
1/ Le nuage de points :



2/ L'équation de la droite de régression de Yen X : Y = aX+b

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
2	1	2	4	1
3	3	9	9	3
5	7	35	25	49
7	11	77	49	121
9	15	135	81	225
10	17	170	100	289
$\sum = 36$	$\sum = 54$	428	268	694

a = $\frac{cov\,(xy)}{v\,(x)}$ avec Cov (xy) = $\frac{\sum X_iY_i}{N} - \bar{X}\,\bar{Y}$ et V(x) = $\frac{\sum X_i^2}{N} - \bar{X}^2$

Calculons d'abord les moyennes marginales : $\bar{X}$ et $\bar{Y}$

$\bar{X} = \frac{\sum X_i}{N} = \frac{36}{6} = 6$ et $\bar{Y} = \frac{\sum Y_i}{N} = \frac{54}{6} = 9$

Cov (xy)= $\frac{428}{6} - (6)\times(9) = 17{,}33$ et V(x) = $\frac{268}{6} - (6)^2 = 8{,}66$

a= $\frac{17{,}33}{8{,}66} = 2$. On trouve le coefficient b comme suit : on a Y= aX+b comme la droite d'ajustement passe par le point moyen ($\bar{X}, \bar{Y}$) $\Rightarrow$ $\bar{Y}$= a$\bar{X}$+b $\Rightarrow$ b = $\bar{Y}$− a$\bar{X}$

b= 9− (2) ×(6) = −3

L'équation est Y= 2X−3 (on peut la représenter sur le nuage de points)

3/ La droite de régression de X en Y : X= $\acute{a}$Y +$\acute{b}$ avec $\acute{a} = \frac{cov\,(xy)}{v(y)}$, calculons la variance de Y :

v(y) $= \frac{\sum Y_i^2}{N} - \bar{Y}^2 = \frac{694}{6} - (9)^2 = 34{,}66$

$\acute{a} = \frac{17{,}33}{34{,}66} = 0{,}5$ et $\acute{b} = \bar{X} - \acute{a}\bar{Y}$ $\Rightarrow$ $\acute{b} = 6 - 0{,}5(9) = 1{,}5$

L'équation est X= 0,5 Y+1,5

4/ Coefficients de corrélation (r) et de détermination r^2 :

$r = \frac{cov\,(xy)}{\sigma(x)\times\sigma(y)} = \frac{17{,}33}{\sqrt{v(X)}\,\sqrt{v(Y)}} = \frac{17{,}33}{\sqrt{8{,}66}\,\sqrt{34{,}66}} = \frac{17{,}33}{2{,}943\times5{,}887} \cong 1$ ou $r = \sqrt{a\acute{a}} = \sqrt{2\times 0{,}5} = 1$

r est égal à 1, il y a une corrélation maximale entre les résultats du test et le rendement des employés.

$r^2 = (1)^2 = 1$ ou 100%. Cela signifie que le rendement des employés est expliqué totalement (à 100%) par les résultats du test.

5/ Si X=4 ; Y= ?, nous avons Y=2X−3 donc Y= 2(4) −3 = 5.

Exercice avec corrige

Exercice n°1 : Le tableau suivant donne la répartition des dépenses mensuelles (10^3 DA), (notées Y_j), des employés d'une entreprise selon le nombre d'enfants (noté X_i)

X_i \ Yj	[0 ; 20[	[20 ; 40[	[40 ; 60[	[60 ; 80[	[80 ; 100[	Totaux
[0 ; 2[	10	6	4	2	0	22
[2 ; 4[	8	6	4	1	0	19
[4 ; 6[	1	2	6	4	3	16
[6; 8 [	0	1	2	4	6	13
[8 ; 10[	0	0	1	1	3	5
Totaux	19	15	17	12	12	75

1-
a/ Que signifie le nombre 6 de la première ligne et de la deuxième colonne ?
b/ Que signifie le nombre 1 de la cinquième ligne et de la troisième colonne ?
2-Déterminer la distribution marginale de X.
3-Déterminer la distribution marginale de Y
4-Calculer la dépense moyenne.
5- Calculer la variance marginale de X.
6-Quelle est la dépense moyenne des employés ayant entre deux et quatre enfants ?
7-Quel est le nombre d'enfants moyen pour les salariés qui dépensent entre 40 000 DA et 60 000 DA ?
8- Calculer les fréquences partielles suivantes : f_{14} ; f_{32}
9- Calculer les fréquences marginales suivantes : $f_{2.}$; $f_{.3}$
10- Calculer les fréquences conditionnelles suivantes : $f_{3/1}$ pour i fixé ; $f_{2/4}$ pour j fixé.

Exercice n°2 : Soit la série bivariée suivante :

X_i	2	6	10	14	18	20
Y_i	16	12	9	12	2	0

1-Représenter le nuage de points.
2- Calculer les moyennes échelonnées (ordre 3) et les moyennes mobiles (ordre 3). Donner uniquement les coordonnées des points.
3-Trouver l'équation de la droite d'ajustement de Y en X par la méthode des moindres carrés.
4-Représenter la droite d'ajustement de Y en X.
5-Calculer les coefficients de corrélation et de détermination. Interpréter les résultats.

6-Estimer la valeur de Y pour X=8

Exercice n°3 : Au cours d'une expérience agricole, on a cherché à connaitre l'influence du facteur eau sur les rendements de betterave à sucre. L'expérience a donné les résultats suivants où X désigne les quantités d'eau et Y les rendements de betterave sucrière correspondants:

Quantités d'eau (X_i) (m^3/ha)	3	5	6	8	9	11	13	15
Rendements de betterave Y_i : Q_x/ha	5	7	8	10	12	14	17	20

1-Trouver l'équation de la droite d'ajustement de Y en X par la méthode des moindres carrés.
2-Représenter le nuage de points et la droite d'ajustement de Y en X
3-Trouver l'équation de la droite d'ajustement de X en Y.
4-Calculer le coefficient de corrélation. Interpréter le résultat.
5-Calculer le coefficient de détermination. Interpréter le résultat.

CORRIGE

Exercice 1 :

1/

a/ le chiffre 6 de la première ligne et de la deuxième colonne (**n_{12}**) représente le nombre d'employés ayant moins de deux enfants et qui dépensent entre 20000 et 40000DA par mois.
b/ Le nombre 1 de la cinquième ligne et de la troisième colonne (**n_{53}**) représente le nombre d'employés ayant de 8 à 10 enfants et qui dépensent entre 40000 et 60000DA par mois.

2/ La distribution marginale de X :

X_i nombre d'enfants	[0-2[	[2-4[	[4-6[	[6-8[	[8-10[	Total
$n_{i.}$	22	19	16	13	05	75

3/ La distribution marginale de Y :

Yj (dépenses)	[0-20[	[20-40[	[40-60[	[60-80[	[80-100[	Total
$n_{.j}$	19	15	17	12	12	75

4/ La dépense moyenne $\overline{Y}$

Yj (dépenses)	[0-20[	[20-40[	[40-60[	[60-80[	[80-100[	Total
$n_{.j}$	19	15	17	12	12	75
Y_j (centre de classe)	10	30	50	70	90	
$n_{.j}\times Y_j$	190	450	850	840	1080	3410

$\bar{Y} = \frac{\sum Yj \times n.j}{n..} = \frac{\sum Yj \times n.j}{N} \Longrightarrow \bar{Y} = \frac{3410}{75} = 45,46\times10^3$ DA.

5/ La variance marginale du caractère X : V(x) = $\frac{\sum X_i^2 \text{ ni.}}{N} - \bar{X}^2$

X_i	[0-2[	[2-4[	[4-6[	[6-8[	[8-10[	Total
ni.	22	19	16	13	05	75
Xi (centre de classe)	1	3	5	7	9	
$X_i \times n_{i.}$	22	57	80	91	45	295
X_i^2	1	9	25	49	81	
$X_i^2 \times n_{i.}$	22	171	400	637	405	1635

On calcule d'abord la moyenne $\bar{X}$: $\bar{X} = \frac{\sum Xini.}{n..} = \frac{295}{75} = 3,93$ enfants.

V(x) = $\frac{1635}{75} - (3,93)^2 = 6,36$ enfants.

6/La dépense moyenne des employés ayant entre 2 et 4 enfants : il s'agit de calculer la moyenne conditionnelle $\bar{Y}$/ X ∈ [2-4[:

Y/X∈ [2-4[	[0-20[	[20-40[	[40-60[	[60-80[	[80-100[	Total
n_{2j}	8	6	4	1	0	19
Yj	10	30	50	70	90	
$n_{2j} \times Y_j$	80	180	200	70	0	530

$\bar{Y}$/ X ∈ [2-4[$= \frac{\sum n_{2j} \times Yj}{\sum n_{2j}} = \frac{530}{19} = 27,89.10^3$DA.

7/ Le nombre d'enfants moyen pour les salariés qui dépensent entre 40000 DA et 60000DA : il s'agit de calculer la moyenne conditionnelle $\bar{X}$/Y∈ [40-60[:

X/ Y ∈ [40-60[	[0-2[	[2-4[	[4-6[	[6-8[	[8-10[	Total
n_{i3}	4	4	6	2	1	17
X_i	1	3	5	7	9	
$n_{i3} \times X_i$	4	12	30	14	9	69

$\bar{X}$/Y∈ [40-60[$= \frac{\sum n_{i3} \times X_i}{\sum n_{i3}} = \frac{69}{17} = 4,06$ enfants.

8/Les fréquences relatives partielles sur l'effectif total :

f_{14} $= \frac{n_{14}}{n..} = \frac{2}{75} = 0,0266\times 100 = 2,66\%$. C'est le pourcentage des employés ayant moins de deux enfants et qui dépensent entre 60000 et 80000DA.

f_{32} $= \frac{n_{32}}{n..} = \frac{2}{75} = 0,0266\times 100 = 2,66\%$. C'est la proportion des employés ayant de 4 à 6 enfants et qui dépensent entre 20000 et 40000 DA par mois.

9/ Les fréquences marginales :

$f_{2.}$ $= \frac{n_{2.}}{n..} = \frac{19}{75} = 0,2533\times 100 = 25,33\%$. C'est la proportion des employés ayant de 2 à 4 enfants (quel que soit le salaire).

$f_{.3}$ $= \frac{n_{.3}}{n..} = \frac{17}{75} = 0,2266$ ou 22,66%. Il y a 22,66% des employés qui dépensent entre 40000 et 60000DA (quel que soit le nombre d'enfants).

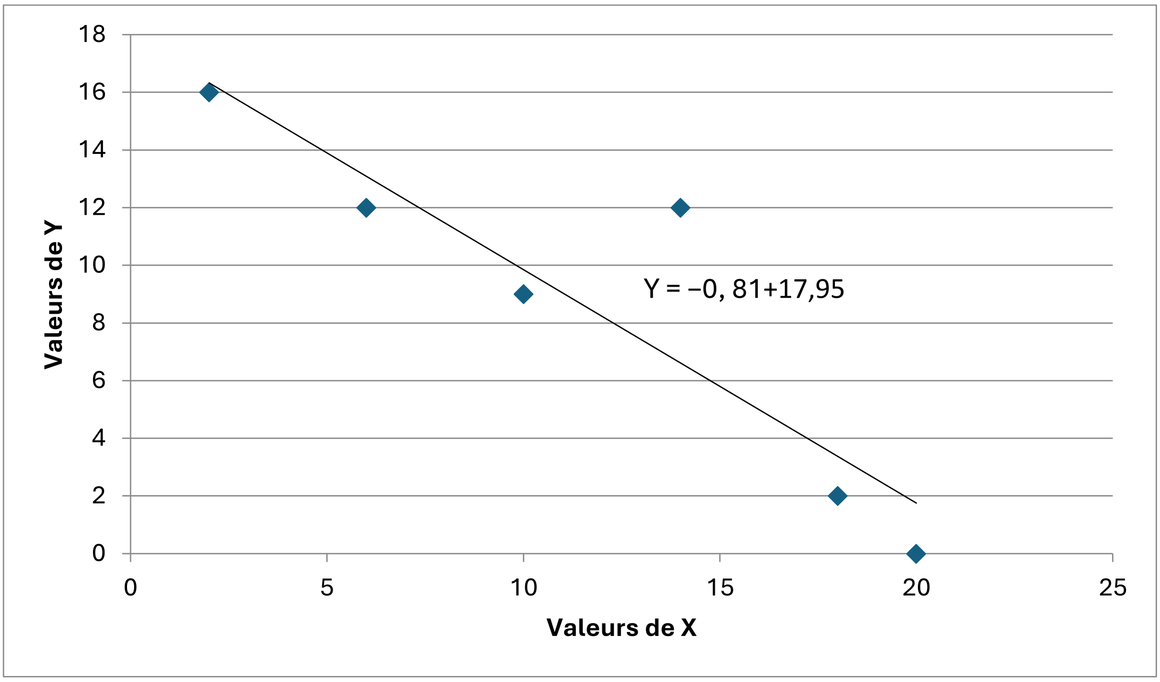
10/ Les fréquences conditionnelles :

$f_{3/1}$ pour i fixé = $\frac{n_{13}}{n_{1.}} = \frac{4}{22}$ = 0, 1818 ou 18, 18%. Cela signifie que parmi les employés ayant moins de 2 enfants, il y en a 18,18% qui dépensent entre 40000 et 60000DA.
$f_{2/4}$ pour j fixé = $\frac{n_{24}}{n_{.4}} = \frac{1}{12}$ = 0,0833 ou 8,33%. Cela signifie que parmi les employés qui dépensent entre 60000 et 80000DA, il y en a 8,33% qui ont entre 2 et 4 enfants.

<u>Exercice 2 :</u>

X_i	2	6	10	14	18	20
Y_i	16	12	9	12	2	0

1/ Nuage de points : sur l'axe des abscisses, on représente les valeurs de la variable indépendante (explicative) X et sur l'axe des ordonnées on représente les valeurs de la variable dépendante (expliquée) Y.



Le nuage de points suggère une relation linéaire négative entre X et Y.
2/ Moyennes échelonnées et moyennes mobiles :
A/ Les moyennes échelonnées d'ordre 3 :
Il s'agit de calculer la médiane (Me) pour les valeurs de X et la moyenne arithmétique ($\bar{Y}$) pour les valeurs de Y comme suit :
Me entre (2 ; 6, 10) = 6 ; Me entre (14 ; 18 ; 20)= 18
$\bar{Y}_1 = \frac{16+12+9}{3} = 12{,}33$; $\bar{Y}_2 = \frac{12+2+0}{3} = 4{,}66$
On obtient les points suivants **: P_1 (6 ; 12,33) ; P_2 (18 ; 4,66).**

B/Les moyennes mobiles d'ordre 3 : on calcule Me pour les valeurs de X et $\bar{Y}$ pour les valeurs de Y comme suit :
<u>Valeurs de X</u> :

Me entre (2 ; 6 ; 10) = 6 ; Me entre (6 ; 10 ; 14) = 10 ; Me entre (10 ; 14 ; 18) = 14 ;
Me entre (14 ; 18 ; 20) = 18.

Valeurs de Y :

$\bar{Y}_1 = \frac{16+12+9}{3} = 12{,}33$; $\bar{Y}_2 = \frac{12+9+12}{3} = 11$; $\bar{Y}_3 = \frac{9+12+2}{3} = 7{,}66$; $\bar{Y}_3 = \frac{12+2+0}{3} = 4{,}66$

On obtient les points suivants : **P_1 (6 ; 12,33) ; P_2 (10 ; 11) ; P_3 (14 ; 7,66) ; P_4 (18 ; 4,66).**

3/ L'équation de la droite d'ajustement (droite de régression) de Y en X par la méthode des moindres carrés : Y= ax+b

a $=\frac{COV(XY)}{V(X)}$ avec COV (XY) $= \frac{\sum X_i Y_i}{N} - \bar{X}\bar{Y}$ et V(X) $= \frac{\sum X_i^2}{N} - \bar{X}^2$

On calcule d'abord $\bar{X}$ et $\bar{Y}$:

$\bar{X} = \frac{\sum X_i}{N} = \frac{70}{6} = 11{,}67$; $\bar{Y} = \frac{\sum Y_i}{N} = \frac{51}{6} = 8{,}5$

On construit le tableau suivant :

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
2	16	32	4	256
6	12	72	36	144
10	9	90	100	81
14	12	168	196	144
18	2	36	324	4
20	0	0	400	0
70	51	398	1060	629

COV(XY) $= \frac{398}{6} - (11{,}67) \times (8{,}5) = 66{,}33 - 99{,}195 = -32{,}865$

V(X) $= \frac{1060}{6} - (11{,}67)^2 = 176{,}67 - 136{,}19 = 40{,}48$

a $= \frac{-32{,}865}{40{,}48} = -0{,}81$

Calculons b : b $= \bar{Y} - a\,\bar{X} \Longrightarrow$ b $= 8{,}5 - (-0{,}81 \times 11{,}67) = 17{,}95$

L'équation est : Y = − 0,81X + 17,95

4/ Représentation de la droite d'ajustement de Y en X: sur le nuage de points, on trace la droite de régression en utilisant l'équation trouvée précédemment (question 3).

5/ Coefficients de corrélation (r) et de détermination (r^2) :

$$\mathbf{r} = \frac{Cov\,(XY)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{COV\,(XY)}{\sigma_x \sigma_y}$$

On calcule la variance de la variable Y : $V(Y) = \frac{\sum Y_i^2}{N} - \bar{Y}^2 \Longrightarrow$

$V(Y) = \frac{629}{6} - (8{,}5)^2 = 32{,}58$

$\mathbf{r} = \frac{-32{,}865}{\sqrt{40{,}48}\,.\,\sqrt{32{,}58}} = \frac{-32{,}865}{6{,}362\times 5{,}708} = \mathbf{-0{,}90}$

$r\ est\ proche\ de$
$-\ 1\ donc\ il\ existe\ une\ forte\ corrélation\ linéaire\ négative\ entre\ X\ et\ Y.$

$r^2 = (-\ 0,\ 90)^2 = 0{,}81$ ou 81% $\Longrightarrow$ Cela signifie que 81% de variation de Y est expliquée par la variation de X.

6/ Si X = 8 ; Y= ? Nous avons Y= − 0,81X + 17,95 $\Longrightarrow$ Y = −0,81 (8) + 17,95 = 11,47.

Exercice 3 :

1/ L'équation de la droite de régression de Y en X : Y= ax+b

a $= \frac{COV(XY)}{V(X)}$ avec COV (XY) $= \frac{\sum X_i Y_i}{N} - \bar{X}\bar{Y}$ et V(X) $= \frac{\sum X_i^2}{N} - \bar{X}^2$

On construit le tableau de calcul suivant :

X_i	Y_i	X_iY_i	X_i^2	Y_i^2
3	5	15	9	25
5	7	35	25	49
6	8	48	36	64
8	10	80	64	100
9	12	108	81	144
11	14	154	121	196
13	17	221	169	289
15	20	300	225	400
70	93	961	730	1267

On calcule $\bar{X}$ et $\bar{Y}$: $\bar{X} = \frac{\sum X_i}{N} = \frac{70}{8} = 8{,}75$; $\bar{Y} = \frac{\sum Y_i}{N} = \frac{93}{8} = 11{,}625.$

COV (XY) $= \frac{961}{8} - (8,\ 75\times 11{,}625) = 18{,}406$

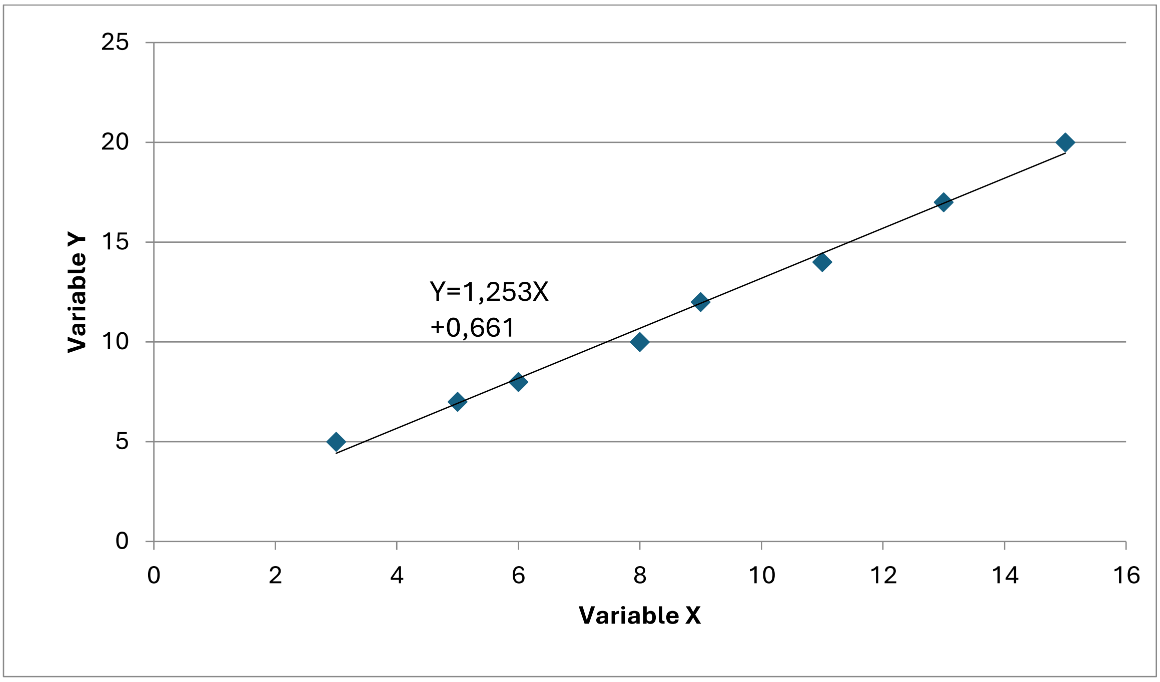
V(X) $= \frac{730}{8} - (8,\ 75)^2 = 14{,}687$

a $= \frac{18{,}406}{14{,}687} = 1{,}253$

b $= \bar{Y} - a\bar{X} \Longrightarrow$ b = 11,625−1,253 (8, 75) = 0,661.

L'équation est : Y = 1,253X+ 0,661

2/ Le nuage de points et la droite de régression de Y en X :



Le nuage de points suggère une relation linéaire positive entre les variables X et Y.

3/ L'équation de la droite de régression de X en Y : X= $\grave{a}$Y + $\grave{b}$:

On peut estimer les valeurs de X à partir de Y, c'est- à- dire que X devient la variable expliquée (dépendante) et Y la variable explicative (indépendante).
On calcule $\grave{a}$ et $\grave{b}$ comme suit :

$\grave{a} = \frac{COV\ (XY)}{V(Y)}$; on calcule d'abord la variance de Y : V(Y) = $\frac{\sum Y_i^2}{N} - \bar{Y}^2$

V(Y) = $\frac{1267}{8}$ – $(11, 625)^2$ = 23, 234

$\grave{a} = \frac{18,406}{23,234}$ = 0,792

$\grave{b} = \bar{X} - \grave{a}\bar{Y} \Longrightarrow \grave{b}$ = 8, 75 – (0,792) (11,625) = – 0, 457

L'équation est : X= 0,792Y – 0, 457

4/ Le coefficient de corrélation r:

$$r = \frac{COV\ (XY)}{\sigma_x \sigma_y} = \frac{18,406}{\sqrt{14,687}.\sqrt{23,234}} = \frac{18,406}{3,8324 \times 4,8201} = 0,996$$

r est très proche de +1 donc il existe une très forte corrélation linéaire positive entre X et Y.
Ou bien, on peut calculer r comme suit : r = $\sqrt{a \times \grave{a}} = \sqrt{1,253 \times 0,792}$ = 0,996

5/ Le coefficient de détermination r^2 :

r^2 = a×$\grave{a}$ = 1,253×0,792 = 0,9923 = 99,23%. Cela signifie que 99,23% de la variation de rendements de la betterave sucrière est expliquée par la variation des quantités d'eau utilisées.

CONCLUSION GENERALE

A travers la lecteur et l'examen des chapitres du présent cours l'étudiant aura ainsi pris connaissance, de façon assez élargit, des soubassements, notions, formules usuelles, règles et des techniques de la statistique descriptive.
Ces connaissances sont indispensables et nécessaires, tant pour la poursuite de son cursus LMD que dans sa future vie professionnelle.
Le cours débute par une introduction au domaine de la statistique descriptive, en exposant les définitions et les notions de base, puis vers la présentation des données avant de passer au calcul des paramètres de mesure des variations des données.
Suite à cela on élargit le champ d'analyse en étudiant les nombres indices et, en dernier lieu, les distributions à deux caractère et l'analyse de la relation entre elles. Nous sommes ainsi passé du simple au complexe.
Les chapitres exposés de cette façon, c'est à dire enchaînée et suivant un raisonnement cohérent et logique, répondant largement aux principes de la statistique descriptive, ont pour but de faciliter à l'étudiant l'assimilation des connaissances proposées et de l'imprégner du raisonnement de la statistique qui est, on ne peut plus, rationnel et scientifique. Cela permet d'accompagner l'étudiant dans sa transition de l'enseignement du secondaire à l'enseignement supérieur.
Au terme de ces chapitres, l'étudiant peut passer à l'étape plus complexe de l'analyse statistique, à savoir ; les probabilités et les variables aléatoires. C'est l'objet de l'enseignement dispensé en Statistique 2 au second semestre.

EXERCICES SUPPLEMENTAIRES

Exercice 01 :

Pour chacun des caractères suivants, préciser la nature :

-Chiffre d'affaires.
-Nationalité.
-Vitesse.
-Poids.
-Origine géographique.
-Age.
-Couleur des cheveux.
-Nombre de langues parlées.

Exercice 02 :

Répondre par vrai ou faux aux expressions suivantes :

1-Il existe quatre quartiles divisant la série statistique en quatre groupes de même effectif.
2-La somme des fréquences relatives est toujours supérieure à un.
3-En statistique, la population désigne un ensemble d'êtres humains.
4- La variance d'une série statistique X est égale à la racine carrée de l'écart-type.
5-Le nombre minimal de modalités d'un caractère est égale à deux.
6- L'intervalle interquartile contient 75% des valeurs centrales de la série.

Exercice 3 : Le tableau suivant donne la répartition des PME privées en Algérie, en 2020, par secteurs d'activité.

Secteurs d'activité	**Nombre**
Agriculture	7 447
Hydrocarbures, Energie et Mines	3 088
Bâtiments, Travaux Publics et Habitat (BTPH)	191438
Industries	104529
Services	622044
Artisanat	280710
Total	1 209252

1- Déterminer : la population statistique, l'unité statistique, le caractère et ses modalités.
2- Représenter graphiquement ces données et faire un commentaire.

Exercice 04 : Au poste de péage, on compte le nombre de voitures se présentant sur une période de 5 min. Sur 50 observations de 5 min, on obtient les résultats suivants :

Nombre de voitures	1	2	3	4	5
Nombre d'observation	2	8	14	20	6

1-Déterminer avec précision : la population statistique, l'unité, le caractère, et ses modalités.

2-Calculer le salaire moyen, le salaire le plus fréquent et le salaire médian,

3-Determinez la médiane graphiquement.

Exercice 05: Une étude statistique auprès d'un échantillon de 2000 étudiants diplômés sur le nombre de voyages effectués à l'étranger a donné les résultats suivants :

Nombre de voyages	0	1	2	3	4	5
Nombre d'étudiants	600	766	282	193	94	65

1-Quel est le caractère étudié ? Préciser sa nature et ses modalités.

2-Calculer les paramètres de tendance centrale : $\overline{X}$, Mo, Me.

3-Représenter graphiquement le mode et la médiane.

Exercice 06 : Un opérateur de téléphonie mobile s'intéresse aux fréquences des appels téléphoniques d'un échantillon de 1000 abonnés durant une journée ordinaire et obtient les résultats suivants :

- 250 abonnés n'ont effectué aucun appel ;
- 378 abonnés ont effectué un seul appel ;
- 218 abonnés ont effectué deux appels ;
- 114 abonnées ont effectué quatre appels ;
- Le reste des abonnés ont effectué cinq appels ou plus.

01- Définir avec précision la population statistique, le caractère étudié la nature du caractère et ses modalités.
02- Construire le tableau statistique complet (y compris les effectifs et les fréquences cumulés croissants)
03- Quelle est la proportion des abonnés ayant effectués au moins trois appels.
04- Déterminer le mode et la médiane graphiquement.

Exercice 07 : Une étude, sur le budget consacré aux vacances d'été par les ménages, a donné les résultats suivants :

Budget (10^3 DA)	[40-60[	[60-80[	[80-100[	[100-120[	[120-140[	[140-160[
Nombre de ménages	8	10	16	30	9	27

1- Déterminer : la population statistique, l'unité statistique, le caractère et sa nature.
2- Quelle est la proportion des ménages :
-Dont le budget consacré aux vacances est inférieur à 100000 DA ?
-Dont le budget consacré aux vacances est supérieur ou égale 120000 ?
3- Calculer le mode et la médiane
4- Calculer Q_1 et C_{50}
5-Calculer la variance par la méthode de changement de variable (avec x_0 =110 et a=20)
6- Calculer l'indice de Gini et tracer la courbe de concentration.

Exercice 08 : Le tableau suivant donne la répartition des petites entreprises d'une commune selon le chiffre d'affaires annuel (en millier de dinars).

CA (10^3DA)	[500-1000[	[1000-1500[	[1500-2000[	[2000-5000[	[5000-10000[
Nombre d'entreprises	70	38	32	18	12

1- Déterminer : la population statistique, l'unité statistique, le caractère et sa nature.
2- Quelle est la proportion des petites entreprises ayant un chiffre d'affaires inférieur à 1.500.000 DA ?
3- Quelle est la proportion des petites entreprises ayant un chiffre d'affaires supérieur ou égal à 1.500.000 DA ?
4- Calculer le mode et interpréter le résultat.
5- Calculer la moyenne arithmétique.
6- Calculer l'écart- type de cette distribution en utilisant la formule de changement de variable. On pose $x_i' = \frac{x_i - x_0}{a}$, avec x_0=750 et a=500.
7- Calculer la médiane et donner sa représentation graphique.
8- Calculer le troisième quartile et le quatrième décile.
9- Calculer la médiale et analyser la concentration avec ΔM/e.

Exercice 09 : Soit le tableau suivant qui présente les dépenses des ménages d'une commune donnée dans l'entretient de leurs résidences principales.

Dépenses en 10^3 DA	effectifs
[0-4[	06
[4-8[	25
[8-12[	24
[12-e_4[	17
[e_4-22[	14
[22-30[	11
[30-42[	03

1. Définir avec précision la population et l'unité statistiques le caractère étudié et sa nature.
2. Démontrer que la borne manquante est égale à 16, sachant que la moyenne arithmétique est égale à 13.
3. Calculer le mode et la médiane.
4. Déterminer le mode graphiquement
5. Calculer D_5 et C_5.
6. Calculer l'écart absolu moyen et l'écart type.

Analyser la concentration par ΔM/e.

Exercice 10: Le prix d'un article était en 2006 de 20% supérieur à son prix en 2005, de 20% inférieur à son prix en 2004 et de 50% supérieur à son prix en 2007.

Transformer ces données en indices élémentaires en prenant 2005 comme année de référence.

Exercice 11 : Un touriste veut étudier l'évolution de ses dépenses d'hôtel et de restauration pour les années 2006 et 2010.

	2006		**2010**	
	Prix	Quantité	Prix	Quantité
Hôtels	120	80	90	100
Restaurants	200	50	150	70
Cafés	40	10	50	30

1) Calculer les indices élémentaires (simples) des prix en 2010, base 100 en 2006.

2) Calculer l'indice des dépenses ou valeurs globales en 2010, base 100 en 2006.

3) Calculer $\mathbf{L^{P}_{2010/2006}}$ en utilisant la formule de définition. Interpréter le résultat.

4) Calculer $\mathbf{P^{P}_{2010/2006}}$ en utilisant la formule développée (simplifiée).

5) Déduire à partir des résultats de la question 2 et 3 l'indice Paasche des quantités en 2010, base 100 en 2006.

Exercice 12: Le tableau suivant résume l'état des ventes de voitures d'un garage l'an dernier en fonction de leur prix de vente (xi) et de leur cylindrée (yi).

(Y) Cylindrée 10^2 cm^3 / (X) Prix 10^3 \$	**[6-10[**	**[10-20[**	**[20-30[**	**Total**
[9-15[	35	10	5	50
[15-19[	10	60	20	90
[19-21[	0	5	25	30
Total	45	75	50	**170**

Questions :

1-Que signifie le nombre 60 de la $2^{éme}$ ligne et de la $2^{éme}$ colonne ?

2-Que signifie le nombre 90 de la $2^{éme}$ ligne et de la marge de droite ?

3-Que signifie le nombre 50 de la $3^{éme}$ colonne et de la marge d'en bas ?

4-Calculer les fréquences suivantes : f_{21}, f_{12}, $f_{2.}$, $f_{.3}$

5-Calculer le prix moyen.

6-Calculer la variance de Y.

7-Calculer le prix moyen pour les voitures dont la cylindrée est comprise entre **[20-30[** x 10^2 cm^3.

Exercice 13 : Soit la série bi variée suivante où $\mathbf{X_i}$ représente le revenu mensuel et $\mathbf{Y_i}$ la dépense mensuelle.

$\mathbf{X_i}$	2	4	6	8	10
$\mathbf{Y_i}$	2	3	5	6	9

Questions :

1-Représenter le nuage de points.

2-Calculer les moyennes échelonnées (ordre 3) et mobiles (ordre 3). Représenter les graphiquement.

3-Trouver l'équation de la droite d'ajustement de Y en X par la méthode des moindres carrés (y = ax + b). Faites sa représentation graphique.

3- Calculer les coefficients de corrélation (r) et le coefficient de détermination (r^2). Commenter vos résultats

BIBLIOGRAPHIE (Ouvrages disponible à la Bibliothéque de la Faculté)

Anderson, R., Dennis J. Sweeney, D.J., & Williams, T.A. (2006). Statistiques pour l'économie et la gestion. 2e Edition. De Boeck, Paris. 779 p.

Anderson, D.R., Camm, J., Cochran, J.J., Sweeney, D.J, & Williams, T.A. (2015). Statistiques pour l'économie et la gestion.5e Edition. De Boeck Supérieur, Paris. 944 p.

Bailly, P. (1993). L'Economie et les chiffres : exercices corrigés de statistique descriptive. OPU, Alger. 141 p.

Benmessaoud, M., & Oukacha, B. (2008). Statistiques descriptives et calculs des probabilités : cours et exercices corrigés. Pages bleues. Alger.

Boudia, M C. (2008). Statistique descriptive. Casbah, Alger. 315 p.

Boukella-Bouzaouane, M. (2001). Statistique descriptive : rappels de cours avec exercices corrigés. Casbah. Alger. 171 p.

Boursin, J-L. (1991). Comprendre la statistique descriptive. AC. Paris. 163 p.

Boursin, J-L. (2000). L'essentiel de la statistique pour l'économie et la gestion. Gualino. Paris. 127 p.

Bressoud, É., & Kahané, J-C. (2009). Statistique descriptive. Collection Synthex. Pearson Education France. 258p.

Chauvat, G. (1992). Statistiques descriptives. Paris : AC. 205 p.

Dhuin, C : Problèmes corrigés de statistiques : posés aux examens du Deug de sciences économiques (1ère et 2ère année). Paris : Ellipses, 192 p.

Dussaix, A-M. (1995). Statistique pour la gestion. Alger : Chihab. 340 p.

Duthil, G. Initiation à la statistique descriptive. Paris : Ellipses. 191p.

Grais, B. (1998). Statistique descriptive avec rappels de cours. Paris : Dunod. 234 p.

Golfarb, B., & Paradoux, C. (2011). Introduction à la méthode statistique, 5ème édition. Dunod. 16.

Grais, B. (2000). Manuel statistiques descriptive . 3e Ed. Paris : Dunod. 280 p.

Hamdani, H. (1988). Statistique descriptive et expression graphique. Alger : OPU. 381 p.

Hamdani, H. (2006). Statistique descriptive avec initiation aux méthodes d'analyse de l'information économique : exercices corrigés. 5éme éd.. Alger : OPU. 259 p.

Hubler, J. (2007). Statistique descriptive : appliquée à la gestion et à l'économie. 2ème éd. Paris : Bréal. 219 p.

Hurlin, C., & Mignon, V. (2015). Statistique et probabilités en économie-gestion. Dunod. Paris. 370p.

Hurlin, C., & Mignon, V. (2018). Statistique et probabilités en économie-gestion. Dunod. Paris. 370p.

Hurlin, C., & Mignon, V. (2022). Statistique et probabilités en économie-gestion. Dunod. 2ème éd°. Collection : Openbook. Paris. 416p.

Janvier, M. (1999). Statistique descriptive avec ou sans tableur : cours et exercices corrigés. Paris : Dunod. 276 p.

Labenne, C. (1995). Introduction à la statistique descriptive et probabilités. Paris : Economica. 197 p.

Labrousse, C. (1987). Statistique : exercices corrigés avec rappels de cours / Christian Labrousse. - 4°éd.. - Paris : Dunod, 1987. - 292 p.

Labrousse, C. (1991). Statistique : exercices corrigés avec rappels de cours. 5°éd. Paris : Dunod. 292 p.

Lasary, J. (2001). La statistique descriptive à portée de tous. Alger : [s.n.]. 188 p.

Lecoutre, J-P. (1990). Statistique descriptive : exercices corrigés avec rappels de cours. Paris : Masson, 1990. - 222 p. ; 24 cm.

Leboucher, L., & Voisin, M-J. (2011). Introduction à la statistique descriptive (cours et exercices avec tableur). Cépaduès-Éditions. 208p.

Doane, G.P., & Seward, L. (2016). Applied Statistics in Business and Economics, Fifth Edition. MacGraw-Hill. 864 p.

Mazerolle, F. (2006). Statistique descriptive : séries statistiques à une et deux variables; séries chronologiques ; indices. Paris : Gualino. 172 p.

Monino, J.L., Kosianski, J.M., & Le Cornu, F. (2000). Statistique descriptive rappels de cours, questions de réflexion, exercices d'entrainement, Annales corrigés. Paris : Dunod. 248 p.

Monino, J.L. (2017). TD de statistique descriptive - 5e éd. Dunod. Paris. 353p.

Py, B. (1994). Exercices corrigés de statistique descriptive : problèmes exercices et Q.C.M. 2e éd. Paris : Economica. 177 p.

Py, B. (1996). Statistique descriptive : nouvelle méthode pour bien comprendre et réussir. 4 ° éd. Paris : Economica. 353 p.

Py, B. (2007). Statistique descriptive : nouvelle méthode pour bien comprendre et réussir / 5éme éd. Paris : Economica. 353 p.

Verlant, B. Saint-Pierre, G. (2008). Statistique et probabilité : Manuel de cours -exercice corrigé-sujets d’examens. BERTI Edition. 304 p.