

Ministère de l'enseignement supérieur et
de la recherche scientifique

Université Djillali Liabes Sidi Bel Abbès

Faculté des sciences économiques,
commerciales et sciences de gestion



وزارة التعليم العالي والبحث العلمي

جامعة جيلالي ليابس بسيدي بلعباس

كلية العلوم الاقتصادية والتجارية

وعلوم التسيير

Département Sciences de Gestion

قسم علوم التسيير

Polycopies-en :

Statistique I

Première Année Licence

Auteur :



Année Universitaire 2021-2022

Sommaire

Table des matières

| | |
|---|-----------|
| Sommaire..... | I |
| Introduction : | 1 |
| Conférence n°01 | 2 |
| I.Généralités sur la statistique..... | 3 |
| Définition :..... | 3 |
| I.1 Les statistiques | 3 |
| I.2 La statistique | 3 |
| II.TERMINOLOGIE STATISTIQUE | 4 |
| II.1.Définitions concepts de base de la statistique descriptive | 4 |
| II.1.1.Population : | 4 |
| II.1.2.Individu | 4 |
| II.1.3.caractère ou variable..... | 4 |
| II.1.4.Modalité | 4 |
| II.1.5.série statique | 4 |
| II.2.Les différents types de caractères et de variables statistiques..... | 4 |
| II.2.1.caractère qualitatif | 4 |
| II.2.2.caractère quantitatif..... | 5 |
| Conférence n°02 | 6 |
| III. DISTRIBUTIONS STATISTIQUES A UN CARACTERE..... | 7 |
| III.1.Tableaux et représentations graphiques..... | 7 |
| III.1.1 les tableaux statistiques et les différents types de fréquences. | 7 |
| Conférence n°03 | 11 |
| III.1.2 Représentation graphique..... | 12 |
| Exercices proposées | 26 |
| Conférence n°04 | 28 |
| IV. Paramètres de position | 29 |
| IV.1.Le mode..... | 29 |
| IV.1.1.Cas Variable Statistique Discrète | 30 |
| IV.1.2.Cas Variable Statistique Continue | 30 |
| IV.2.La médiane (Me) | 31 |
| IV.2.1.Détermination pratique | 32 |
| IV.2.2.Cas d'une variable statistique discrète | 32 |

| | |
|--|-----------|
| Conférence n°05 | 34 |
| IV.2.3.Cas d'une variable statistique continue | 35 |
| IV.3.La moyenne | 37 |
| IV.3.1.La moyenne arithmétique..... | 37 |
| Conférence n°06 | 40 |
| IV.3.2.Moyenne géométrique | 41 |
| IV.3.3.Moyenne harmonique H | 42 |
| IV.4. Quantiles : | 44 |
| IV.4.1.Cas .Variables Statistique Discrète (V.S.D) | 46 |
| IV.4.2.Cas variable statistique continue | 47 |
| Conférence n°07 | 48 |
| V. Les paramètres de dispersion | 49 |
| V.1.L'Etendue : | 49 |
| V.2.L'écart absolu moyen : | 50 |
| V.2.1L'écart absolu moyen par rapport à la moyenne arithmétique X..... | 50 |
| V.2.2.L'écart des moyen par rapport à la médiane | 51 |
| V.3.L'écart type et variance : | 51 |
| V.3.1.La variance :..... | 51 |
| V.3.3.Détermination de L'écart type et variance par calcule : | 51 |
| V.4.Le coefficient de la variation de Pearson :..... | 54 |
| V.5.L'intervalle interquartile | 54 |
| V.6.Propriétés de la moyenne arithmétique et de la variance..... | 55 |
| V.6.1.Propriétés de la moyenne arithmétique | 55 |
| V.6.2.Propriétés de la variance et de l'écart-type..... | 55 |
| Conférence n°08 | 56 |
| VI. Les moments :..... | 57 |
| VI.1. Moment simple..... | 57 |
| VI.1. Moment centré..... | 57 |
| Conférence n°09 | 59 |
| VII. Caractéristiques de forme | 60 |
| VII.1.Asymétrie : | 60 |
| VII.2 APLATISSEMENT | 62 |
| Exercices résolues | 65 |
| Solutions des exercices | 68 |
| Conférence n°10 | 81 |
| VIII. Distribution à deux dimensions | 82 |

| | |
|--|------------|
| VIII.1.Présentation en tableau | 83 |
| VIII.2.Distribution marginal et indépendance statistique | 83 |
| VIII.3.Indépendance statistique | 84 |
| VIII.4.Paramètres d'une distribution à deux dimensions : | 84 |
| VIII.4.1.Moyennes marginales | 84 |
| VIII.4.2.Variance marginales | 84 |
| VIII.4.3.Covariance | 84 |
| Conférence n°11..... | 86 |
| IX. Corrélation et régression linéaire | 87 |
| IX.1. La théorie de la corrélation..... | 87 |
| IX.2.Régression linéaire simple : | 89 |
| IX.2.1.Détermination de la droite de régression | 89 |
| IX.2.2.Détermination de la corrélation : | 90 |
| IX.2.3. La covariance entre X et Y ($cov(xy)$) : | 92 |
| Solution des exercices | 96 |
| Exercices proposées | 100 |
| Conférence n°12..... | 102 |
| X. Caractéristique de concentration : | 103 |
| X.1.Valeur globales : | 103 |
| X.1.1Valeurs globales et valeurs globales relatives | 104 |
| X.1.2.Valeurs globales relatives cumulées croissantes : | 104 |
| X.2.Médiale | 106 |
| X.2.1. Détermination graphique : | 106 |
| X.2.2. Détermination algébrique : | 107 |
| Conférence n°13..... | 110 |
| X.3. LA COURBE DE LORENTZ (COURBE DE CONCENTRATION) : | 111 |
| X.3.1Construction de la courbe de Lorentz : | 111 |
| X.3.2.L'INDICE DE CONCENTRATION OU COEFFICIENT DE GINI | 113 |
| Conférence n°14..... | 116 |
| XI. Les indices élémentaires : | 117 |
| XI.1.Condition de réversibilité | 118 |
| XI.2.Condition de transférabilité o(u de circularité) | 118 |
| XII. Les indices synthétiques | 119 |
| XII.1.INDICES DE LASPEYRES ET DE PAASCHE..... | 120 |
| XII.2. L'indice de Fisher | 123 |
| Exercices résolues | 124 |

| | |
|-------------------------------------|------------|
| Solution des exercices | 125 |
| Exercice proposées..... | 130 |
| Bibliographie..... | 131 |

Liste des tableaux

| | |
|--|-----------|
| Tableau 1.1 : tableau théorique de dénombrement..... | 05 |
| Tableau 1.2 : la Distribution Statistique du nombre de pièces par logement..... | 08 |
| Tableau 1.3 : calcul des fréquences cumulées pour le nombre de pièces par logement | 08 |
| Tableau 1.4 : calcul des fréquences (effectifs) cumulées croissantes..... | 08 |
| Tableau 1.5 : La distribution de la population d'une ville donnée selon l'âge..... | 10 |
| Tableau 1.6 : ménage selon le nombre de personnes au recensement | 13 |
| Tableau 1.7 : Fréquence cumulée croissante en % des ménages selon le nombre de personnes..... | 13 |
| Tableau 1.8 : des résultats d'une enquête concernant les loyers mensuels des appartements..... | 16 |
| Tableau 1.9 : Les montants des ventes réalisées par une librairie..... | 17 |
| Tableau 2.1 : Nombre de jours d'absences des étudiants..... | 30 |
| Tableau 2.2 : Revenu agricole des exploitations individuelles..... | 30 |
| Tableau 2.3 : Le nombre de pièces par logement..... | 31 |
| Tableau 2.4 : Les résultats obtenus en mathématiques dans un centre d'examen..... | 32 |
| Tableau 2.5 : Calcul de la moyenne arithmétique (cas V.S.D)..... | 36 |
| Tableau 2.6 : Calcul de la moyenne arithmétique (cas V.S.C)..... | 38 |
| Tableau 2.7 : les taux de croissance annuels du PIB..... | 41 |
| Tableau 2.8 : Vitesse moyenne des camions..... | 42 |
| Tableau 2.9 : Comparaison des indicateurs de position..... | 43 |
| Tableau 2.10 : Calcul du quartile (Cas V.S.D)..... | 46 |
| Tableau 2.10 : Calcul du quartile (Cas V.S.C)..... | 47 |
| Tableau 3.1 : Calcul de l'étendu (Cas V.S.D)..... | 50 |
| Tableau 3.2 : Calcul de l'étendu (Cas V.S.C)..... | 50 |
| Tableau 3.4 : Calcul de la variance (Cas V.S.D)..... | 52 |
| Tableau 3.5 : Calcul de la variance (Cas V.S.C)..... | 52 |
| Tableau 4.1 : Calcul des moments..... | 57 |
| Tableau 4.2 : Calcul des moments centrés..... | 58 |
| Tableau 4.3 : Autre méthode du calcul des moments centrés..... | 63 |

| | |
|---|------------|
| Tableau 5.1 : Méthodes d'analyse selon la nature du variable..... | 82 |
| Tableau 5.2 : Tableau de données ponctuelles..... | 83 |
| Tableau 5.3 : Tableau à double entrée (ou de contingence)..... | 83 |
| Tableau 5.4 : Tableau à double entrée (ou de contingence) de X et Y..... | 84 |
| Tableau 5.5 : Distribution marginale de X..... | 85 |
| Tableau 5.6 : Distribution marginale de Y..... | 85 |
| Tableau 5.7 : Les critères de linéarité et de corrélation..... | 87 |
| Tableau 6.1 : Les ventes et les dépenses publicitaires pendant 6ans..... | 92 |
| Tableau 6.2 : Calcul intermédiaire pour les coefficients de la droite d'ajustement | 93 |
| Tableau 7.1 : La distribution des salaires nets annuels des agents de l'état..... | 103 |
| Tableau 7.2 : Les valeurs globales, les valeurs globales relatives et les valeurs globales relatives cumulées croissantes..... | 104 |
| Tableau 7.3 : Les salariés d'une entreprise sont répartis selon le salaire mensuel | 115 |
| Tableau 7.4 : Les prix et les quantités de deux produits 1 et 2..... | 121 |
| Tableau 7.5 : Calcul des indices de Laspeyres t/t_0 et de Paasche t/t_0 | 121 |

Liste des figures

| | |
|--|-----|
| Figure 1.1 : Diagramme à bandes (verticales)..... | 12 |
| Figure 1.2 : Diagramme circulaire..... | 12 |
| Figure 1.3 : Diagramme en bâtons..... | 13 |
| Figure 1.4 : Diagramme cumulatif..... | 14 |
| Figure 1.5 : Histogramme et polygone des effectifs..... | 16 |
| Figure 1.6 : Courbes des fréquences cumulées croissantes | 17 |
| Figure 2.1 : Détermination graphique du Mode..... | 31 |
| Figure 2.2 : Détermination graphique de la médiane..... | 35 |
| Figure 2.3 : Diagramme en boîte..... | 45 |
| Figure 2.4 : Construction du diagramme en boîte..... | 46 |
| Figure 4.1 : Les formes d'asymétrie d'une distribution statistique | 60 |
| Figure 4.2 : La boîte à moustache pour déterminer la forme d'une distribution..... | 62 |
| Figure 4.3 : Les formes d'aplatissements d'une distribution statistique..... | 63 |
| Figure 6.1 : Linéarité et corrélation..... | 87 |
| Figure 6.2 : La détermination de la droite \hat{Y} | 90 |
| Figure 6.3 : la droite de régression linéaire des ventes Y en fonction des dépenses X..... | 93 |
| Figure 7.1 : L'estimation graphique de la médiane..... | 105 |
| Figure 7.2 : Les salaires des agents : courbes des fréquences cumulées croissantes et des valeurs globales relatives cumulées croissantes | 108 |
| Figure 7.3 : La courbe de concentration..... | 112 |
| Figure 7.4 : Les salaires des agents de l'état : Courbe de concentration..... | 112 |
| Figure 7.5 : Type de concentration..... | 113 |
| Figure 7.6 : La courbe de Lorentz..... | 114 |
| Figure 7.7 : La courbe pour calculer l'indice de Gini | 115 |

Introduction

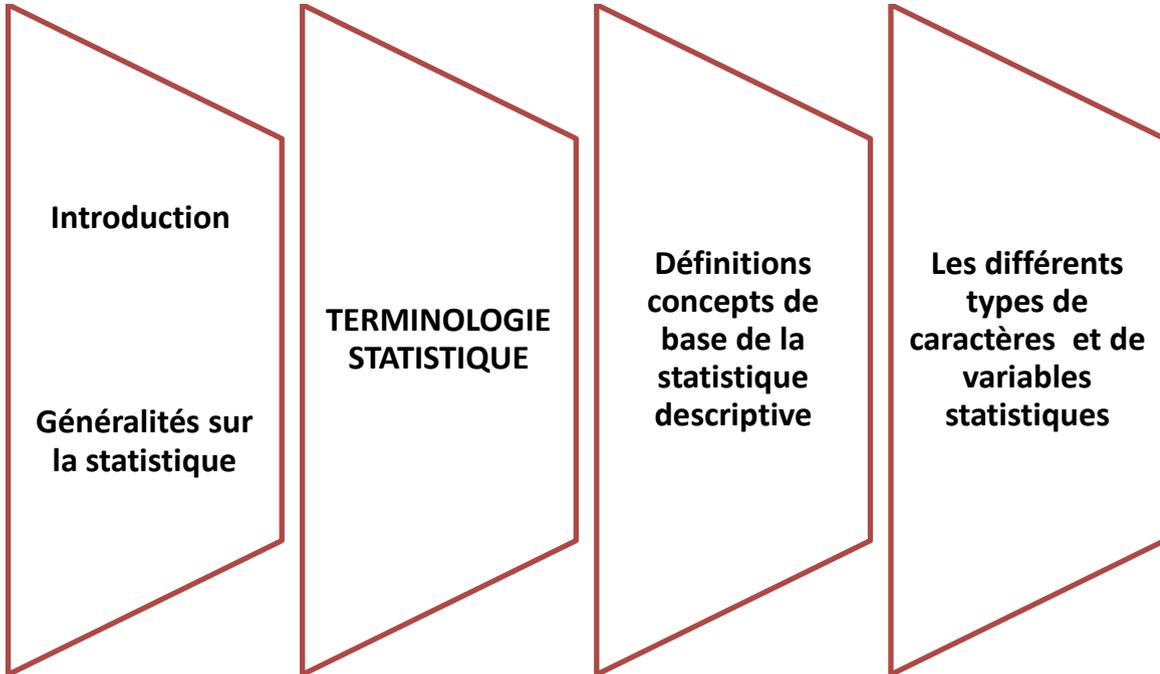
Pour fixer les idées, on commence par donner une définition très générale de la statistique. Ensuite on précise la différence entre statistique descriptive (objet de ce document) et statistique inférentielle (non traitée ici). On termine par la définition d'une dizaine de termes indispensables à la bonne compréhension de la suite de ce cours.

Ce document permet à l'étudiant de découvrir tout d'abord les différents vocabulaires de la statistique et leurs synonymies, et par la suite, l'initier à l'organisation des différentes données univariées générées lors d'une expérimentation sur terrain ou une base donnée existante, le calcul des principaux paramètres et la représentation sous forme de tableaux ou graphiques des résultats, afin de synthétiser l'information.

Le polycopié est structuré selon un enchaînement logique de plusieurs sections, dont la première est consacrée à la définition de la statistique et sa composante descriptive. Ensuite, il a été expliqué la signification des différents concepts et notions adoptées dans la démarche statistique, suivi d'un détail sur la nature des caractères étudiées en statistique et les différentes techniques d'organisation et de structuration d'une série de données. Ce polycopié expose également les différents paramètres descriptifs d'une série statistique simple à savoir les paramètres de position, les paramètres de dispersion et les paramètres de forme d'une distribution de fréquences, puis une analyse statistique bi-variées pour étudier le lien existant entre deux variables et à la fin nous aborderons les caractéristique de concentration qui a pour objectif de mettre en évidence et de mesurer des éventuelles inégalités de répartition. Elle doit permettre en particulier de compléter l'analyse de la « dispersion relative » d'une distribution. Les domaines d'application sont nombreux en économie : concentration des salaires, concentration des revenus ou du patrimoine, de l'emploi, concentration économique d'un secteur. Après cette dernière section on abordera les indices élémentaires et synthétiques qui sert à évaluer des évolutions dans le temps de grandeurs représentatives de phénomènes économiques et sociaux tels que les prix, la production, le chômage, les salaires.

Chaque section est accompagnée d'un ensemble d'exemples explicatifs et d'exercices résolus et proposés.

Conférence n°01



I. Généralités sur la statistique

Définition :

Il n'est pas commode, dans une introduction, de donner une définition précise du concept de statistique, alors que son contenu sera en partie élaboré dans la suite de ce cours. Nous nous contenterons donc, pour fixer les idées, d'en donner une définition volontairement assez vague.

L'origine du mot « statistique » remonte au latin classique « status » (état) qui, par une série d'évolutions successives, aboutit au mot français « statistique », attesté pour la première fois en 1771.¹

C'est vers la même époque que « statistik » apparut en allemand, alors que les anglophones utilisaient l'expression « political arithmetic » jusqu'en 1798, date à laquelle le mot « statistics » fit son entrée dans cette langue.

A l'origine, cette discipline concernait exclusivement les affaires de l'Etat. Actuellement, on distingue généralement les statistiques (au pluriel) de la statistique (au singulier).

Il faut distinguer les données statistiques qui sont les résultats d'observations recueillies lors de l'étude d'un phénomène, et la méthode statistique qui a pour objet l'étude rationnelle des données.

I.1 Les statistiques : représente un ensemble de renseignements numériques découlant des recensements de population ; des données de registres d'état civile et d'enquêtes appropriées. Les statistiques économiques sont par exemple celles relative au prix aux chiffre d'affaires d'entreprise au taux de croissance de l'économie.

I.2 La statistique : L'ensemble de techniques et de méthodes visant au rassemblement à la présentation et à l'analyse des données quantitative. Elle permet de tirer des conclusions logiques, mais aussi de prendre de la décision à partir des analyses effectuées.

La statistique descriptive univariée est un premier niveau d'analyse incontournable. Elle permet de prendre contact avec les données. Les informations sont présentées à partir de tableaux, de diagrammes, et d'indicateurs de tendance et de dispersion qui synthétisent les résultats.

On notera que les termes de statistique descriptive, statistique exploratoire et analyse des données sont quasiment synonymes.

La statistique inférentielle : Ce terme regroupe les méthodes dont l'objectif principal est de préciser un phénomène sur une population globale, à partir de son observation sur une partie restreinte de cette population (penser aux sondages). D'une certaine manière, il s'agit donc d'inférer du particulier au général. Le plus souvent, ce passage ne pourra se faire que moyennant des hypothèses de type probabiliste.

¹ http://www.astro.ulg.ac.be/cours/magain/STAT/Stat_Main_Fr/Download/StatChap1.pdf

Les termes de statistique inférentielle, statistique mathématique et statistique inductive sont eux aussi quasiment synonymes.

II.TERMINOLOGIE STATISTIQUE

II.1.Définitions concepts de base de la statistique descriptive.

II.1.1.Population :

Désigne un ensemble d'individus concernés lors de l'étude d'un sujet particulier. la population étudiée doit être défini avec précision

Exemple : la population d'une ville nous sommes tenues de préciser (inclus les étudiants internes ; les malades de l'hôpital ; etc.

II.1.2.Individu :

on appelle un individu chaque élément de la population ou de l'échantillon on utilise le terme (Unité Statistique) pour désigner un individu.

II.1.3.caractère ou variable :

Le caractère est l'aspect particulier que l'on désire l'étudier, nous décrire les individus qui composent une populations on un échantillon, nous pouvons nous intéresser à leur âge ; à leur sexe et à leur taille ; nationalité.

Exemple : chiffre d'affaire d'une entreprise.

II.1.4.Modalité : sont les différentes situations dont lesquels les individus peuvent se trouver à l'égard du caractère considéré.

Exemple : sexe à deux modalités (F.M)

Nombre d'enfants pour famille 0,1,2,3

II.1.5.série statique :

On appelle série statistique l'ensemble des différents données associés à un certains nombres d'individus.

II.2.Les différents types de caractères et de variables statistiques.

II.2.1.caractère qualitatif : on dit d'un caractère qu'il est qualitatif si les modalités ne s'expriment pas en nombre Ex : religion, sexe, etc.

Les modalités de caractère qualitatif peuvent se présente sous deux formes nominale ou ordinale.

II.2.1.1.la forme qualitative nominale : quand ses modalités ne peuvent pas se ranger dans une graduation logique c'est-à-dire se range qu'à travers des modalités entre lesquelles il n'existe aucune relation d'ordre.

Car on peut exprimer leurs modalités par bleu-vert noir ou noir, l'ordre n'est pas important.

II.2.1.2.La forme qualitative ordinale : possibilité de positionner les modalités entre eux selon la valeur attaché à leur caractère donc il sera possible de ranger dans une graduation logique. D'une façon générale, elle fait référence à des caractère non mesurables mais dont on soit que les modalités fait référence à une relation d'ordre.

Exemple : catégorie socioprofessionnelle.

II.2.2.caractère quantitatif : on dit d'un caractère qu'il est quantitatif si ses modalités sont mesurables c'est-à-dire s'expriment par un nombre.

Le caractère quantitatif : ses diverses modalités sont mesurables c'est à dire : possibilité de faire correspondre un nombre à chaque modalité (poids, âge,.....) ce nombre qui varie d'une modalité à l'autre est appelé variable statistique, celle-ci peut être discrète ou continue.

II.2.2.1.La Variable Statistique (VS) discrète noté Xi (caractère. quantitatif. discret)

Une V.S discrète lorsqu'elle ne peut prendre que des valeurs isolées souvent des nombres entiers

Exemple : le nombre d'enfant par ménage, le nombre de pièce par logement

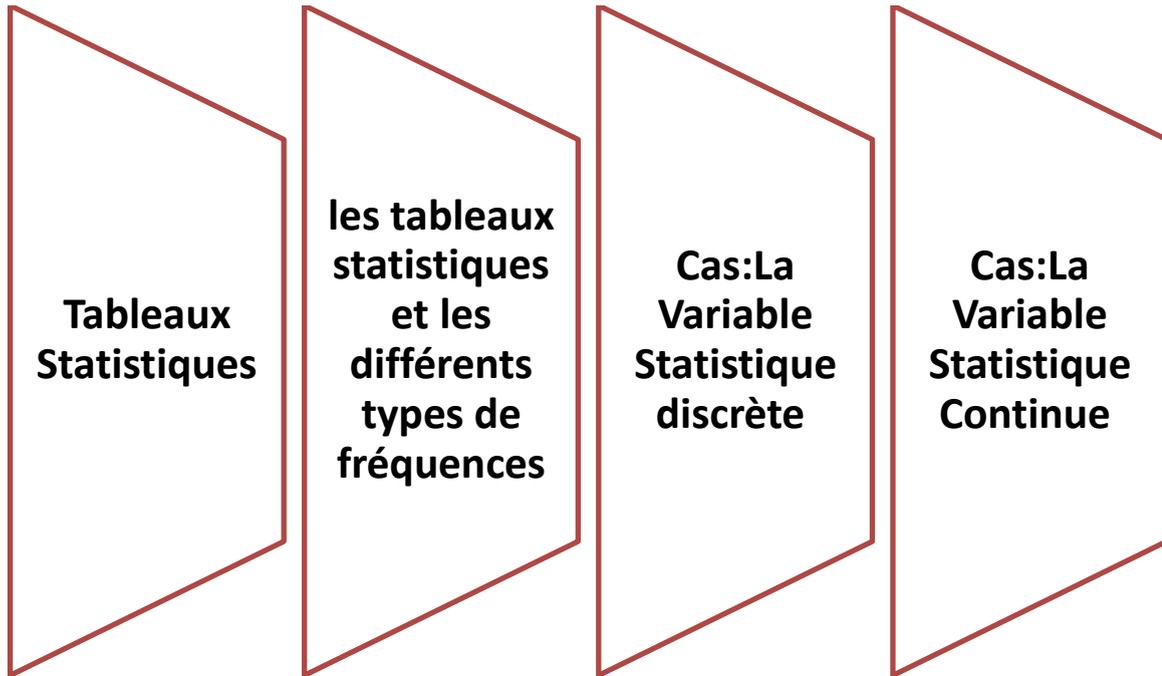
II.2.2.2.La Variable Statistique (V.S) continue note X (caractère. quantitatif. continue) .

La V.S continue peut prendre toute les valeurs appartenant à un intervalle de variation ou encore lorsque les valeurs prise par celle-ci sont en nombre indéterminé.

Exemple : la taille des nouveaux nés, l'âge d'un individu, vitesse....

Conférence n°02

DISTRIBUTIONS STATISTIQUES A UN CARACTERE



III. DISTRIBUTIONS STATISTIQUES A UN CARACTERE

III.1. Tableaux et représentations graphiques

III.1.1 les tableaux statistiques et les différents types de fréquences.

On appelle donc tableau statistique un tableau dont la première colonne comporte l'ensemble des k observations distinctes de la variable X. Ces observations sont rangées par ordre croissant et non répétées; nous les noterons $\{x_i ; i = 1, \dots, k\}$. Dans une seconde colonne, on dispose, en face de chaque valeur x_i , le nombre de réplifications qui lui sont associées. Ces réplifications sont appelées effectifs et notées n_i

Notation :

X : le caractère ; x_i : Les modalités ; N : le nombre total de population ; K : le nombre totale de modalités.

Considérant une population statistique composée de N unités statistiques décrites suivant X, constitué de K modalité (x_1, x_2, \dots, x_k) .

Nous pouvons construire le tableau théorique suivant quel soit la nature du caractère qualitatif on quantitatif.

Tableau 1.1 : tableau théorique de dénombrement

| | | |
|-----------------------------|----------------------------------|-------|
| Modalité X_i du caractère | $X_1, X_2 \dots \dots \dots X_k$ | Total |
| Nombre d'individus X_i | $n_1, n_2 \dots \dots \dots n_k$ | N |

Les observations ordonnées forme une distribution ou série statistique, lorsque est constituée de l'ensemble des modalités et des effectifs d'un caractère

$$D.S = \{X_i, n_i\}$$

L'effectif total de la population statistique est désigné par N

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

n_i désignée effectif

Les fréquence absolue se définit comme étant le nombre d'unités statistique ou d'individus correspondant à chaque modalité du caractère.

III.1.1.1. La Variable Statistique discrète : les effectifs de n_i correspondant a des valeurs isole x_i de la V.S X les quelles sont ordonnées dans le tableau de manière croissant ou décroissante)

Exemple : Etant donné la Distribution Statistique du nombre de pièces par logement

Tableau 1.2 : la Distribution Statistique du nombre de pièces par logement

| | | | | | | |
|--|----|----|---|----|----|-------|
| <i>nombre de pieces x_i</i> | 1 | 2 | 3 | 4 | 5 | Total |
| <i>effectif n_i</i> | 10 | 15 | 5 | 20 | 25 | 75 |
| Nombre de logement | | | | | | |

Le nombre de fois qu'une valeur x_i de la V.S X apparait est appelé effectif de cette valeur par exemple, le nombre de logement de 3 pièces apparait 5 fois.

Exemple : Soit la distribution statistique du nombre de pièces par logement

Tableau 1.3 : calcul des fréquences cumulées pour le nombre de pièces par logement

| | | | | | |
|------------------|------|-----|------|-----|-------|
| Nombre de pièces | 1 | 2 | 3 | 4 | Total |
| x_i | | | | | |
| n_i | 15 | 20 | 25 | 40 | 100 |
| f_i | 0.15 | 0.2 | 0.25 | 0.4 | 1 |
| $F_i\%$ | 15% | 20% | 25% | 40% | 100% |

On dit que 25% de logement ont 3 pièces.

La notion de fréquence cumulée permet de répondre à la question combien de logement ont plus de 2 pièces ou moins de 3 pièces.

La fréquence relative cumulé se définit comme étant la somme (Σ) des fréquences relatives

$$F(x_i) = f_1 + f_2 + \dots + f_k = \sum_{i=1}^k f_i$$

La fréquence absolue (n_i) ou relatives (f_i) s'additionnent successivement jusqu'à d'obtenir leur somme totale.

Tableau 1.4: calcul des fréquences (effectifs) cumulées croissantes

| | | | | | | | |
|-------|-------|-------|----------|------------------|--------------------|----------------------------|----------------------------|
| n_i | n_i | f_i | $F(x_i)$ | $N(n_i)\uparrow$ | $N(n_i)\downarrow$ | $f(n_i)\uparrow\downarrow$ | $f(n_i)\uparrow\downarrow$ |
| 1 | 15 | | | 0 | 180 | 0 | 1 |
| 2 | 20 | | | 15 | | 0.15 | |
| 3 | 25 | | | 35 | | | |
| 4 | 40 | | | | 0 | 1 | 0 |

On obtient les valeurs successives $F(x_i)$ sur le tableau statistique de la variable en cumulant progressivement les fréquences f_i ces valeurs sont écrites entre les lignes pour bien montrer qu'elles

rappellent à l'intervalle séparant deux valeurs cette fonction $F(x)$ est définie par $F(x) = \sum_{i=1}^k f_i$, pour tout $x \in [x_k; x_{k+1}]$

La fonction F est nulle pour tout $x \leq$ à la plus petite valeur possible de x_k et $=1$ pour tout $x >$ à la plus grande valeur possible x_{k+1}

$$\forall x < x_1, \text{ on a } F(x) = 0 \text{ et } \forall x > x_j \text{ on a } F(x) = 1 \quad F(-\infty) = 0 ; F(+\infty) = 1$$

III.1.1.2. La variable statistique continue :

Les modalités du caractère sont constituées de classe de valeurs limitées par deux extrémités e_i, e_{i+1}
Généralement ces classes sont ordonnées de manière croissante.

Une variable quantitative continue peut prendre une infinité de valeurs possibles. Le domaine de la variable est alors \mathbb{R} ou un intervalle de \mathbb{R} . En pratique, une mesure est limitée en précision. La taille peut être mesurée en centimètres, voire en millimètres. On peut alors traiter les variables continues comme des variables discrètes. Cependant, pour faire des représentations graphiques et construire le tableau statistique, il faut procéder à des regroupements en classes.

Le tableau regroupé en classe est souvent appelé distribution groupée. Si $[e_i, e_{i+1}[$ désigne la classe i , on note, de manière générale :

e_i : la borne inférieure de la classe i

e_{i+1} : la borne supérieure de la classe i

$c_i = (e_i + e_{i+1})/2$: le centre de la classe i

$a_i = e_{i+1} - e_i$: l'amplitude de la classe i

n_i : l'effectif de la classe i

$N(e_i)$: l'effectif cumulé de la classe i ,

f_i : la fréquence de la classe i ,

$F(e_i)$ la fréquence cumulée de la classe i .

La répartition en classes des données nécessite de définir a priori le nombre de classes k et donc l'amplitude de chaque classe. En règle générale, on choisit au moins cinq classes de même amplitude jusqu'au 15. Cependant, il existe des formules qui nous permettent d'établir le nombre de classes et l'intervalle de classe (l'amplitude) pour une série statistique de n observations.

– La règle de Sturge : $k = 1 + (3.3 \log_{10}(n))$.

– La règle de Yule : $k = 2,5 \sqrt[4]{n}$.

L'intervalle de classe est obtenue ensuite de la manière suivante : longueur de l'intervalle = $(x_{\max} - x_{\min})/k$, où x_{\max} (resp. x_{\min}) désigne la plus grande (resp. la plus petite) valeur observée.²

Remarque : Il faut arrondir le nombre de classe k à l'entier le plus proche. Par commodité, on peut aussi arrondir la valeur obtenue de l'intervalle de classe.

A partir de la plus petite valeur observée, on obtient les bornes de classes en additionnant successivement l'intervalle de classe (l'amplitude).

Exemple : La distribution de la population d'une ville donnée selon l'âge la V.S désigne l'âge et l'effectif n_i désigne le nombre de personnes

Tableau 1.5 : La distribution de la population d'une ville donnée selon l'âge

| Age n_i | Effectif (n_i) |
|-----------|--------------------|
| 0-5 ans | 100 |
| 5-10 ans | 200 |
| 10-20 ans | 400 |
| 20-30 ans | 300 |
| 30-40 ans | 500 |
| 40 et + | 300 |
| | 1800 |

La fréquence absolue de la classe d'âge [10.20] est égale à 400, cela signifie que il ya 400 personnes ou individués dont leur âge est comprise dans cette classe.

On désigne par e_1, e_2, \dots, e_k les extrémités de classe, dans l'exemple précédent $e_1 = 0, e_2 = 5, \dots$

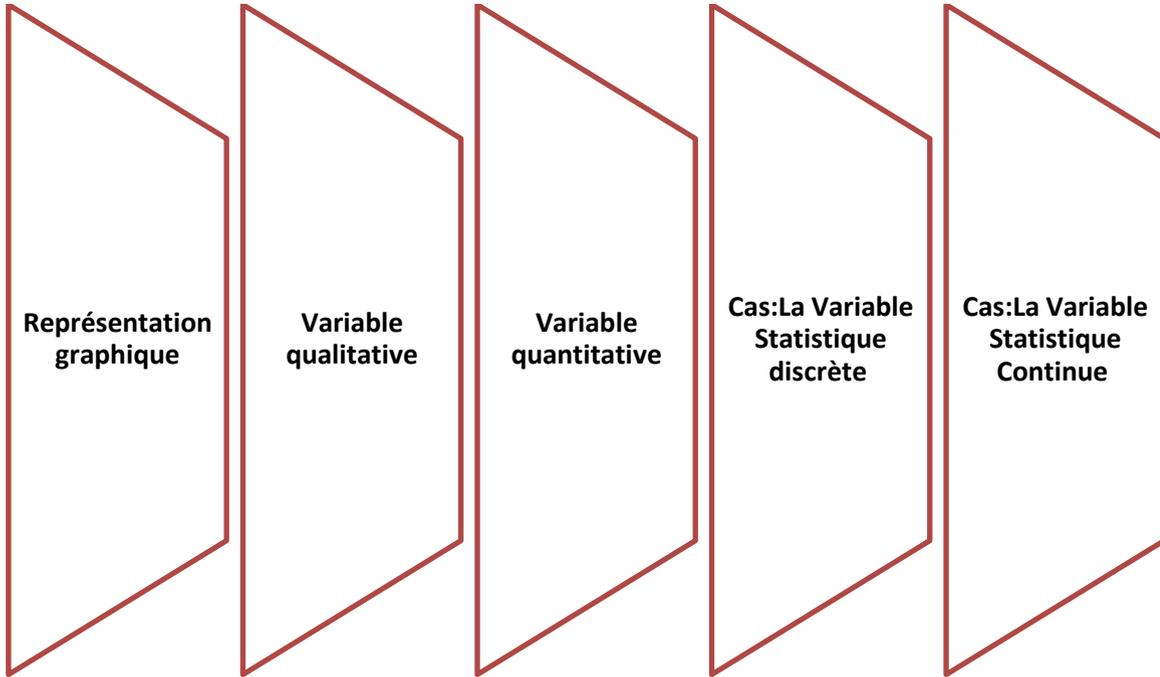
Les effectifs sont strictement inférieurs à l'extrémité supérieure mais égaux à l'extrémité inférieure, on écrit $e_i \leq n_i < e_{i+1}$

- **Les fréquences relatives**

La fréquence relative notée f_i est égale au rapport des individus n_i présentant la modalité x_i du caractère X sur l'effectif total N , on a donc $f_i = \frac{n_i}{N} \times 100$

² <http://mathsv-ressources.univ-lyon1.fr/cours/stats/chap5/c5p4/c5p4.html>

Conférence n°03



III.1.2 Représentation graphique

Organisation des données sous forme de tableaux permet de structurer l'information statistique collectée. L'objet d'une représentation graphique est de proposer une image plus élaborée et plus synthétique de l'ensemble des observations.

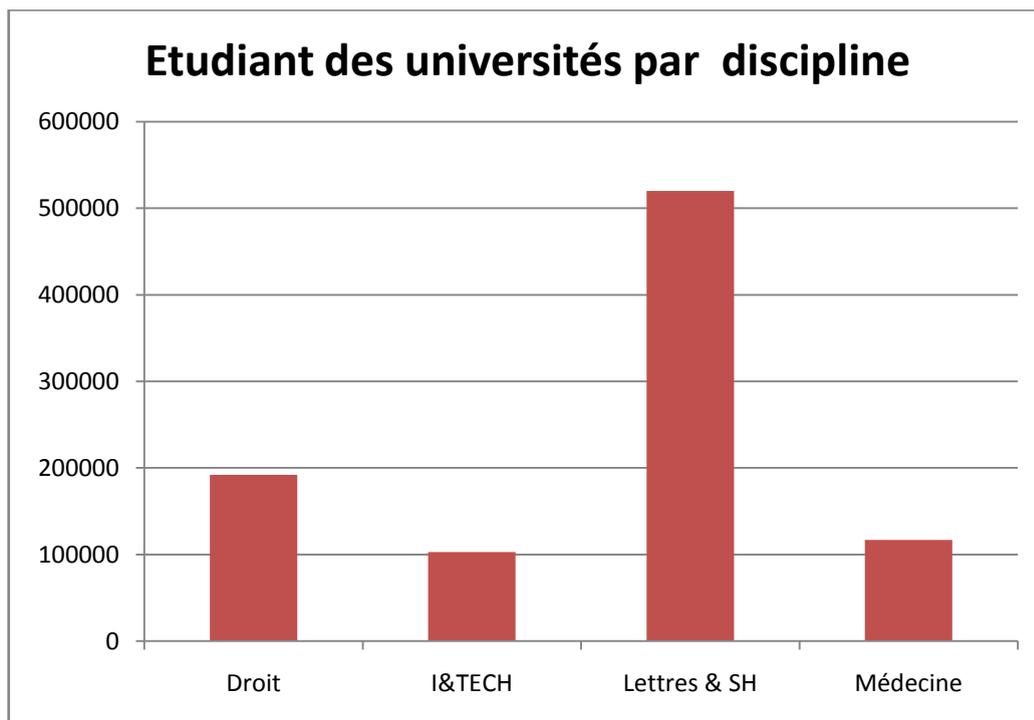
En fonction de la nature de la variable, plusieurs modèles de représentations sont possibles.

III.1.2.1. Variable qualitative

III.1.2.1.1. Diagrammes à bandes

On appelle diagramme à bandes(ou diagramme à barres) un graphique qui à chaque modalité de la variable associé un rectangle (de base constante) dont la hauteur est proportionnelle à l'effectif

Figure1.1 : Diagramme à bandes (verticales).



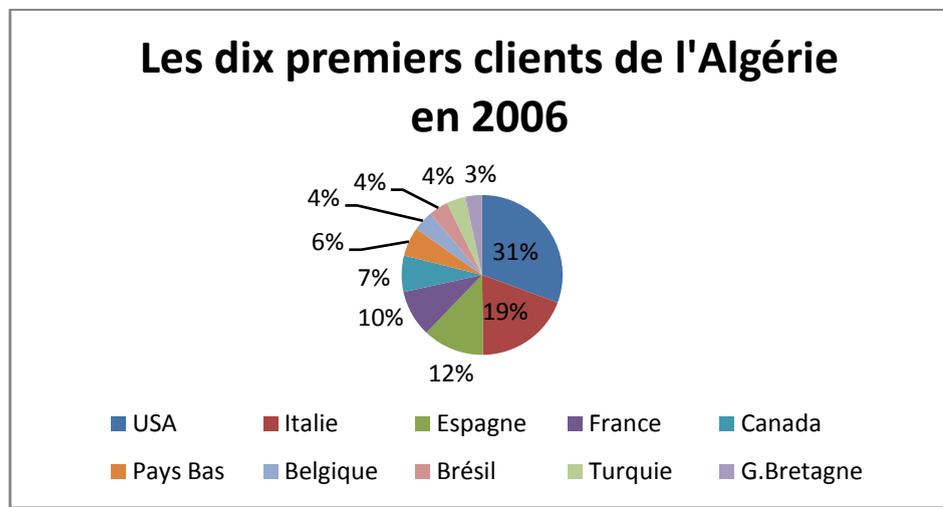
III.1.2.1.2. Diagramme circulaire

Pour une modalité donnée n_i , d'effectif f_i , l'angle au centre α_i correspondant est donnée en degrés par ;

$$\alpha_i = \frac{n_i}{n} \times 360 = f_i \times 360 = 3.6 \times f_i (\%)$$

Exemple : Les dix premiers clients de l'Algérie en 2006

Figure 1.2 : Diagramme circulaire



III.1.2.2. Variable quantitative :

III.1.2.2.1. Cas Variable statistique discrète

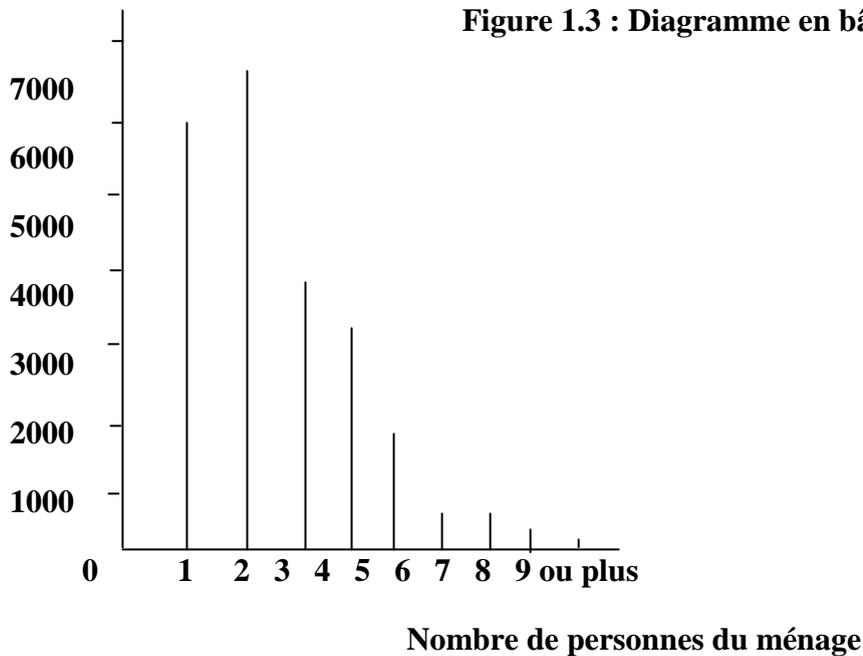
* **Diagramme en bâtons :**

On appelle diagramme en bâtons un graphique en segment (bâton) dont le hauteur est proportionnelle à l'effectif.

Tableau 1.6 : ménage selon le nombre de personnes au recensement

| Nombre de personnes du ménage | Nombre de ménage (effectif en millier) | Fréquence (%) |
|-------------------------------|--|---------------|
| 1 | 5845 | 27,1 |
| 2 | 6369 | 92,6 |
| 3 | 3822 | 17,7 |
| 4 | 3371 | 15,6 |
| 5 | 1439 | 6,7 |
| 6 | 424 | 2,0 |
| 7 | 153 | 0,7 |
| 8 | 65 | 0,3 |
| 9 ou plus | 54 | 0,3 |
| Total | 21542 | 100 |

Figure 1.3 : Diagramme en bâtons



*** Diagramme cumulatif :**

On appelle fonction de répartition d'une variable X , et on note F l'application de l'ensemble R dans l'intervalle $[0,1]$ qui pour une valeur donnée n de la variable X , précise la proportion des individus de la population pour lesquels la valeur de la variable est inférieur (strictement) à n

$F(n) = P(X < n)$ [proportion d'individus de la population pour lesquels la variable X présente une valeur inférieur strictement à n]

On appelle diagramme cumulatif (ou courbe cumulative) la courbe représentative de la fonction de répartition.

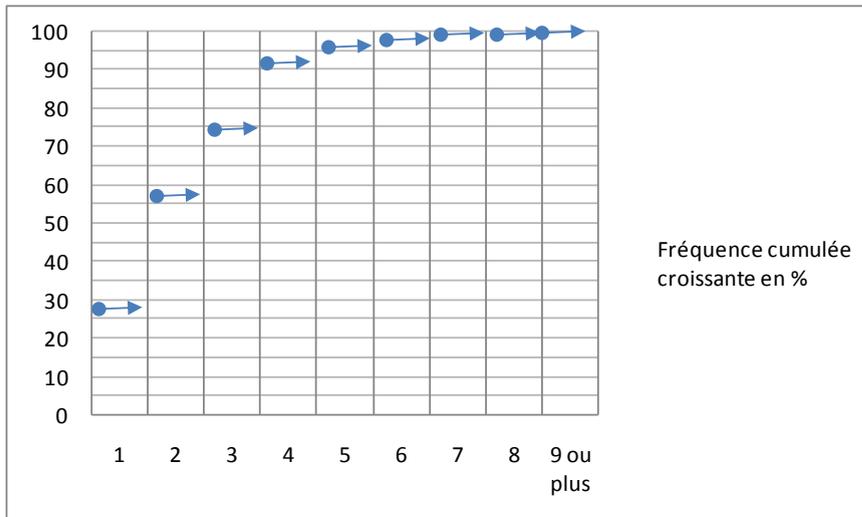
En pratique le diagramme cumulatif est obtenu à partir des fréquences cumulées croissantes.

Lorsque la variable est quantitative discrète, la courbe cumulative se présente, comme une courbe en escalier est ouvert à gauche et fermé adroite (sauf le dernier).

Tableau 1.7 : Fréquence cumulée croissante en % des ménages selon le nombre de personnes

| Nombre de ménage | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 ou plus |
|-----------------------------------|------|------|------|------|------|------|------|------|-----------|
| Fréquence cumulée croissante en % | 27,1 | 56,7 | 74,4 | 90,1 | 96,8 | 98,7 | 99,4 | 99,7 | 100 |

Figure 1.4 : Diagramme cumulatif



Pour tout $X \leq 1$ $F(x)=0$; Pour tout $1 < n \leq 2$ $F(x)=0,271=27,1\%$; (Proportion de ménage comprenant moins de 2 personnes)

Pour tout $8 < n \leq 9$ $f(x)=99,7\%$; Pour tout $9 < n, f(n)=100\%$

III.1.2.2.2.Cas Variable statistique Continue

✓ Histogramme :

On appelle histogramme un diagramme comparé d »un ensemble de rectangle contiguë (adjacents), chaque rectangle, associé à chaque classe ayant une surface proportionnelle à l'effectif (ou à la fréquence) de cette classe. Deux cas peuvent se présenter :

- **Classes d'amplitudes égales :** $(a_i=a)$
- **Classes d'amplitudes égales :** $(a_i \neq a_j)_{i \neq j}$
- **Le polygone des effectifs**

Le polygone des effectifs ou des fréquences : est une autre représentation graphique (en ligne brisée) de la distribution des effectifs ou des fréquences d'une variable quantitative. Pour tracer le polygone, on joint les points milieu du sommet des rectangles adjacents par un segment de droite. Le polygone est fermé aux deux bouts en le prolongeant sur l'axe horizontal.

- **Classes d'amplitudes égales :**

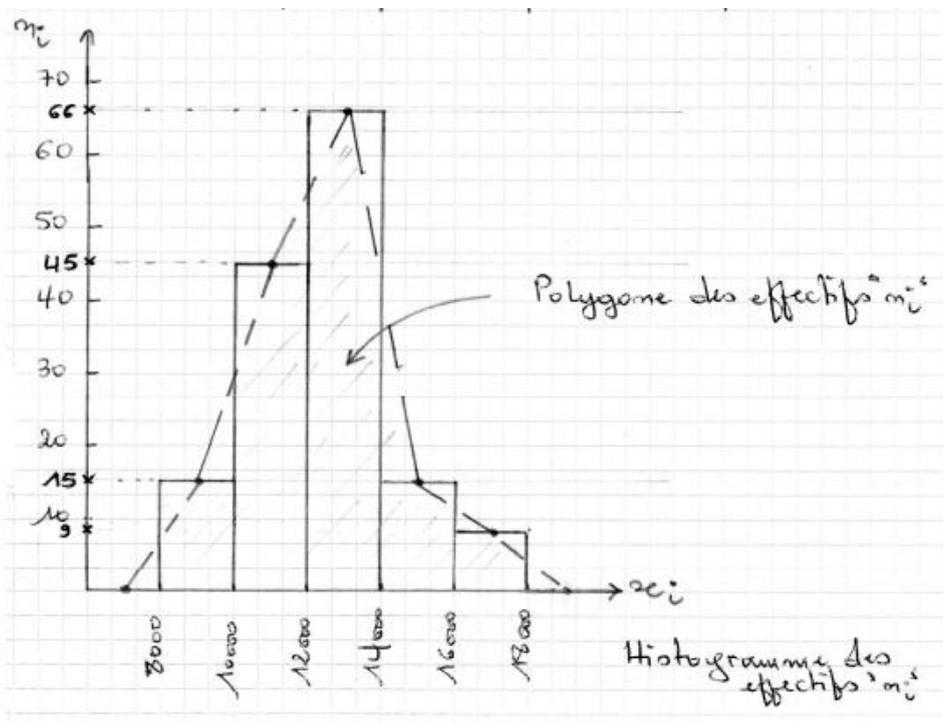
Lorsque les classes sont d'amplitudes égales $(a_i=a)$

On dispose des résultats d'une enquête concernant les loyers mensuels des appartements d'une commune donnée.

Tableau 1.8 : des résultats d'une enquête concernant les loyers mensuels des appartements

| loyers en dinars | [8000; 10000[| [10000; 12000 [| [12000; 14000[| [14000; 16000 [| [16000; 18000 [|
|------------------|---------------|-----------------|----------------|-----------------|-----------------|
| Effectifs | 15 | 45 | 66 | 15 | 9 |

Figure 1.5 : histogramme et polygone des effectifs



• **Classes d'amplitudes inégales : $(a_i \neq a_j)_{i \neq j}$**

Lorsqu'au moins deux classes ont des amplitudes différentes, la hauteur proportionnelle à l'effectif ne permet plus de construire un histogramme. En effet, la surface de chaque rectangle (est plus proportionnelle à l'effectif (conformément à la définition). D'où la nécessité de corriger les fréquences

(ou effectifs). $f_i^c = \frac{f_i}{a_i}$ ou $n_i^c = \frac{n_i}{a_i}$ avec f_i^c : fréquence corrigée; n_i^c : effectif corrigé

Les classes sont d'amplitudes inégales. Pour le traçage de l'historgramme et le calcul du mode, il va falloir corriger les effectifs n_i .

Tableau 1.9 : Les montants des ventes réalisées par une librairie

| Montant en dinars | [1000,1500[| [1000,1500[| [1000,1500[| [1000,1500[| [1000,1500[| [1000,1500[|
|---------------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| n_i | 500 | 3000 | 1500 | 2000 | 1000 | 1000 |
| $n_i^c = \frac{n_i}{a_i}$ | 500/500=1 | 2 | 3 | 2 | 2 | 1 |

Remarque :

Ainsi, la classe modale est [3000, 3500[et le mode se calcule classiquement par sa formule connue. Pour le calcul de la médiane et la représentation par courbe cumulative, on doit utiliser les effectifs n_i et non pas les effectifs corrigés.

Les effectifs corrigés ne sont utilisés que pour le traçage de l’histogramme, le polygone et le calcul du mode.

✓ **Courbes des fréquences cumulées croissantes :**

Définition :

On appelle courbe de fréquences cumulées croissantes la représentation graphique de la fonction de répartition de la variable x .

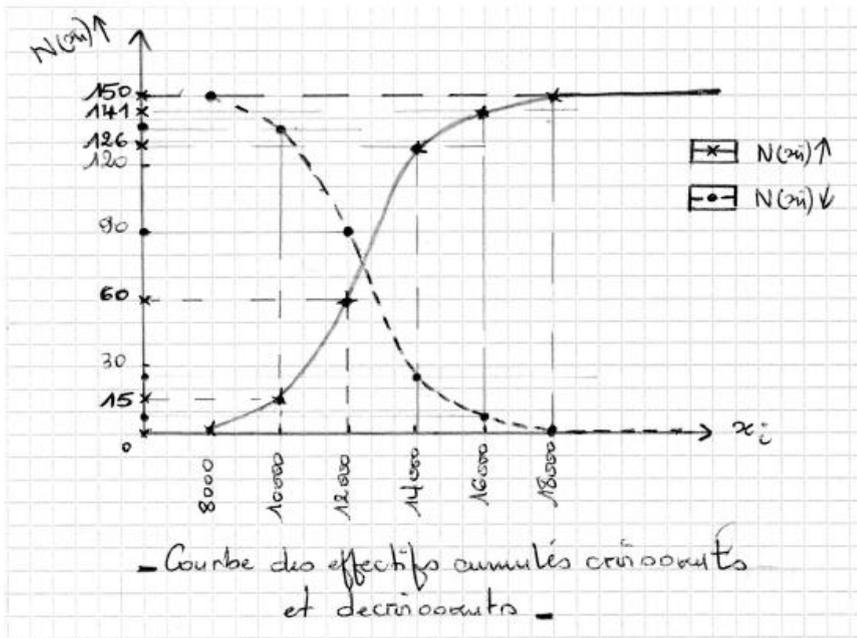
Les données étant groupées en classes, la fréquence cumulée croissante F_i associé à la classe $n^\circ i$ représente la proportion d’individus de la population pour lequel la variable prend une valeur inférieure (strictement) à la limite supérieur e_{i+1} de la classe $n^\circ i$.

En pratique, la courbe des fréquences cumulées croissantes est obtenue en joignant, dans un système d’axes orthogonaux, les points d’abscisse e_{i+1} (extrémité de la classe $n^\circ i$) et d’ordonnée F_i (fréquence cumulée croissante correspondante).

Remarque :

Joindre les points de coordonnées $(e_{i+1}; F_i)$ par des segments revient à faire l’hypothèse d’une répartition uniforme des individus à l’intérieur des classes.

Figure 1.6 : Courbes des fréquences cumulées croissantes



Exercices résolues

Exercice 1.

Déterminer la population statistique, l'unité statistique, la nature du caractère ainsi que la variable statistique dans les cas suivants :

1. Les salaires annuels des enseignants d'une école.
2. Le nombre de salariés d'une entreprise.
3. Le lieu de travail des habitants d'un quartier.
4. Le nombre d'habitants par quartier.
5. Le nombre d'enfants à charge d'une famille.
6. Nombre de personnes employées par l'ensemble des entreprises d'une localité.
7. Le groupe sanguin des employés d'une entreprise.
8. La répartition des enseignants selon l'âge.
9. La distance séparant deux points.
10. La catégorie socioprofessionnelle des clients d'un opérateur téléphonique.
11. La mention au baccalauréat des étudiants de cette promotion.
12. La répartition des variables statistiques selon leur nature.

Exercice 2. Une enquête statistique auprès de N familles a conduit à la répartition suivante selon le nombre d'enfants.

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 7 | 2 | 3 | 5 | 3 | 4 | 8 | 3 | 3 | 4 | 0 | 2 | 6 | 4 | 6 | 1 | 2 | 2 |
| 3 | 6 | 4 | 1 | 0 | 1 | 2 | 2 | 0 | 4 | 6 | 3 | 2 | 3 | 2 | 1 | 8 | 2 | 7 | 3 |
| 2 | 3 | 0 | 3 | 3 | 4 | 1 | 4 | 7 | 2 | 1 | 3 | 2 | 7 | 5 | 0 | 5 | 7 | 4 | 3 |

1. Le caractère est-il qualitatif ou quantitatif ?
2. Déterminer le nombre de familles noté N.
3. Déterminer la population statistique, l'unité statistique et la variable statistique.
4. Trouver les valeurs distinctes de cette variable statistique.
5. Construire le tableau représentatif de cette distribution.

Exercice 3. On a relevé en fin de journée le montant en dinars des chèques reçus dans un magasin.

| | | | | | | | |
|------|--------|------|--------|--------|--------|--------|--------|
| 1270 | 3900 | 4100 | 5400,5 | 1900 | 2800 | 2100 | 4250 |
| 3500 | 7420 | 1760 | 1380,5 | 1200 | 3550,2 | 4720 | 1400 |
| 1702 | 1500,3 | 1700 | 5950 | 7920,5 | 6880,2 | 1000,5 | 6720 |
| 2400 | 2000,1 | 1850 | 2050 | 1750 | 1820,5 | 1050,1 | 1800,1 |

1. Identifier la variable statistique et préciser sa nature.

2. Présenter la distribution des effectifs dans les classes :
 $[0; 2000[$, $[2000; 4000[$, $[4000; 6000[$, $[6000; 8000[$.

Exercice 4. On note par X la variable statistique qui représente le nombre de langues parlées chez 50 employés d'une petite compagnie aérienne. Les résultats sont représentés par le tableau suivant, où n est l'effectif correspondant à x la valeur de X :

| | | | | | |
|--------------------------------------|---|----|----|----|---|
| x (nombre de langues parlées) | 1 | 2 | 3 | 4 | 5 |
| Effectif n | 5 | 10 | 20 | 12 | 3 |

1. Identifier la variable statistique et préciser sa nature.
2. Calculer les effectifs cumulés croissants et décroissants.
3. Calculer les fréquences cumulées croissantes et décroissantes.
4. Tracer le diagramme en bâtons des effectifs n_i . De même pour le polygone.
5. Que représente l'effectif cumulé $N(4)$? De même pour les fréquences cumulées $F(2)$ et $1-F(2)$.
6. Tracer la courbe cumulative des fréquences.
7. Donner le nombre d'employés parlant moins de 4 langues.
8. Dans quelle proportion observe-t-on au plus une langue parlée? De même pour au moins 5.

Exercice 5. Une étude statistique faite sur le prix d'un produit dans 40 magasins, a donné les résultats suivants :

| | | | | |
|-----------------------|-------------|-------------|-------------|-------------|
| Prix en dinars | $[460;480[$ | $[480;500[$ | $[500;520[$ | $[520;540[$ |
| Magasins | 5 | 12 | 14 | 9 |

1. Identifier la variable statistique et préciser sa nature.
2. Combien de magasins vendent le produit entre 480 et 540 dinars?
3. Faire une représentation par histogramme.
4. Représenter le polygone des effectifs.
5. Composez le tableau des effectifs cumulés croissants et décroissants.
6. Tracer la courbe des effectifs cumulés croissants et décroissants.
7. Dans combien de magasins le prix est-il inférieur à 520 dinars?
8. Dans combien de magasins le prix est-il supérieur à 480 dinars?
9. Quel est le pourcentage de magasins qui vendent le produit à moins de 500 dinars?

Exercice 6. Le responsable d'un magasin a relevé au cours d'une période de 10 jours le nombre d'articles vendus. Il a obtenu les résultats suivants,
 300, 100, 500, 300, 100, 1000, 900, 700, 100, 100.

1. Quelle est dans ce cas la population statistique?
2. Que signifie par exemple l'observation 900?

3. Identifier la variable statistique et préciser sa nature.
4. Calculer les fréquences cumulées croissantes et décroissantes.

Exercice 7. On étudie les revenus (mensuels en dinars) d'un ensemble de familles d'un quartier.

| Revenus | [7000;9000[| [9000;11000[| [11000;13000[| [13000;15000[| [15000;17000[|
|-----------|-------------|--------------|---------------|---------------|---------------|
| Effectifs | 30 | 160 | 60 | 70 | 80 |

1. Quel est le nombre de familles dont le revenu est compris entre 7000 et 9000 dinars?
2. Quelle est la proportion des familles dont le revenu est compris entre 9000 et 15000 dinars?
3. Quelle est la proportion de familles dont le revenu est supérieur à 13000 dinars? De même si c'est inférieur à 15000 dinars.

Solution des exercices

Corrigé de l'exercice 01 :

| | Population statistique | Unité Statistique | Caractère | Variable statistique/modalité |
|----|----------------------------|--------------------------|---|--|
| 1 | Les enseignants | Un enseignant | Quantitatif | Variable statistique continue (Salaires) |
| 2 | Les salariés | Un salarié | Quantitatif | Variable statistique discrète (Le nombre de salariés) |
| 3 | Les habitants | Un habitant | Qualitatif (Le lieu) | Modalité nominale |
| 4 | Les quartiers | Un quartier | Quantitatif | Variable statistique discrète (Le nombre d'habitants) |
| 5 | Les enfants | Un enfant | Quantitatif | Variable statistique discrète (Le nombre d'enfants) |
| 6 | Les entreprises | Une entreprise | Quantitatif | Variable statistique discrète (Le nombre de personnes employés) |
| 7 | Les employés | Un employé | Qualitatif (Groupe sanguin) | Modalité nominale |
| 8 | Les enseignants | Un enseignant | Quantitatif | Variable statistique continue (L'âge) |
| 9 | Les deux points | Un seul point | Quantitatif | Variable statistique continue (La distance) |
| 10 | Les clients | Un client | Qualitatif (La catégorie socio-prof) | Modalité nominale |
| 11 | Les étudiants | Un étudiant | Qualitatif (Mention) | Modalité ordinale |
| 12 | Les variables statistiques | Une variable statistique | Qualitatif | Modalité nominale |

Corrigé de l'exercice 2.

1. Il s'agit d'un caractère quantitatif (nombre d'enfants).
2. $N = 60$ familles (chaque case représente une famille + son nombre d'enfants).
3. Population statistique : les 60 familles ; Unité statistique : une famille

Variable statistique : Il s'agit de la variable statistique discrète, le nombre d'enfants

4. Les valeurs distinctes de la variable statistique : 0, 1, 2, 3, 4, 5, 6, 7, 8.

5. Tableau représentatif :

| | | | | | | | | | | |
|-----------------------|---|---|----|----|---|---|---|---|---|-------|
| Nombre d'enfants "xi" | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
| Familles "ni" | 6 | 7 | 12 | 13 | 8 | 3 | 4 | 5 | 2 | N=60 |

Notations :

X : La variable statistique

x i : les valeurs prises par la variable statistique X

n i : les effectifs (ou répétition)

$$N = \sum_i n_i \text{ (l'effectif total)}$$

$$f_i = \frac{n_i}{N} \text{ (Les fréquences).}$$

Exemple : On a 3 familles ayant 5 enfants. Ainsi 5 % de ces 60 familles ont 5 enfants.

Corrigé de l'exercice 3.

1. Variable statistique continue (salaires mensuels).

2. Tableau statistique

| | | | | | |
|--------------------------------|----------|-------------|-------------|-------------|--------------|
| Salaires en dinars "xi" | [0;2000[| [2000;4000[| [4000;6000[| [6000;8000[| Total |
| Employés "ni" | 15 | 8 | 5 | 4 | N=32 |

Corrigé de l'exercice 4.

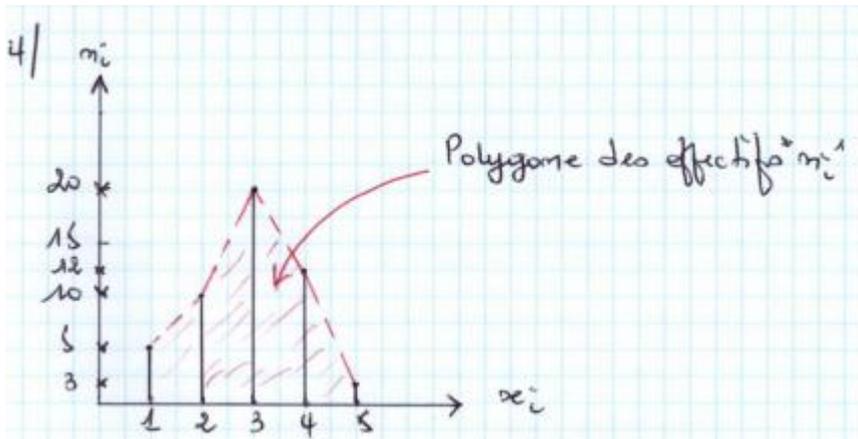
1- La variable statistique X est : « le nombre de langues parlées » sa nature est discrète.

| | | | | | | | |
|---------------------------------------|---------------------|-----|-----|------|------|------------------|----|
| xi (nombre de langues parlées) | 1 | 2 | 3 | 4 | 5 | Total | |
| Effectif ni | 5 | 10 | 20 | 12 | 3 | N=50 | |
| $f_i = \frac{n_i}{N}$ | 0,1 | 0,2 | 0,4 | 0,24 | 0,06 | $\sum_i f_i = 1$ | |
| 2 | $N(x_i) \uparrow$ | 0 | 5 | 15 | 35 | 47 | 50 |
| | $N(x_i) \downarrow$ | 50 | 45 | 35 | 15 | 03 | 0 |
| 3 | $F(x_i) \uparrow$ | 0 | 0,1 | 0,3 | 0,7 | 0,94 | 1 |
| | $F(x_i) \downarrow$ | 1 | 0,9 | 0,7 | 0,3 | 0,06 | 0 |

Pour calculer $N(x_i) \downarrow$ et $F(x_i) \downarrow$ on utilise le fait que :

$$N(x_i) \downarrow = N - N(x_i) \uparrow \text{ et } F(x_i) \downarrow = 1 - F(x_i) \uparrow$$

4- Le digramme en bâtons :



5- Rappel :

Les effectifs (fréquences) cumulés croissants répondent à la question :

Combien d'effectifs (fréquences) ont moins de...

Les effectifs (fréquences) cumulés décroissants répondent à la question :

Combien d'effectifs (fréquences) ont plus de...

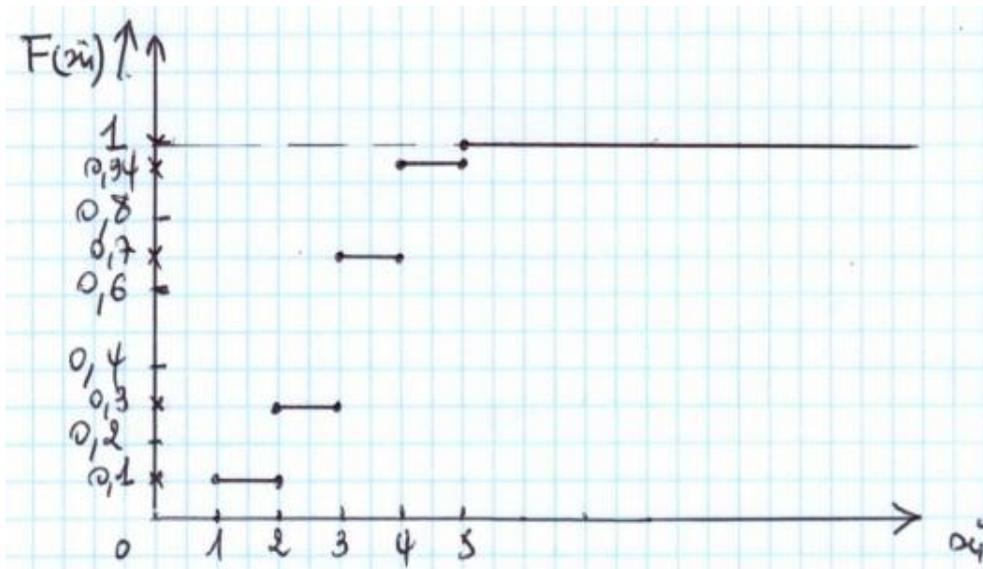
Ainsi :

$N(4) = N(4) \uparrow = 47 \Rightarrow 47$ employés parlent moins de 5 langues ou 47 employés parlent au plus de 4 langues

$F(2) = F(2) \uparrow = 0,3 \Rightarrow 30\%$ des employés parlent moins de 3 langues ou 30% des employés parlent au plus de 2 langues

$1 - F(2) = 1 - F(2) \uparrow = F(2) \downarrow = 0,7 \Rightarrow$ donc 70% des employés parlent plus de 2 langues (ou 70% des employés parlent au moins 3 langues

6. Traçage de la courbe cumulative des fréquences.



7. moins de 4 langues : (5+10+20=35).

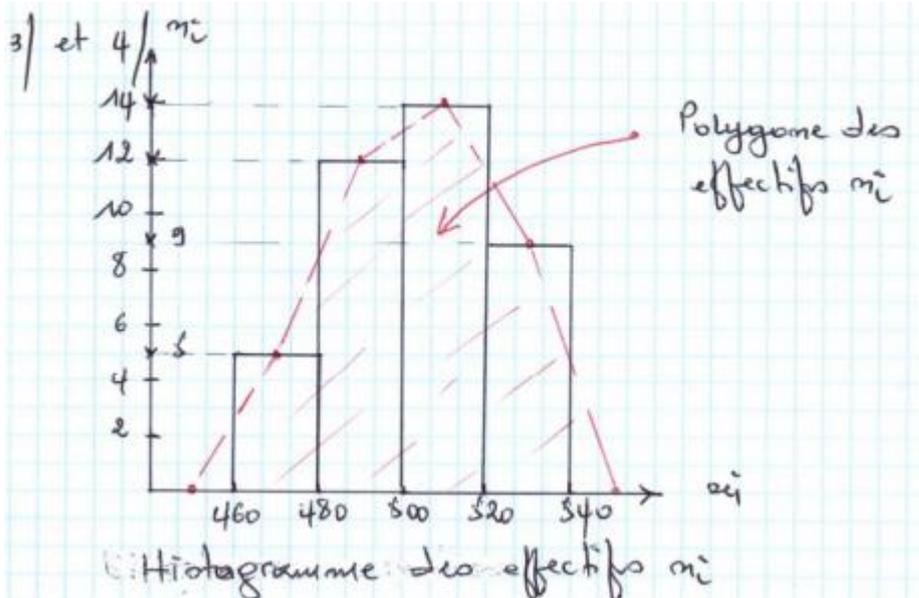
35 employés parlent moins de 4 langues ou $(N(3) \uparrow = 35$.

8. Au plus une langue : $0,1=10\%$ des employés ou $F(1) \uparrow = 0,1$

9. Au moins 5 langues parlées : $f_5=0,06=6\%$ ou $F(4) \downarrow = 0,06$

Corrigé de l'exercice 05 :

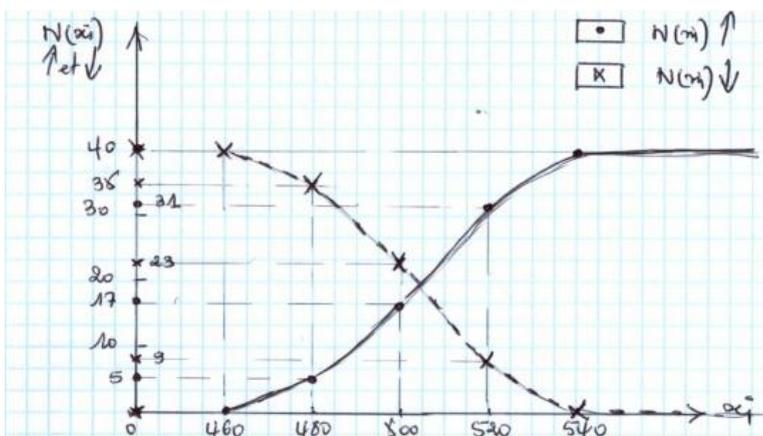
1. Il s'agit d'une variable statistique continue (le prix).
2. 35 magasins qui vendent le produit entre 480 et 540 ($12+14+9=35$)
3. La représentation par histogramme. Et la représentation du polygone des effectifs.



5.

| Prix en dinars(x_i) | [460;480[| [480;500[| [500;520[| [520;540[| Total |
|-------------------------|-----------|-----------|-----------|-----------|--------|
| Magasins(n_i) | 5 | 12 | 14 | 9 | $N=40$ |
| $N(e_i) \uparrow$ | 5 | 17 | 31 | 40 | |
| $N(e_i) \downarrow$ | 40 | 23 | 9 | 0 | |

6. Traçage de la courbe des effectifs cumulés croissants et décroissants.



7. 31 magasins qui vendent le produit inférieur à 520 dinars :

$(5+12+14=31)$ ou $N(520) \uparrow = 31$

8. 35 magasins qui vendent le produit supérieur à 480 dinars :

$$(12+14+9=35) \text{ ou } N(480) \downarrow = 35$$

9. le pourcentage de magasins qui vendent le produit à moins de 500 dinars est 42,5% $(5+12)/40=17/40=0,425=42,5\%$ ou

$$N(500) \uparrow = 0,425$$

Corrigé de l'exercice 6 :

Le responsable d'un magasin a relevé au cours d'une période de 10 jours le nombre d'articles vendus. Il a obtenu les résultats suivants,

300, 100, 500, 300, 100, 1000, 900, 700, 100, 100.

1. La population statistique est 10 jours.
2. L'observation 900 signifie que 900 articles sont vendus en une journée.
3. La variable statistique est le nombre d'articles vendus et sa nature est discrète.
4. Calcul des fréquences cumulées croissantes et décroissantes.

| | | | | | | | |
|---|-----|-----|-----|-----|-----|------|------------------|
| xi (le nombre d'articles vendus) | 100 | 300 | 500 | 700 | 900 | 1000 | Total |
| les jours ni | 4 | 2 | 1 | 1 | 1 | 1 | N=10 |
| $f_i = \frac{n_i}{N}$ | 0,4 | 0,2 | 0,1 | 0,1 | 0,1 | 0,1 | $\sum_i f_i = 1$ |
| $N(x_i) \uparrow$ | 0 | 4 | 6 | 7 | 8 | 9 | 10 |
| $N(x_i) \downarrow$ | 10 | 6 | 4 | 3 | 2 | 1 | 0 |
| $F(x_i) \uparrow$ | 0 | 0,4 | 0,6 | 0,7 | 0,8 | 0,9 | 1 |
| $F(x_i) \downarrow$ | 1 | 0,6 | 0,4 | 0,3 | 0,2 | 0,1 | 0 |

Corrigé de l'exercice 7. On étudie les revenus (mensuels en dinars) d'un ensemble de familles d'un quartier.

| | | | | | | |
|------------------|-------------|--------------|---------------|---------------|---------------|-------|
| Revenus | [7000;9000[| [9000;11000[| [11000;13000[| [13000;15000[| [15000;17000[| Total |
| Effectifs | 30 | 160 | 60 | 70 | 80 | N=400 |

1. Le nombre de familles dont le revenu est compris entre 7000 et 9000 dinars est de 30.
2. La proportion des familles dont le revenu est compris entre 9000 et 15000 dinars est de 72,5% $((160+60+70/400) \times 100 = 0,725 \times 100 = 72,5\%)$
3. La proportion de familles dont le revenu est supérieur à 13000 dinars est 37,5%. Et de 80% si c'est inférieur à 15000 dinars.

| | | | | | | |
|------------------|-------------|--------------|---------------|---------------|---------------|-------|
| Revenus | [7000;9000[| [9000;11000[| [11000;13000[| [13000;15000[| [15000;17000[| Total |
| Effectifs | 30 | 160 | 60 | 70 | 80 | 400 |
| fi | 0,075 | 0,4 | 0,15 | 0,175 | 0,2 | 1 |
| F(xi) ↑ 0 | 0,075 | 0,475 | 0,625 | 0,8 | 1 | |
| F(xi) ↓ 1 | 0,925 | 0,525 | 0,375 | 0,2 | 0 | |

Exercices proposées

Exercice n°1 :

Déterminer la population statistique, l'unité statistique, la nature du caractère ainsi que la variable statistique dans les cas suivants :

- 1- 473 garçons âgés de 12 à 17 ont été interrogés sur de leur durée moyenne d'écoute de la télévision.
- 2- on étudie la répartition des actifs occupés en Algérie selon la catégorie socioprofessionnelle.
- 3- On dispose des données suivantes sur la répartition des députés à l'Assemblée nationale en fonction de l'appartenance politique en 2007.
- 4- Répartition des familles en Algérie en 2008 selon le nombre d'enfants.
- 5- Des salaires dans les secteurs public et privé
- 6- La répartition des logements selon la catégorie.
- 7- Les exploitations agricoles se répartissaient selon la surface agricole utilisée.

Exercice n°2 :

On a noté l'âge (arrondi à l'année près) des 48 salariés d'une entreprise. Les données sont listées ci-dessous :

43 29 57 45 50 29 37 59 46 31 46 24 33 38 49 31

62 60 52 38 38 26 41 52 60 49 52 41 38 26 37 59

57 41 29 33 33 43 46 57 46 33 46 49 57 57 46 43

1. Identifier la variable statistique et préciser sa nature.
2. Présenter la distribution des effectifs dans les classes :
[20; 30[, [30; 40[, [40; 50[, [50; 60[, [60; 70[.

Exercice n°3 : Une entreprise de services a relevé au cours des cent derniers jours ouvrés le nombre de plaintes de clients par jour. Le tableau suivant synthétise les résultats :

| | | | | | | | | |
|--------------------|----|----|----|----|----|---|---|---|
| Nombre de plaintes | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Nombre de jours | 11 | 22 | 19 | 16 | 11 | 9 | 7 | 5 |

- 1- Identifier la variable statistique et préciser sa nature.
- 2- Déterminer la population statistique et l'unité statistique.
- 3- Calculer les effectifs cumulés croissants et décroissants.
- 4- Calculer les fréquences cumulées croissantes et décroissantes.
- 5- Tracer le diagramme en bâtons des effectifs n_i . De même pour le polygone.
- 6- Tracer la courbe cumulative des fréquences.
- 7- Donner le nombre de client qui ont déposé moins de 4 plaintes.
- 8- Quelle la proportion des clients ayant au plus une plainte déposée? De même pour au moins 5.

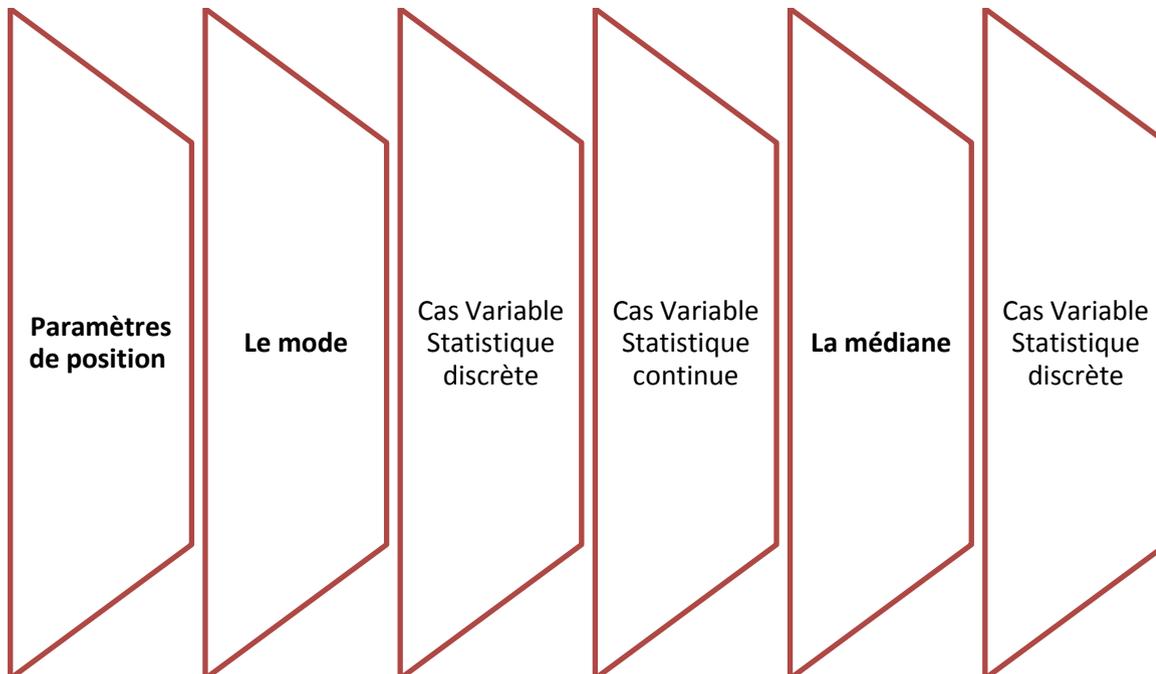
Exercice n°4 :

On a mesuré 20 personnes et les résultats sont (en cm) :

| La taille | Effectifs |
|------------------|------------------|
| [130-140[| 1 |
| [140-150[| 6 |
| [150-160[| 1 |
| [160-170[| 2 |
| [170-180] | 10 |

1. Identifier la variable statistique et préciser sa nature.
2. Combien de personnes ont une taille entre 140 et 170 cm?
3. Faire une représentation par histogramme.
4. Représenter le polygone des effectifs.
5. Composez le tableau des effectifs cumulés croissants et décroissants.
6. Tracer la courbe des effectifs cumulés croissants et décroissants.
7. Dans combien de personne ont une taille inférieure à 170 cm?
8. Dans combien de personne ont une taille supérieur à 150 cm?
9. Quel est le pourcentage de personne qui ont une taille à moins de 160 cm?

Conférence n°04



IV. Paramètres de position

Introduction :

Lorsque la variable étudiée est quantitative, le traitement statistique des informations ne se limite pas aux seules représentations graphiques (dont l'interprétation reste toujours plus au moins subjective on cherche à résumer les informations par des indicateurs numériques bien choisis. Ces indicateurs indiquent la caractéristique la plus représentative de tous les individus du groupe en la ramenant à un individu type qui se situerait au « centre » de la distribution. Ce sont des indices qui donnent des informations sur la position générale de la distribution et satisfaisant certaines propriétés comme il a signalé Yule.

- **Conditions de Yule³**

Le statisticien britannique Georges U. Yule a énoncé un certain nombre de propriétés souhaitées pour les indicateurs des séries statistiques ; ceux-ci doivent être d'une part, des résumés "maniabiles" et d'autre part, les plus exhaustifs possibles relativement à l'information contenue dans les données.

Dans son schéma, une caractéristique statistique doit être une valeur-type :

1. définie de façon objective et donc indépendante de l'observateur,
2. dépendante de toutes les observations,
3. de signification concrète pour être comprise par les non-spécialistes,
4. simple à calculer,
5. peu sensible aux fluctuations d'échantillonnages,
6. se prêtant aisément aux opérateurs mathématiques classiques.

En réalité, on ne dispose pas de caractéristiques répondant simultanément à ces six conditions. Le choix d'un indicateur sera l'objet d'un compromis guidé par la spécificité de l'étude en cours.

Les plus usités sont le mode ; la médiane et la moyenne arithmétique elles en fournissent un ordre de grandeur. Les modes de détermination de ces paramètres différents en fonction du type discret ou continu de la variable quantitative étudiée.

IV.1. Le mode

Lorsqu'un fabricant de vêtement décide de ne fabriquer qu'une taille d'un modèle particulier, il a intérêt à choisir la taille la plus fréquemment portée

³ B. Goldfard et C. Pardoux, 2013; « Introduction à la méthode statistique, Statistique et probabilités, Cours et exercices corrigés » © Dunod, Paris, pp 13-14

IV.1.1.Cas Variable Statistique Discrète

On appelle mode d'une variable statistique discrète et on note M_o

La modalité la plus fréquente (valeur de la variable correspondant à l'effectif maximale d'où on le note $n_i \uparrow$)

Note : pour une distribution discrète, le mode satisfait aux conditions 1, 3 et 4 de Yule.

Exemple :

Tableau 2.1 : Nombre de jours d'absences des étudiants

| | | | | | | | | |
|---------------------------|---|----------|---|---|---|----|----|-------|
| Nombre de jours d'absence | 0 | 3 | 5 | 7 | 8 | 11 | 13 | Total |
| L'effectif | 1 | 4 | 2 | 1 | 1 | 2 | 1 | 12 |

La valeur du variable « nombre de jours d'absence » qui est la plus fréquente ex (effectif correspondant : 4) les autres valeurs n'apparaissent que 2 ou 1 fois le mode de la série est donc 3 ; $M_o = 3$ jours

Remarque : si l'effectif maximal se répète deux (ou plusieurs fois) pour des valeurs de X on parle de série statistique bimodale (ou plurimodale).

IV.1.2.Cas Variable Statistique Continue

On appelle mode d'une V.S.C et on note M_o la valeur de la variable associé à l'effectif(ou a la fréquence) le plus élevé par unité d'amplitude, la détermination se fait en deux étapes :

- Détermination d'une classe modale
- Détermination du mode (valeur particulière de cette classe modale)

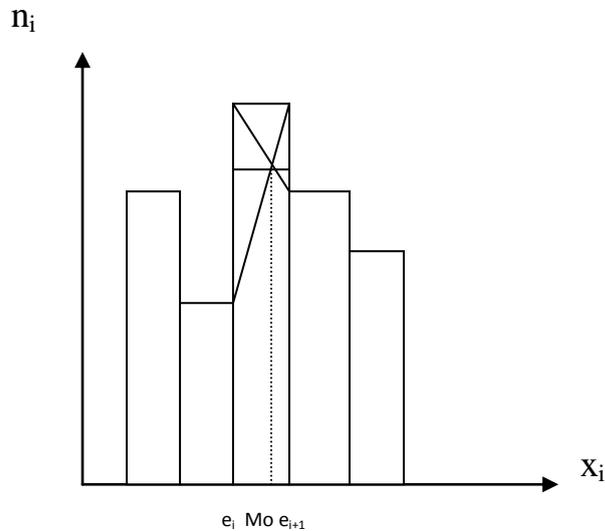
La classe modale est la classe d'effectif(ou de fréquence) maximal $n_i \uparrow$.

Exemple : revenu agricole des exploitations individuelles.

Tableau 2.2 : Revenu agricole des exploitations individuelles

| Revenu en D | Nombre d'exploitation |
|-------------------------------|---------------------------|
| [-100 à -50[| 7344 |
| [-50 à -0[| 39587 |
| Classe modale [0 à 50[| 164497 ↔ effectif maximum |
| [50 à 100[| 125821 |
| [100 à 150[| 44245 |
| [150 à 200[| 17614 |
| [200 à 250[| 5086 |
| total | 404194 |

Figure 2.1 : Détermination graphique du Mode



Soit $[e_1, e_2]$ [la classe modale, l'estimation du mode est obtenue en traçant les diagonales AC et BD joignant les sommets A et D du rectangle modale aux sommets B et C des rectangles contigus. Le Mode est l'abscisse du point et l'intersection E des deux diagonales =

Détermination par calcul : théorème de Thalès)

$$\text{ABE et CDE triangle } \frac{EF}{AB} = \frac{EG}{CD} \Rightarrow \frac{M_o - e_1}{\alpha_1} = \frac{e_2 - M_o}{\alpha_2} \Rightarrow \mathbf{M_o = e_i + a_i \times \frac{\alpha_1}{\alpha_1 + \alpha_2}}$$

e_i : représente l'extrémité inférieure de la classe modale

e_i : représente l'extrémité supérieure de la classe modale

$a_i = e_{i+1} - e_i$ (a_i amplitude de la classe modale)

α_1 : La différence entre l'effectif de la classe modale et celui de la classe précédente

α_2 : La différence entre l'effectif de la classe modale et celui de la classe suivante

On prenant notre exemple :

$$M_o \in [0; 50] \text{ car } n_i \uparrow = 164497$$

$$M_o = 0 + (50 - 0) \times \frac{(164497 - 39587)}{(164497 - 39587) + (164497 - 125821)} = 38,18$$

IV.2. La médiane (Me)

La médiane d'une distribution statistique, notée Me, est la valeur de la variable statistique qui sépare la population en deux parties égales telle que le nombre des observations qui présentent une valeur inférieure à Me soit égal au nombre des observations qui présentent une valeur supérieure à Me.

Remarque : La médiane est la valeur Me de la variable statistique pour laquelle la fréquence cumulée (ou effectif cumulé) est égale à $1/2$, i.e.

$$F(Me) = 1/2 \text{ (ou } N(Me) = N/2)$$

Elle satisfait les conditions 1, 3, 4 et 5 de Yule. Contrairement à la moyenne arithmétique, la médiane résiste aux valeurs extrêmes.

IV.2.1.Détermination pratique

Nous commençons par ordonner la série deux cas peuvent se présenter :

Le cas pair et le cas impair

1- Le cas nombre d'observation est pair :

Dans le cas d'une série statistique comportant $2k$ observations, il n'y a pas à proprement parler de médiane. Ainsi, on introduit la notion « d'intervalle médian » dont les extrémités correspondent aux valeurs de la $k^{\text{ième}}$ et de la $(k + 1)^{\text{ième}}$ observations. On prend généralement la moyenne arithmétique de ces deux extrémités. La médiane sera alors la demi somme de la $k^{\text{ième}}$ et de la $(k+1)^{\text{ème}}$ observation c'est à dire $(\frac{k+(k+1)}{2})^{\text{ème}}$ valeur

Exemple : Dans la série : 4,5,8,8, 9,11, 12,14,17,19

$$N=10=2k ; k = \frac{N}{2} = \frac{10}{2} = 5, \text{ donc } M_e = \frac{5^{\text{ème}} \text{ obs} + 6^{\text{ème}} \text{ obs}}{2} ; M_e = \frac{9+11}{2} = \frac{20}{2} = 10$$

2- Le cas nombre d'observation est pair :

Si la série est impair possède un nombre impair $(2k+1)$, la médiane $M_e = (k+1)^{\text{ème}}$ valeurs

Exemple : Dans la série : 1, 2, 4, 4,5

$$N=5=2k+1 \Rightarrow k=2 \text{ donc } M_e = (2+1)^{\text{ème}} \text{ obs} = 3^{\text{ème}} \text{ obs} = 4$$

Remarque : Cette méthode n'est applicable que lorsque le nombre d'observation est faible.

IV.2.2.Cas d'une variable statistique discrète

Dans la pratique, on détermine la médiane soit à partir des fréquences cumulées de tableaux statistiques, soit à partir du graphique cumulatif.

A partir de la construction du tableau des fréquences cumulée absolues ou relatives ,on va déterminer les valeurs x_i celle qui correspond à $\frac{N}{2}$ ou $F(x) = 0.5$

Reprenant l'exemple du nombre de pièces par logement

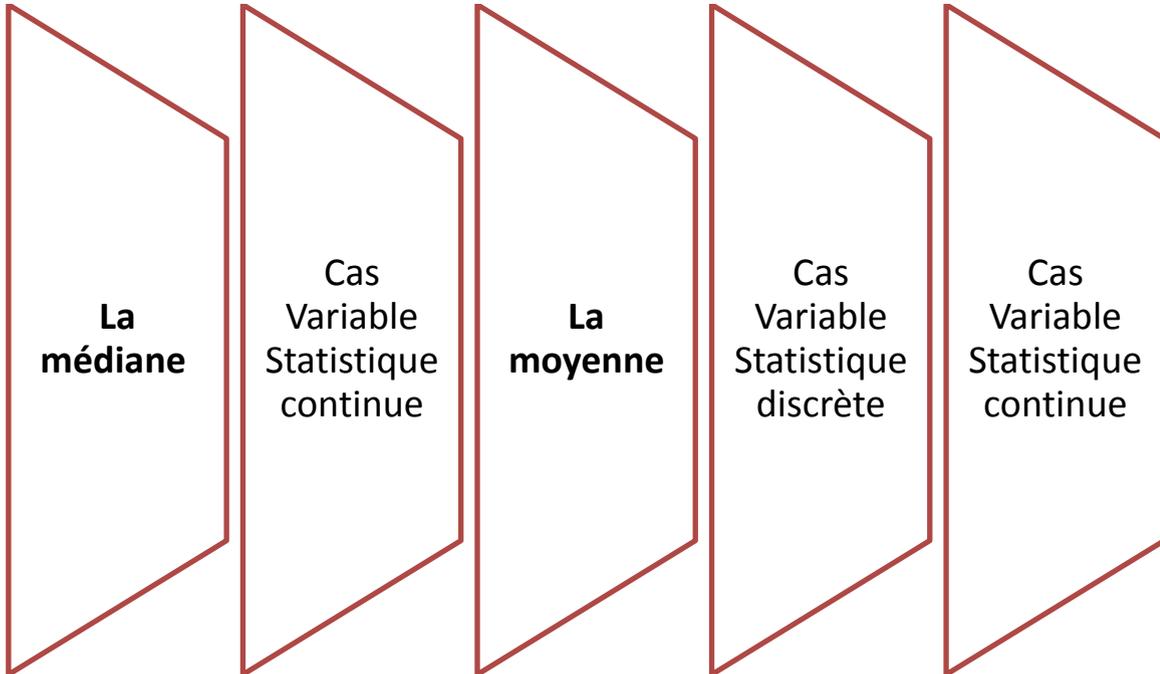
Tableau 2.3 : Le nombre de pièces par logement

| x_i | n_i | $f_i = \frac{n_i}{N}$ | $N(x) \uparrow$ | $N(x) \downarrow$ | $F(x) \uparrow$ | $F(x) \downarrow$ |
|----------|-------|------------------------|-----------------|-------------------|-----------------|-------------------|
| 1 | 15 | 0,15 | 0 | 100 | 0 | 1 |
| 2 | 20 | 0,20 | 15 | 85 | 0,15 | 0,85 |
| 3 | 25 | 0,25 | 35 | 65 | 0,35 | 0,65 |
| 4 | 40 | 0,40 | 60 | 40 | 0,60 | 0,40 |
| | | | 100 | 0 | 1 | 0 |
| Total | N=100 | $\sum_{i=1}^k f_i = 1$ | | | | |

$$\frac{N}{2} = \frac{100}{2} = 50$$

Dans ce cas, la 50^{ième} observations se situe à la valeur 3, qui correspond, dans la distribution des effectifs cumulés, à toutes les observations situées entre la 35^{ième} observation et 60^{ième} observation (intervalle contenant l'observation médiane qui est la 50^{ième}). La valeur de la médiane se détermine de façon directe. D'où $Me = 3$ pièces par logement, c'est-à-dire qu'il ya d'autant de logement ayant moins de 3 pièces que de logements ayant plus de 3 pièces.

Conférence n°05



IV.2.3.Cas d'une variable statistique continue

Dans le cas d'une variable statistique continue la médiane est définie avec exactitude. Mais, en raison du regroupement par classe on ne peut généralement que la situer à l'intérieur d'une classe qu'on qualifiera de classe médiane.

IV.2.3.1.La médiane par calcul :

Si la moitié de l'effectif total $\frac{N}{2}$ correspond à une valeur de l'extrémité de la classe, la médiane est une valeur exacte.

| e_i | | | N ↗ | F ↗ |
|---------|-------|--|-----|-----|
| [10,20[| | | 10 | |
| [20,30[| | | 50 | 0.5 |
| [30,40[| | | 100 | |
| | N=100 | | | |

→ $M_e = 30$

On aura la moitié des effectifs est comprise entre deux extrémités de classe e_i et e_{i+1} , on a une classe médiane, on utilise la méthode d'interpolation linéaire

$$M_e = e_i + a_i \frac{\frac{N}{2} - N(x_i) \uparrow}{n_i}$$

$$\text{Ou } M_e = e_i + a_i \frac{0.5 - F(x_i) \uparrow}{f_i}$$

e_i : l'extrémité inférieure de la classe médiane
 a_i : l'amplitude
 n_i : effectif la classe médiane
 N : effectif total
 $N(x_i) \uparrow$: effectif cumulé croissant de la classe médiane
 f_i : fréquence de la classe média $F(x_i)$: fréquence Cumulée de la classe médiane.

Exemple : on prenant les résultats obtenus en mathématiques dans un centre d'examen

Tableau 2.4 : Les résultats obtenus en mathématiques dans un centre d'examen

| classes | Centre (c_i) | Effectifs n_i | eff. cumul. croiss. $N \uparrow$ | eff. cumul. décroiss. $N \downarrow$ |
|---------------|------------------|-----------------|----------------------------------|--------------------------------------|
| [0,2[| 1 | 4 | 4 | 150 |
| [2,4[| 3 | 11 | 15 | 146 |
| [4,6[| 5 | 14 | 29 | 135 |
| [6,8[| 7 | 29 | 58 | 121 |
| [8,10[| 9 | 26 | 84 | 92 |
| [10,12[| 11 | 38 | 122 | 66 |
| [12,14[| 13 | 15 | 137 | 28 |
| [14,16[| 15 | 8 | 145 | 13 |
| [16,18[| 17 | 2 | 147 | 5 |
| [18,20] | 19 | 3 | 150 | 3 |
| Total | | 150 | | |

On détermine la classe médiane :

$$N(\text{Me}) = N/2 = \frac{150}{2} = 75$$

$$58 < 75 < 84 \Rightarrow \text{Me} \in [8; 10]$$

Donc la classe médiane est [8; 10] on applique la formule on aura

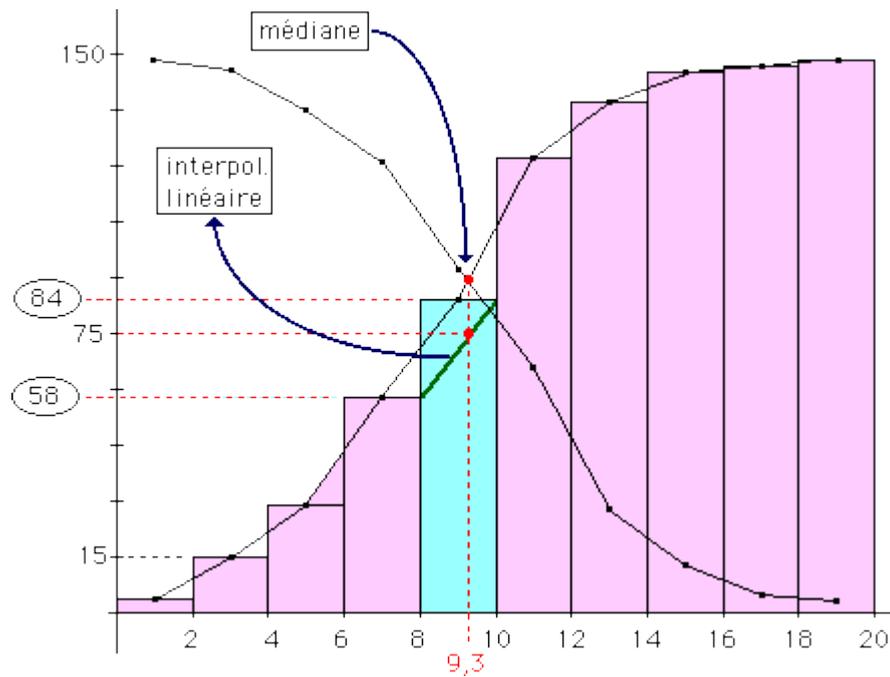
$$\text{Me} = e_i + a_i \frac{\frac{N}{2} - N(x_i)^{\uparrow}}{n_i} \quad \text{Me} = 8 + 2 \times (75 - 58)/(26) \cong 9,3$$

IV.2.3.2. Détermination graphique :

On construit la courbe cumulative de effectifs cumulés $N \uparrow$ (ou fréquences cumulées $F \uparrow$; Et on déterminé la moitié d'effectif total (ou 0.5 de fréquence cumulée).

La médiane s'interprète comme abscisse du point d'intersection des courbes des effectifs (ou des fréquences) cumulé(e)s croissant(e)s et décroissant(e)s. On la retrouve à l'intersection de l'horizontale d'ordonnée 75 (moitié de l'effectif de la population) et de la droite d'interpolation passant par les points d'entrée et sortie de la classe médiane.⁴

Figure 2.2 : Détermination graphique de la médiane



⁴ <http://serge.mehl.free.fr/exos/histog.html>

IV.3. La moyenne

Définition :

La moyenne est la caractéristique de tendance centrale la plus réputée et n'est définie que pour des variables statistiques quantitatives.

Plusieurs ouvrages la présentent comme une valeur mathématique et non physique, ils la définissent par sa formule :

- La moyenne arithmétique est définie comme étant la somme des valeurs observées divisée par le nombre d'observations
- La moyenne harmonique est définie comme étant l'inverse de la moyenne arithmétique des inverses des observations
- La moyenne géométrique est définie comme étant la racine nième des produits des n observations

IV.3.1. La moyenne arithmétique

Définition : la moyenne arithmétique notée \bar{X} (se lit "x barre") d'une V.S.X se définit comme étant le rapport de la somme des valeurs prises par cette variable sur l'effectif total N

On distingue 2 types de moyennes : la moyenne arithmétique simple et la moyenne arithmétique pondérée.

la moyenne arithmétique simple est simple lorsque les valeurs prises par la variable X n'apparaissent

qu'une seule fois, elle est égale donc à $\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{N} = \frac{1}{N} \sum_{i=1}^k x_i$

IV.3.1.1. Cas Variable Statistique Discrète (V.S.D) :

La moyenne arithmétique pondérée lorsque les valeurs prises par la V.S X($x_1; x_2 \dots x_k$) ou les fréquences d'apparition respectives (n_1, n_2, \dots, n_k), la moyenne arithmétique est égale

$$\bar{X} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_k \cdot x_k}{n_1 + n_2 + \dots + n_k} = \frac{\sum_{i=1}^k n_i x_i}{\sum_{i=1}^k n_i} = \frac{1}{N} \sum_{i=1}^k n_i x_i$$

Propriété: Pour obtenir la moyenne d'une série statistique:

1. on multiplie chaque valeur (ou centre de classe) par l'effectif partiel correspondant;
2. on additionne les produits ainsi obtenus;
3. on divise cette somme par l'effectif total.

On parle, dans ce cas, de moyenne pondérée par les effectifs.

Lorsque on utilise les fréquences relatives, la moyenne arithmétique pondérée est égale à \bar{X}

$$= \sum_{i=1}^k f_i x_i$$

Exemple : On considère la série statistique suivante :

Tableau 2.5 : Calcul de la moyenne arithmétique (cas V.S.D)

| | | | | | | |
|----------------------|----|----|----|----|----|-------|
| Notes x_i | 9 | 10 | 11 | 12 | 14 | Total |
| Coefficient n_i | 6 | 4 | 8 | 5 | 1 | 24 |
| $x_i.n_i$ | 54 | 40 | 88 | 60 | 14 | 256 |

Pour calculer la moyenne pondérée, on effectue le calcul suivant:

$$\bar{X} = \frac{\sum_{i=1}^k x_i \times n_i}{N} = \frac{9 \times 6 + 10 \times 4 + 11 \times 8 + 12 \times 5 + 14 \times 1}{24} = \frac{256}{24} = 10,7$$

Note : on doit ajouter une nouvelle colonne ($n_k x_k$) au tableau pour procéder facilement au calcul de la moyenne arithmétique

IV.3.1.2.Cas d'une Variable Statistique Continue (V.S.C):

Dans le cas d'une variable continue, cette moyenne pondérée n'est qu'une valeur approchée de la vraie valeur moyenne de la série car on remplace chaque x_i par le centre de la classe à laquelle il appartient. On va calculer d'abord les centres C_i de chaque classe

$$C_i = \frac{\text{extr émit é inferieur} + \text{extr émit é superieur}}{2} = \frac{e_{i-} + e_{i+}}{2}$$

La moyenne arithmétique pondérée dans ce cas est égale à

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k n_i \times c_i$$

Exemple des notes : Tableau 2.6 : Calcul de la moyenne arithmétique (cas V.S.C)

| classes | Centres (c_i) | Effectifs n_i | $c_i \times n_i$ |
|---------|----------------------|-----------------|------------------|
| [0,2[| 1 | 4 | 4 |
| [2,4[| 3 | 11 | 33 |
| [4,6[| 5 | 14 | 70 |
| [6,8[| 7 | 29 | 203 |
| [8,10[| 9 | 26 | 234 |
| [10,12[| 11 | 38 | 418 |
| [12,14[| 13 | 15 | 195 |
| [14,16[| 15 | 8 | 120 |
| [16,18[| 17 | 2 | 34 |
| [18,20] | 19 | 3 | 57 |
| Total | | 150 | 1368 |

$$\bar{x} = \frac{1}{150} (1368) = 9,12$$

- **Pourquoi utiliser la moyenne arithmétique?**

Elle a été choisie parmi d'autres types de moyenne (géométrique, harmonique...) car elle possède une propriété extrêmement intéressante: Lorsqu'on se livre à des observations scientifiques, les mesures ne sont pas toujours exactement identiques d'une fois sur l'autre, même lorsque les conditions semblent être similaires. Il se produit ce que l'on appelle une erreur d'observation.

On a la relation suivante :

Valeur observée = valeur exacte + erreur d'observation

avec: x_i = valeur observée x_e = valeur exacte
 $x_i - x_e$ = erreur d'observation

On décide alors de prendre pour x_e la valeur qui minimise les erreurs d'observation, en fait la moyenne des carrés de ces erreurs (critère des moindres carrés) . Le calcul prouve que la meilleure valeur estimant x_e suivant ce critère est \bar{x} .

Propriété :

La moyenne \bar{x} des valeurs observées d'une grandeur x correspond à la meilleure estimation de x_e . Elle conserve la somme totale et satisfait à toutes les conditions de Yule, sauf la 5ème, car elle est une mesure sensible aux valeurs extrêmes.

Conférence n°06

**Moyenne
géométrique**

**Moyenne
harmonique**

**Quantiles
(quartiles)**

IV.3.2.Moyenne géométrique

Beaucoup moins fréquente, La moyenne géométrique s'utilise, par exemple, quand on veut calculer la moyenne de taux d'intérêt.

IV.3.2.1.Moyenne géométrique simple

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

$$G = \sqrt[N]{x_1 \times x_2 \times \dots \times x_n}$$

$$\log G = \frac{1}{N} \sum_i \log x_i = \frac{1}{N} (\log x_1 + \log x_2 + \dots + \log x_n)$$

Exemple : Au cours des 5 dernières années, les taux de croissance annuels du PIB ont été comme suit :

Tableau 2.7 : les taux de croissance annuels du PIB

| Année | 2015 | 2016 | 2017 | 2018 | 2019 |
|---------|------|------|------|------|------|
| Taux(%) | 3,7 | 3,2 | 1,3 | 1,2 | 0,8 |

Quel est le taux moyen de croissance du PIB au cours de ces dernières années ?

$$G = \sqrt[5]{103,7 \times 103,2 \times 101,3 \times 101,2 \times 100,8} = 1,71$$

Au cours de ces dernières années, le taux moyen de croissance du PIB est de 1,71%

IV.3.2.2.Moyenne géométrique pondérée:

Dans ce cas la formule de la moyenne géométrique devient :

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

$$\log G = \frac{1}{\sum_i n_i} \sum_i n_i \log x_i = \frac{1}{\sum_i n_i} (n_1 \log x_1 + n_2 \log x_2 + \dots + n_k \log x_k)$$

Exemple :

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

Quel est le taux de croissance annuel moyen des bénéfices réalisés au cours de ces 9 années ?

$$G = \sqrt[9]{150^4 \times 117^3 \times 70^2}$$

$$\log G = \frac{1}{9}(4\log 150 + 3\log 117 + 2\log 70) = 2,06$$

D'où $G=10^{2,06}=114,8\%$; l'augmentation moyenne est de 14,8%.

IV.3.3.Moyenne harmonique H

Définition :

La moyenne harmonique d'une variable X est égale à l'inverse de la moyenne arithmétique des inverses des valeurs observées de cette variable (aucune des valeurs observées ne devant être nulle).

IV.3.3.1.Moyenne harmonique simple

Si chaque valeur de la variable est observée une seule fois, la moyenne harmonique est dite simple et s'écrit :

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

Exemple : vitesse moyenne

Un automobiliste effectue un aller-retour entre deux villes. À l'aller sa vitesse moyenne est de 130 km/h ; au retour, elle n'est que de 70 km/h. Quelle a été sa vitesse moyenne sur l'ensemble du parcours ?

La réponse n'est pas la moyenne arithmétique des vitesses : $(130 + 70)/2 = 100$! Pourquoi ? Parce que la vitesse moyenne est un rapport, le rapport de la distance totale parcourue au temps nécessaire pour la parcourir. Pour la calculer, notons d la distance entre les deux villes. Le numérateur du rapport, la distance totale parcourue, est égal à 2d.

D'après la définition de la vitesse moyenne, le temps nécessaire pour parcourir un trajet est le rapport de la distance à la vitesse moyenne, soit $d/130$ pour l'aller et $d/70$ pour le retour. Le dénominateur du rapport est donc égal à $\frac{d}{130} + \frac{d}{70}$

Le calcul de la vitesse moyenne est le suivant :

$$\text{Vitesse moyenne} = \frac{\text{distance}}{\text{temps}} = \frac{2d}{\frac{d}{130} + \frac{d}{70}} = \frac{2}{\frac{1}{130} + \frac{1}{70}} = 91 \text{ km/h.}$$

La vitesse moyenne est une moyenne harmonique simple :

$$H = \frac{N}{\frac{1}{x_1} + \frac{1}{x_2}}, \text{ avec } N = 2, x_1 = 130 \text{ et } x_2 = 70$$

IV.3.3.2.Moyenne harmonique pondérée :

Si x_1 a été observé n_1 fois, x_2 n_2 fois, ..., x_k n_k fois, par définition la moyenne harmonique est égale à :

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

La moyenne harmonique est employée dans le calcul des moyennes de rapports.

Exemple : Une entreprise de transport a 10 camions qui font des rotations entre Oran et Sidi Bel-Abbes.

Au cours d'une de celle-ci, le trajet a été couvert par ces camions aux vitesses moyennes présentées dans le tableau suivant :

Tableau 2.8 : Vitesse moyenne des camions

| Vitesse moyenne (km/h) | 50 | 60 | 70 |
|------------------------|----|----|----|
| Nombre de camions | 4 | 2 | 4 |

Quelle est la vitesse moyenne globale ?

$$H = \frac{\sum_i n_i}{\sum_i \frac{n_i}{x_i}} = \frac{10}{\frac{4}{50} + \frac{2}{60} + \frac{4}{70}} = 58,66 \text{ km/h}$$

Remarque : lorsque, pour une série statistique, chacune des trois moyennes existe, on peut démontrer que : $H < G < \bar{X}$

Les trois paramètres les plus usées (mode, moyenne arithmétique et la médiane ont des avantages et des inconvénients sont présentés dans le tableau comme suit :

Tableau 2.9 : Comparaison des indicateurs de position

| | Avantages | Inconvénients |
|--------------------------------|---|--|
| la moyenne arithmétique | <ul style="list-style-type: none"> - Facile à calculer, - Répond au principe des moindres carrés. | <ul style="list-style-type: none"> - Fortement influencée par les valeurs extrêmes de la variable, - Représente mal une population hétérogène (polymodale). |
| La médiane | <ul style="list-style-type: none"> - Pas influencée par les valeurs extrêmes de la variable, - Peu sensible aux variations d'amplitude des classes, - Calculable sur des caractères cycliques (saison, etc.) où la moyenne a peu de signification. | <ul style="list-style-type: none"> - Se prête mal aux calculs statistiques, - Suppose l'équi-répartition des données - Ne représente que la valeur qui sépare l'échantillon en 2 parties égales. |
| Le mode | <ul style="list-style-type: none"> - Pas influencée par les valeurs extrêmes de la variable, - Calculable sur des caractères cycliques (saison, etc.) où la moyenne a peu de signification, - Bon indicateur de population hétérogène. | <ul style="list-style-type: none"> - Se prête mal aux calculs statistiques, - Très sensible aux variations d'amplitude des classes, - Son calcul ne tient compte que des individus dont les valeurs se rapprochent de la classe modale. |

IV.4. Quantiles⁵ :

Définitions :

Soit une variable quantitative X , discrète ou continue, dont les modalités sont classées par valeurs croissantes.

Le quantile d'ordre α %, noté q_α , est la valeur de la variable telle que α % des valeurs observées sont inférieures à q_α . Donc, par définition, la fréquence cumulée croissante associée au quantile d'ordre α % est α %, soit: $F(q_\alpha) = \alpha$ %

Le quantile d'ordre α % peut aussi être défini comme la valeur de la variable qui partage la population en deux sous-populations, telles que dans la première on a $(\alpha \%)n$ individus et dans la seconde $(100 - \alpha \%)n$ individus. L'effectif cumulé croissant associé au quantile d'ordre α % est donc égal à $(\alpha \%)n$, soit : $N(q_\alpha) = (\alpha \%)n$

Principaux quantiles

Les quantiles les plus couramment utilisés sont les quartiles, les déciles et les centiles :

– les quartiles sont les trois valeurs de la variable notées q_{25} , q_{50} et q_{75} qui partagent les observations en quatre groupes d'effectifs égaux : 25 % (resp. 50 % et 75 %) des individus de la population ont une valeur de la variable inférieure à q_{25} (resp. q_{50} et q_{75}) ;

La notation q_α a le mérite d'être identique pour tous les quantiles et d'indiquer, par le α présent en indice, la part de la population qui présente une valeur de la variable inférieure à ce quantile. Il existe cependant d'autres notations fréquemment adoptées : Q_1 , Q_2 , Q_3

Pour les déciles et les centiles ne sont pas traités dans ce cours.

Les quartiles :

Les quartiles divisent l'effectif de la série préalablement ordonnée par ordre croissant, en quatre parties égales. Le 1^{er} quartile Q_1 est tel que 25% des observations lui sont inférieures et 75% lui sont supérieures. $N(Q_1) = \frac{N}{4}$ ou $F(Q_1) = 25\%$. Le 2^{ème} quartile Q_2 est la médiane. Cette dernière apparaît donc aussi comme une caractéristique de position ($F(Q_2) = 50\%$) $N(Q_2) = \frac{N}{2}$. Le 3^{ème} quartile Q_3 est tel que 75% de l'observation lui sont inférieures et 25% lui sont supérieures $F(Q_3) = 75\%$ ou $N(Q_3) = \frac{3N}{4}$.

⁵ Elisabeth Olivier (2008), **L'essentiel de statistique descriptive** Edition Gualino, collection Les carrés

Note : Les quartiles non plus ne subissent pas l'influence des valeurs extrêmes. Comme la médiane, les quartiles satisfont aux conditions : 1, 3, 4 et 5 de Yule.

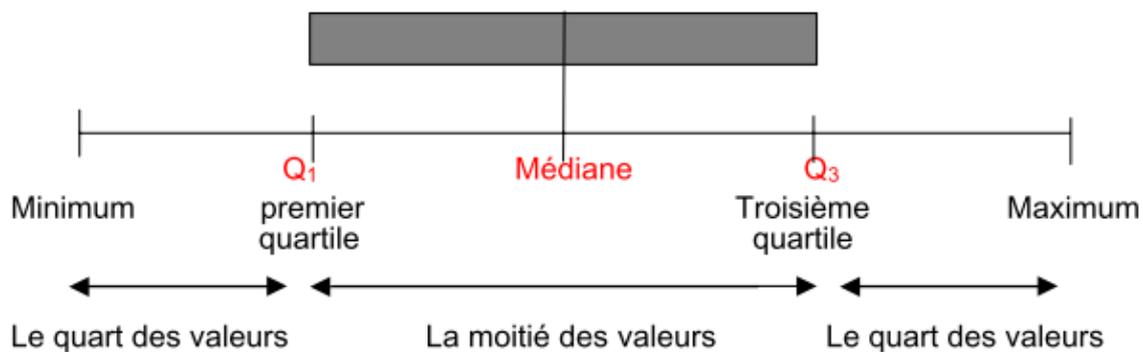
Remarque : Le deuxième quartile n'est autre que la médiane, qui correspond à 50% des effectifs de la série statistique.

La méthode d'évaluation diffère selon que les données sont ou non groupées dans un tableau statistique.

- **Les données non groupées**

En pratique : Soit une série statistique. Pour chercher les quartiles, on partage la série en 4 groupes de même effectif. On peut représenter ces quartiles dans un diagramme en boîte comme suit :

Figure 2.3 : Diagramme en boîte



Exemple : Trouver le premier et le troisième quartile de la série et construire son diagramme en boîte :
20 ; 52 ; 31 ; 4 ; 78 ; 5 ; 34 ; 4 ; 9 ; 10 ; 45 ; 12.

1°) Recherche des quartiles

- Je range les valeurs par ordre croissant : 4 ; 4 ; 5 ; 9 ; 10 ; 12 ; 20 ; 31 ; 34 ; 45 ; 52 ; 78.
- Je compte le nombre n de valeurs : Il y en a 12.
- Je divise n par 4 et j'obtiens le rang du 1^{er} quartile : $12 \div 4 = 3$.

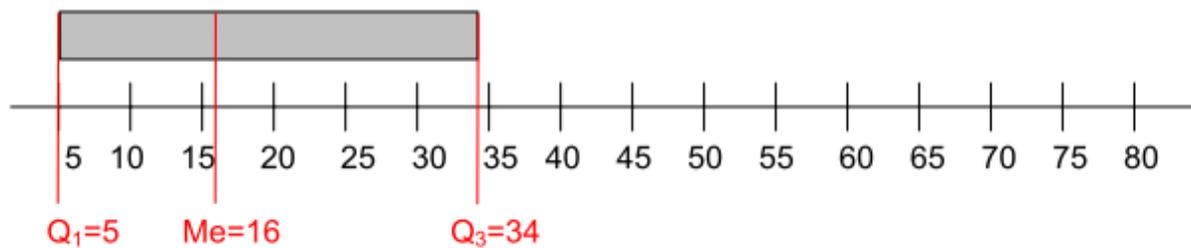
Donc le 1^{er} quartile est la troisième valeur. $Q_1=5$.

- J'obtiens le rang du 3^{ème} quartile par le calcul : $(n \div 4) \times 3 = (12 \div 4) \times 3 = 9$.

Donc le 3^{ème} quartile est la neuvième valeur. $Q_3=34$.

- Bien évidemment, comme le nombre de valeurs est pair, la médiane est égale à la moyenne des deux valeurs centrales : $m = (12+20) \div 2 = 16$. Donc $m = 16$.

Figure 2.4 :Construction du diagramme en boîte :



Le diagramme en boîte⁶ (ou boîte à moustaches, ou “box-and-whisker plot”)

Il s’agit d’un graphique très simple qui résume la série à partir de ses valeurs extrêmes, de ses quartiles et de sa médiane. La “boîte” est la partie du graphique comprise entre les premier et troisième quartiles. La médiane, nécessairement située à l’intérieur de la boîte, est représentée par un trait horizontal. Dans les parties basse et haute du graphique figurent les “moustaches”, joignant le minimum au premier quartile (en bas) et le troisième quartile au maximum (en haut). Ce graphique fournit donc très simplement quelques valeurs caractéristiques de la variable considérée.

On notera qu’il est encore possible de réaliser le diagramme en boîte de façon horizontale (on permute les axes horizontaux et verticaux), ce que font certains logiciels.

- Les données groupées

IV.4.1.Cas .Variables Statistique Discrète (V.S.D)

Exemple : répartition des enfants dans 500 familles

Tableau 2.10 : Calcul du quartile (Cas V.S.D)

| xi | ni | N(xi)↑ | N(xi)↓ | |
|----|-----|--------|--------|-------------------------|
| 0 | 50 | 0 | 500 | |
| 1 | 120 | 50 | 450 | N(Q ₁) =125 |
| 2 | 130 | 170 | 330 | N(Q ₂) =250 |
| 3 | 110 | 300 | 200 | N(Q ₃)= 375 |
| 4 | 90 | 410 | 90 | |
| | 500 | 500 | 0 | |

$$N(Q_1) = \frac{N}{4} = \frac{500}{4} = 125 \Rightarrow 50 < 125 < 170 \text{ D'où } Q_1 = 1 \text{ enfant}$$

$$N(Q_2) = \frac{N}{2} = 250 \Rightarrow 170 < 250 < 300 \Rightarrow Q_2 = 2 \text{ enfants}$$

$$N(Q_3) = \frac{3N}{4} = 375 \Rightarrow 300 < 375 < 410 \Rightarrow Q_3 = 3 \text{ enfants}$$

⁶ Alain Baccini (2010) ; **Statistique Descriptive Élémentaire** ; Publications de l’Institut de Mathématiques de Toulouse ; p 13.

IV.4.2.Cas variable statistique continue

On détermine la classe qui contient les quartiles :(même étape à suivre comme le calcul de la médiane qui est le deuxième quartile)

$$\text{Pour le 1}^{\text{ere}} \text{ quartile : } Q_1 = e_i + \frac{\frac{N}{4} - N(e_i)\uparrow}{n_i} \times ai$$

$$\text{Pour le 3}^{\text{eme}} \text{ quartile : } Q_3 = e_i + \frac{\frac{3N}{4} - N(e_i)\uparrow}{n_i} \times ai$$

Exemples :

Tableau 2.11 : Calcul du quartile (Cas V.S.C)

| $x_i \in [e_i, e_{i+1}[$ | n_i | $N(e_i)\uparrow$ | $N(e_i)\downarrow$ |
|--------------------------|-------|------------------|--------------------|
| [2,3-4,7[| 7 | 0 | 51 |
| [4,7-7,1[| 15 | 7 | 44 |
| [7,1-9,5[| 9 | 22 | 29 |
| [9,5-11,9[| 12 | 31 | 20 |
| [11,9-14,3[| 5 | 43 | 8 |
| [14,3-16,7[| 2 | 48 | 3 |
| [16,7-19,1[| 1 | 50 | 0 |
| Total | 51 | 51 | |

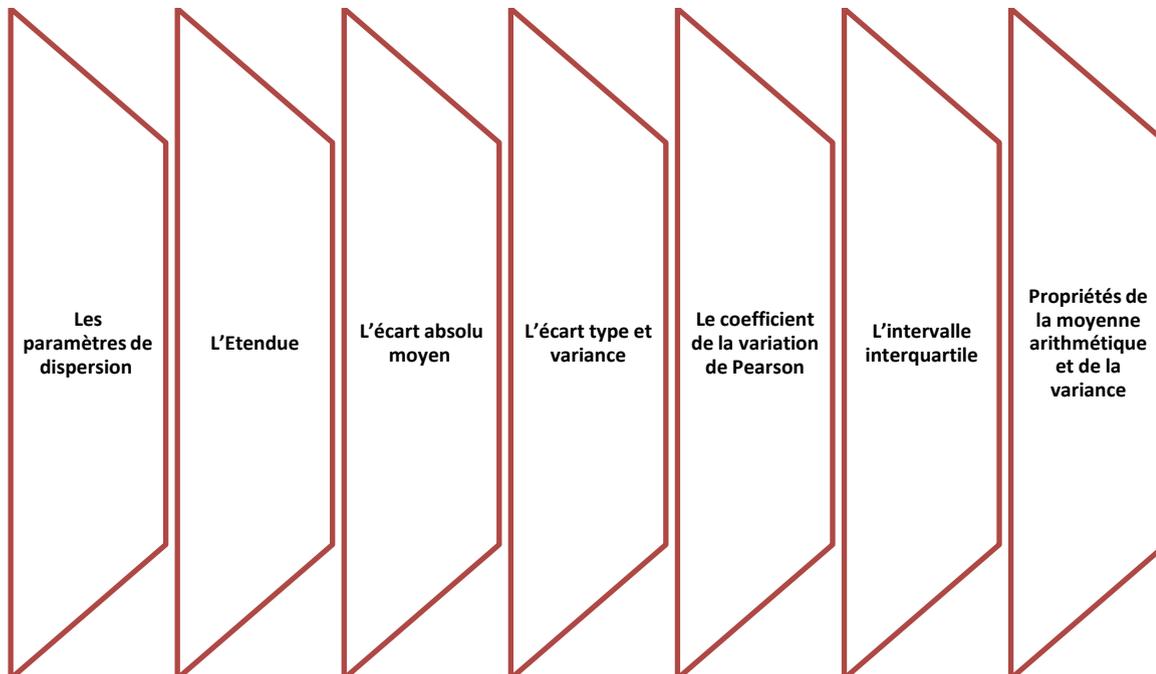
$$N(Q_1) = 12,5 \approx 13 \text{ obs} \Rightarrow 7 < 13 < 22 \Rightarrow Q_1 \in [4,7; 7,1]$$

$$Q_1 = 4,7 + \frac{13-7}{15} \times 2,4 = 5,66$$

$$N(Q_3) \approx 39 \Rightarrow 31 < 39 < 43 \Rightarrow Q_3 \in [9,5; 11,9]$$

$$Q_3 = 9,5 + \frac{39-31}{12} \times 2,4 \Rightarrow Q_3 = 11,09$$

Conférence n°07



V. Les paramètres de dispersion

Définitions :

Exprime l'étendue de la variabilité des observations. Ils donnent des renseignements sur la dispersion et la variabilité dans un groupe, à savoir à quel point les valeurs de la distribution sont homogènes (si les valeurs sont proches de la moyenne ou pas) et hétérogènes (si écart entre la moyenne et les valeurs extrêmes est trop important). Et on peut donner autre définition :

Une caractéristique de dispersion est un nombre qui permet d'estimer dans quelle mesure des observations s'écartent les unes des autres (ou s'écartent de leur valeur centrale).

V.1.L'Etendue :

L'intervalle de variation (IV) ou l'étendue de la série est simplement une façon de résumer le minimum et le maximum de la série en un seul chiffre noté E. On l'obtient ainsi :

$E =$ la plus grande valeur de la variable – la plus petite valeur :

E : Maximum-minimum

Note : L'étendue est ainsi très simple à calculer, mais présente l'inconvénient majeur de ne dépendre que des valeurs extrêmes de la série et ne satisfait pas aux conditions 2 et 5 de Yule. L'étendue ne constitue qu'une mesure très imparfaite de la dispersion d'une série.

Exemple : Un enseignant a corrigé les copies des étudiants de deux groupes de travaux dirigés (Groupe 1 et Groupe 2) de 30 étudiants chacun. Les notes classées par valeurs croissantes sont les suivantes :

Groupe1

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 1 | 2 | 2 | 4 | 4 | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 11 | 11 | 11 | 11 | 12 | 12 | 13 | 14 | 14 | 14 | 15 | 15 | 16 | 17 | 18 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

Groupe2

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 6 | 7 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 9 | 9 | 10 | 11 | 11 | 11 | 11 | 11 | 11 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 12 | 13 | 13 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|

Une observation rapide montre que les notes de la première série vont de très faibles à très élevées tandis que, dans la deuxième, les notes ne descendent pas en dessous de 6, mais ne dépassent pas 13.

Une première manière simple de traduire cette observation est de calculer l'étendue de chaque série.

L'étendue est égale à la différence entre la plus grande et la plus petite des valeurs observées. Dans notre exemple, l'étendue de la 1^{ère} série est 19 alors que celle de la deuxième est 7 ; la dispersion des notes est donc nettement plus forte dans le premier groupe que dans le deuxième.

Variable Statistique discrète : $E = x_{\max} - x_{\min}$

Exemple : Distribution de 500 familles selon le nombre d'enfant

Tableau 3.1 : Calcul de l'étendu (Cas V.S.D)

| x_i | n_i | $N(n_i) \uparrow$ | $N(n_i) \downarrow$ |
|-------|-------|-------------------|---------------------|
| 0 | 50 | 50 | 500 |
| 1 | 120 | 170 | 450 |
| 2 | 130 | 300 | 330 |
| 3 | 110 | 410 | 200 |
| 4 | 90 | 500 | 90 |
| | 500 | | |

$E = 4 - 0 = 4$ enfants par famille

Variable Statistique Continue : $E = e_k - e_1$

Tableau 3.2 : Calcul de l'étendu (Cas V.S.C)

| classe | n_i | $N(n_i) \uparrow$ | $N(n_i) \downarrow$ |
|-----------|-------|-------------------|---------------------|
| 2,3-4,7 | 7 | 7 | 51 |
| 4,7-7,1 | 15 | 22 | 44 |
| 7,1-9,5 | 9 | 31 | 29 |
| 9,5-11,9 | 12 | 43 | 20 |
| 11,9-14,3 | 5 | 48 | 8 |
| 14,3-16,7 | 2 | 50 | 3 |
| 16,7-19,1 | 1 | 51 | 1 |

$E = 19,1 - 2,3 = 16,8$

V.2.L'écart absolu moyen :

C'est la moyenne arithmétique des écarts par rapport à une valeur centrale la moyenne arithmétique \bar{X} ou la médiane Me

V.2.1L'écart absolu moyen par rapport à la moyenne arithmétique \bar{X}

$$e_{\bar{X}} = \sum_{i=1}^k n_i |x_i - \bar{X}|$$

Si on utilise les fréquences relatives f_i au lieu des f_i , n_i l'écart absolu moyen devient

$$e_{\bar{X}} = \sum f_i |x_i - \bar{X}|$$

V.2.2.L'écart des moyen par rapport à la médiane

$$e_{Me} = \frac{1}{N} \sum_{i=1} n_i |x_i - Me|$$

Si on utilise les fréquences relatives f_i au lieu des fréquences absolues n_i l'écart absolu moyen devient

$$e_{Me} = \sum_{i=1} f_i |x_i - Me|$$

Remarque : dans le cas continue, on utilise le centre des classes.

V.3.L'écart type et variance :

V.3.1.La variance :

Consiste à mesurer l'écart moyen de toutes les valeurs observées par rapport à la moyenne. Mesure de la dispersion des valeurs des individus autour de la moyenne

La variance est définit par :

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

Si on utilise les f_i elle devient $V(X) = \sum f_i (x_i - \bar{X})^2$

V.3.2.L'écart type

L'écart type est définit comme étant les racines carré de la variance

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

V.3.3.Détermination de L'écart type et variance par calcule :

V.3.3.1.Cas de Variable Statistique Discrète :

$$V(X) = \frac{1}{N} \sum n_i (x_i - \bar{X})^2 = \sum f_i (x_i - \bar{X})^2$$

$$\delta_x = \sqrt{V(X)}$$

V.3.3.2.Cas de Variable Statistique Continue :

on calcule d'abord le centre de classe C_i

$$C_i = \frac{e_i + e_{i+1}}{2}$$

La variance devient dans le cas continue

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

Détermination de L'écart type et variance par la formule développée :

Le calcul par la formule développée est simple et plus rapide

a) Pour le cas discret :

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \bar{X}^2 \text{ ou } V(X) = \sum f_i x_i^2 - \bar{X}^2,$$

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

Tableau 3.4 : Calcul de la variance (Cas V.S.D)

| x_i | n_i | x_i^2 | $x_i n_i$ | $x_i^2 n_i$ |
|-------|-------|---------|-----------|-------------|
| 1 | 15 | 1 | 15 | 15 |
| 2 | 20 | 4 | 40 | 80 |
| 3 | 25 | 9 | 75 | 225 |
| 4 | 40 | 16 | 160 | 640 |
| Total | N=100 | | 290 | 960 |

$$\bar{X} = \frac{1}{N} \sum_{i=1}^k n_i x_i = \frac{290}{100} = 2,9 \approx 3$$

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i x_i^2 - \bar{X}^2 = \frac{960}{100} - 2,9^2 = 1,19$$

$$\delta_x = \sqrt{V(X)} = \sqrt{1,19} = 1,09$$

b) Pour le cas continue :

$$V(X) = \frac{1}{N} \sum n_i c_i^2 - \bar{X}^2 \text{ ou } V(X) = \sum f_i c_i^2 - \bar{X}^2,$$

Tableau 3.5 : Calcul de la variance (Cas V.S.C)

| classe | n_i | $c_i = (e_{i+1} + e_i) / 2$ | $c_i x n_i$ | $c_i^2 x n_i$ |
|-----------|-------|-------------------------------|-------------|---------------|
| 2,3-4,7 | 7 | $c_1 = (2,3 + 4,7) / 2 = 3,5$ | 24,5 | 85,75 |
| 4,7-7,1 | 15 | 5,9 | 88,5 | 522,15 |
| 7,1-9,5 | 9 | 8,3 | 74,7 | 620,01 |
| 9,5-11,9 | 12 | 10,7 | 128,4 | 1373,88 |
| 11,9-14,3 | 5 | 13,1 | 65,5 | 858,05 |
| 14,3-16,7 | 2 | 15,5 | 31 | 480,5 |
| 16,7-19,1 | 1 | 17,9 | 17,9 | 320,41 |
| Total | 51 | 71,4 | 430,5 | 4260,75 |

$$\bar{X} = \frac{1}{N} \sum_{i=1}^k n_i c_i = \frac{430,5}{51} = 8,44$$

$$V(X) = \frac{1}{N} \sum_{i=1}^k n_i c_i^2 - \bar{X}^2 = \frac{4260,75}{51} - 8,44^2 = 12,29$$

$$\delta_x = \sqrt{V(X)} = \sqrt{12,29} = 3,50$$

Plus l'écart-type est faible, plus les valeurs de la variable sont proches de \bar{x} , plus la moyenne arithmétique est représentative de l'ensemble des valeurs de la variable. Plus il est élevé, moins la moyenne arithmétique est pertinente pour résumer l'ensemble des données. Il est donc indispensable de compléter le calcul de la moyenne par celui de l'écart-type.

L'écart-type peut être supérieur à la moyenne. Plus la distribution est asymétrique, plus il est probable qu'il en sera ainsi.

En ajoutant l'écart-type à la moyenne, on peut obtenir une valeur supérieure à la plus grande valeur de la variable. En soustrayant l'écart-type à la moyenne, on peut obtenir une valeur inférieure à la plus petite valeur de la variable. En revanche, si $\bar{x} + \sigma$ est supérieur à la plus grande valeur de la variable, alors $\bar{x} - \sigma$ ne peut pas être inférieur à la plus petite valeur de la variable. De même, si $\bar{x} - \sigma$ est inférieur à la plus petite valeur de la variable, alors $\bar{x} + \sigma$ ne peut pas être supérieur à la plus grande valeur de la variable.

Exemple⁷ : Dans une usine, le fait d'avoir un écart-type aussi bas que possible peut constituer un objectif de contrôle de qualité. Soit une entreprise qui fabrique un certain composant et qu'un des éléments du contrôle de la qualité consiste à mesurer le diamètre du composant. Chaque composant aura donc son diamètre mesuré. On calculera ensuite le diamètre moyen, puis l'écart-type. Si l'écart-type est faible, cela signifie que les pièces ont dans l'ensemble un diamètre proche de la moyenne, donc que leur diamètre se ressemblent. À la limite, un écart-type nul signifie que toutes les pièces ont le même diamètre. Inversement, plus l'écart-type est élevé, plus il y a de pièces dont le diamètre s'écarte de la moyenne et qui risque de ne pas cadrer avec le système auxquelles elles sont destinées.

Une autre utilité de l'écart-type est de servir de bornes. On regardera par exemple combien de valeurs de la série sont comprises dans l'intervalle défini par : $[\bar{x} - \sigma ; \bar{x} + \sigma[$

Dans l'exemple de $S = \{4, 0, 1, 1, 2, 2, 2, 3, 3, 4, 2, 3, 4, 5, 2, 1, 3, 3, 4, 5\}$, dont nous avons calculé la moyenne simple (2,7), nous obtenons l'intervalle suivant : $[2,7 - 1,345 ; 2,7 + 1,345] = [1,355 ; 4,045]$

⁷ Fabrice Mazerolle, (2008) ; Statistique descriptive, site internet : <http://www.ofarcy.net/documentation/Statistique-descriptive.pdf>

Pour savoir combien d'unités ont des valeurs comprises dans cet intervalle, il faut classer la série par ordre croissant des valeurs de la variable et ensuite encadrer les valeurs comprises dans l'intervalle. On voit ainsi que 14 unités sur 20 ont des valeurs comprises dans l'intervalle, ce qui représente 70% de la population. Si ces valeurs se rapportaient aux caractéristiques des pièces d'un processus industriel on pourrait juger que la qualité est acceptable ou bien au contraire se fixer comme objectif d'augmenter le nombre de valeurs qui sont dans cet intervalle. On voit que cet intervalle est avec l'intervalle interquartile une autre façon de mesurer la dispersion d'une série.

V.4.Le coefficient de la variation de Pearson :

Indicateur de dispersion utilisée dans l'analyse d'une variable numérique, il est égal au rapport de l'écart -type par la moyenne. Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande. Ce rapport permet de comparer la dispersion autour de la moyenne de variables numériques possédant des échelles de valeurs différentes.

$$\alpha = \frac{\delta(x)}{\bar{X}} > 0$$

Il permet de comparer une Série Statistique (S.S) ou deux S,S

Plus le coefficient de variation est élevé, plus la dispersion est forte et plus le coefficient de variation est faible, plus les valeurs de la variable sont proches de \bar{X} , donc plus la moyenne est bien représentative de l'ensemble des valeurs de la variable.

V.5.L'intervalle interquartile

L'intervalle de variation ne donne qu'une idée imprécise et parfois fautive de la dispersion de la série, car les valeurs extrêmes peuvent être exceptionnelles et le reste de la population statistique être concentré sur un intervalle beaucoup plus restreint. On peut parfaitement s'en rendre compte dans par exemple des statistiques résumées de la population, superficie pour certains pays, où les intervalles de variation sont énormes pour les superficies et pour les populations, car de toutes petites îles sont comparées avec des pays ayant presque la taille d'un continent. D'où l'idée de calculer l'intervalle interquartile qui donne une idée plus précise de la dispersion des valeurs d'une série (ou d'une distribution).⁸

Définition :

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

La formule est donnée par :

$$IQ = Q3 - Q1$$

Note : L'écart interquartile n'est pas sensible aux valeurs extrêmes.

⁸ Fabrice Mazerolle,(2008) ;Statistique descriptive, site internet : <http://www.ofarcy.net/documentation/Statistique-descriptive.pdf>

Le coefficient de variation de l'écart interquartile :

noté $\alpha = \frac{Q_3 - Q_1}{E} = \frac{IQ}{E}$ ($E = X_{\max} - X_{\min}$; E est l'étendue de la série, il est déjà défini auparavant)

V.6. Propriétés de la moyenne arithmétique et de la variance

V.6.1. Propriétés de la moyenne arithmétique

- **La moyenne arithmétique est linéaire :**

Cela signifie que si une variable x subit une transformation linéaire et devient : $y = aX + b$ ou a et b soit des paramètres connus appartiennent à \mathbb{R} ($a, b \in \mathbb{R}$), alors sa moyenne \bar{X} subit la même transformation : $\bar{y} = a\bar{X} + b$

Sachant que \bar{X} est la moyenne arithmétique de X et $V(x)$ est la variance de X alors $\bar{y} = a\bar{X} + b$ est la moyenne de y .

En effet : $y_i = ax_i + b$

$$\bar{y} = \frac{\sum_{i=1}^k n_i y_i}{N} = \frac{\sum_{i=1}^k n_i (ax_i + b)}{N} = \frac{\sum_{i=1}^k a \times n_i \times x_i}{N} + \frac{\sum_{i=1}^k b \times n_i}{N} = a\bar{X} + b$$

- **La somme des écarts à la moyenne est nulle**

Soit : $\sum_{i=1}^k n_i (x_i - \bar{X})$ la somme des écarts à la moyenne. Notons (1) cette somme.

On a : $\bar{X} = \frac{\sum_{i=1}^k n_i x_i}{N}$ et $\sum_{i=1}^k n_i = N$; alors : (1) = $\sum_{i=1}^k (n_i \times x_i - n_i \times \bar{X}) = \sum_{i=1}^k (n_i \times x_i) - \sum_{i=1}^k n_i \times \bar{X} = N\bar{X} - N\bar{X} = 0$

V.6.2. Propriétés de la variance et de l'écart-type

- **multiplication par une constante**

Si on multiplie un caractère x par un facteur a , que devient-il de l'écart-type et de la variance du caractère ? Soit : $y = ax$; $a \in \mathbb{R}$ alors :

$$V(y) = V(ax) = \frac{\sum_{i=1}^k a^2 n_i x_i^2}{N} - a^2 \times \bar{X}^2 = a^2 \times \left[\frac{\sum_{i=1}^k n_i x_i^2}{N} - \bar{X}^2 \right]$$

$V(y) = a^2 V(x)$ est variance de y . L'écart-type est multiplié par le facteur a et la variance par le carré du facteur.

- **Addition d'une constante**

Si on ajoute une constante b à un caractère x , que devient-il de la variance et de l'écart-type du caractère ? Soit : $y = x + b$

$$y - \bar{y} = x_i + b - \bar{X} - b = x_i - \bar{X} \text{ et } \bar{y} = \bar{X} + b$$

$$V(y) = \frac{\sum_{i=1}^k n_i \times (y_i - \bar{y})^2}{N} = \frac{\sum_{i=1}^k (x_i + b - \bar{X} - b)^2}{N} \text{ Soit } V(y) = V(x)$$

L'écart-type et la variance du caractère ne sont pas modifiés

La variance et l'écart-type sont toujours positifs.

Par exemple si $E(x) = 2$ et $V(x) = 14$

Et on a $y = -x + 1$ donc $E(y) = -E(x) + 1 = -1$ et $V(y) = (-1)^2 V(x) = 14$

Conférence n°08

Les
moments

Moment
simple

Moment
centré

VI. Les moments :

Nous pouvons généraliser la notion de moyenne d'ordre k des valeurs des variables centrées, et définir ainsi des moments de la distribution d'une variable X.

VI.1. Moment simple

Moment simple d'ordre K si $X_0=0$ et nous le notons m_k

$$M_1(x) = \frac{\sum n_i x_i}{\sum n_i} \text{ moment du 1}^{\text{er}} = \text{ordre}$$

$$M_k(x) = \frac{\sum n_i x_i^k}{N} \text{ moment du } k^{\text{er}} = \text{ordre}$$

Le moment d'ordre K d'une variable statistique X est la quantité définie par :

$$M_k = \frac{\sum n_i (x_i - x_0)^k}{\sum_{i=1}^n n_i}$$

VI.1. Moment centré

Si $X_0 = \bar{X}$ nous parlons de moment centré d'ordre K et le notons M_k

(K entier positif quelconque).

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

$M_1(X) = 0$ (somme de var centres nulle)

$M_2(X) = \delta^2(X) = V(X)$ variance est le plus important

$$M_2(X) = V(X) = \delta^2(X) = \frac{\sum x_i^2}{N} - \bar{X}^2 = m_2 - m_1^2$$

$$M_3(X) = m_3 - 3m_1 m_2 + 2m_1^3 \quad / \quad M_4(X) = m_4 - 4m_1 m_3 + 6m_1^2 m_2 - 3m_1^4$$

Sont utilisés dans le calcul des caractéristiques de forme

Exemple : calcul des moments

Soit la série statistique :-3,-2,0,1,3,7

Calculer les moments d'ordre 1,2,3,4

Réponse puisque la série contient six valeurs on a : $m_r = \frac{1}{6} \sum_{i=1}^6 x_i^r$

Tableau 4.1 : calcul des moments

| r | x_i^r | x_1^r | x_2^r | x_3^r | x_4^r | x_5^r | x_6^r | Σx_i^r | m_r |
|---|---------|---------|---------|---------|---------|---------|---------|----------------|------------|
| 1 | x_i^1 | -3 | -2 | 0 | 1 | 3 | 7 | 6 | 6/6=1 |
| 2 | x_i^2 | 9 | 4 | 0 | 1 | 9 | 49 | 72 | 72/6=12 |
| 3 | x_i^3 | -27 | -8 | 0 | 1 | 27 | 343 | 336 | 336/6=56 |
| 4 | x_i^4 | 81 | 16 | 0 | 1 | 81 | 2401 | 2580 | 2580/6=430 |

Calculer les moments centrés d'ordre 1,2,3

Réponse $\bar{X} = \frac{-3-2+0+1+3+7}{6} = 1$

Tableau 4.2 : calcul des moments centrés

| r | $(x_1-1)^r$ | $(x_2-1)^r$ | $(x_3-1)^r$ | $(x_4-1)^r$ | $(x_5-1)^r$ | $(x_6-1)^r$ | $\sum_{i=1}^6 (x_i-1)^r$ |
|---|-------------|-------------|-------------|-------------|-------------|-------------|--------------------------|
| 1 | -4 | -3 | -1 | 0 | 2 | 6 | 0 |
| 2 | 16 | 9 | 1 | 0 | 4 | 36 | 66 |
| 3 | -64 | -27 | -1 | 0 | 8 | 216 | 132 |
| 4 | 256 | 81 | 1 | 0 | 16 | 1296 | 1650 |

On a donc $M_1=0/6=0$ $M_3=132/6=22$

$M_2=66/6=11$ $M_4=1650/6=275$

Autre méthode :

$\mu_1=0$ (connu)

$\mu_2=m_2-m_1=12-1=11$

$\mu_3=m_3-3 m_2 m_1+2 m_1^3=56-3 \times 12 \times 1+2 \cdot 1=22$

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

$=430-4 \times 56 \times 1+6 \times 12 \times 1-3 \times 1=430-224-72-3=275$

Conférence n°09

Caractéristiques de forme

- **Asymétrie**
- **APLATISSEMENT**

VII. Caractéristiques de forme

Ces caractéristiques permettent de préciser l'allure de la courbe de effectif (ou de fréquences) sans avoir besoin de la tracer

Par rapport à une courbe (idéale) dite courbe normale, nous définissons deux sortes de caractéristiques :

- Caractéristiques d'asymétrie
 - Caractéristiques d'aplatissement
- } sont définies à partir des moments

VII.1. Asymétrie :

L'asymétrie [Skewness en anglais] : est basée sur le moment centré d'ordre 3.

Pour une distribution symétrique, la moyenne arithmétique est égale à la médiane égale au mode. De plus, les moments centrés d'ordre impair sont nuls pour une distribution symétrique

Une distribution est symétrique si $\bar{X} = M_e = \text{Mode}$.

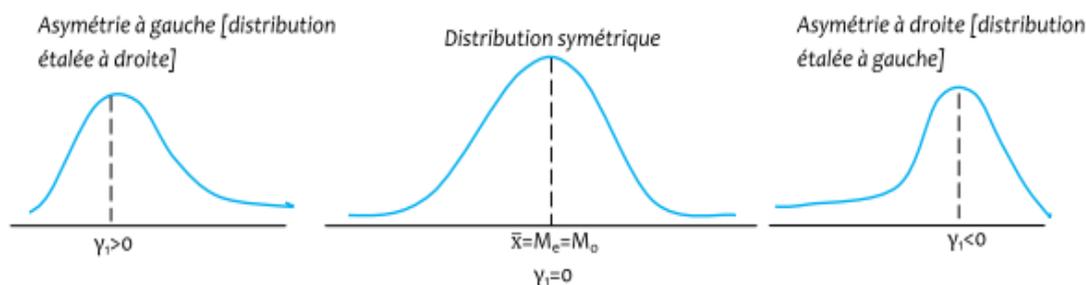
Q1 et Q3 équidistants de la médiane (situés à égale distance).

Elle est représentée par un histogramme dit « en cloche » la position respective de ces éléments va permettre le calcul des coefficients d'asymétrie qui sont nuls en cas de symétrie et s'éloignent d'autant plus de 0 que la distribution est plus asymétrique.

Le coefficient d'asymétrie de Fisher⁹ : $\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$

γ_1 est nul pour une distribution symétrique, telle que la loi normale.

Figure 4.1 : Les formes d'asymétrie d'une distribution statistique



- Coefficient d'asymétrie

Coefficient de fisher : $\gamma_1 = \frac{\mu_3(X)}{\delta^3(X)}$

⁹ https://www.academia.edu/19786928/Econometrie_suite_lareq

symétrique si $\gamma_1=0$

γ_1 est positif lorsque la distribution est plus accentuée à droite

γ_1 est négatif lorsque le mode est situé à gauche de la médiane

La courbe dite oblique (à gauche) et vice-versa

- **Coefficient de Yule :**

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

| | | |
|-------------------|---|----------------------------------|
| C _y =0 | ⇒ | courbe symétrique |
| C _y >0 | ⇒ | étalé à droite Mo à gauche de Me |
| C _y <0 | ⇒ | étalé à gauche Mo à droite de Me |

- **Coefficient de Pearson**

$$B_1 = \frac{\bar{X} - M_0}{S(X)}$$

| | | | |
|---|---|--------------------|----------------------------------|
| B ₁ = $\frac{\bar{X} - M_0}{S(X)}$ | | B ₁ = 0 | symétrie |
| B ₁ > 0 | → | C _y > 0 | étalé à droite Mo à gauche de Me |
| B ₁ < 0 | → | C _y < 0 | étalé à gauche Mo à droite de Me |

$$B_2 = \frac{\mu_3^2(x)}{\mu_2^3(x)}$$

| | |
|------------------------|---|
| B ₂ = 0 | symétrie |
| B ₂ ≠ 0 | asymétrie |
| Le signe de $\mu_3(x)$ | |
| Si $\mu_3(x) > 0$ | → C _y > 0 ⇒ étalé à droite Mo à gauche de Me |
| Si $\mu_3(x) < 0$ | → C _y < 0 ⇒ étalé à gauche Mo à droite de Me |

Utilité de la boîte à moustache pour déterminer la forme d'une distribution¹⁰ :

Suivant la position de la médiane au sein de la boîte, on peut en déduire des informations sur la forme de la distribution.

- 1) Si la médiane est proche du centre de la boîte, c'est que la distribution est symétrique.
- 2) Si la médiane est à gauche du centre de la boîte, c'est que la distribution est étalée à droite.
- 3) Si la médiane est à droite du centre de la boîte, c'est que la distribution est étalée à gauche.

¹⁰ Fabrice MAZEROLLE (2005), « Statistique Descriptive », © Gualino éditeur, EJA – Paris – 2006, ISBN 2 - 84200 - 891 - X, p80

De même, en comparant la longueur respective de chaque moustache, on peut en déduire des informations sur la forme de la distribution.

- 1) Si les moustaches sont à peu près de la même longueur, c'est que la distribution est symétrique.
- 2) Si la moustache de droite est plus longue que la moustache de gauche, c'est que la distribution est étalée à droite.
- 3) Si la moustache de gauche est plus longue que la moustache de droite, c'est que la distribution est étalée à gauche.

Exemple : Soit les trois séries suivantes A,B et C, dont les distributions sont respectivement symétrique (Me=3), étalée à droite (Me = 2) et étalée à gauche (Me = 4) :

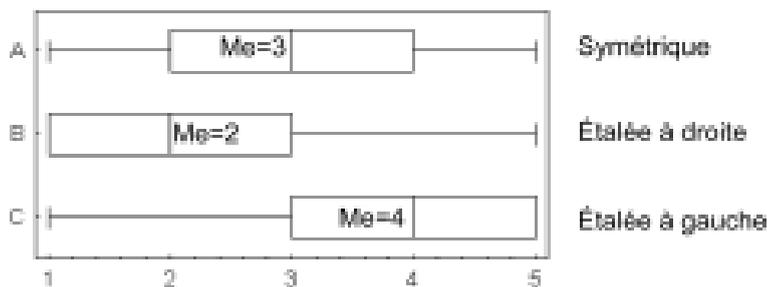
A = {1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5}

B = {1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 5, 5}

C = {1, 1, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5}

Les boîtes à moustaches correspondantes ont bien les caractéristiques précitées :

Figure 4.2 : la boîte à moustache pour déterminer la forme d'une distribution



VII.2 APLATISSEMENT

L'aplatissement [Kurtosis en grec, qui signifie bosse] : est basé sur le moment centré d'ordre 4 et permet de mesurer l'importance des queues d'une distribution ou son aplatissement.

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

Une courbe plus aplatie que la courbe normale est dite «platicurtique »

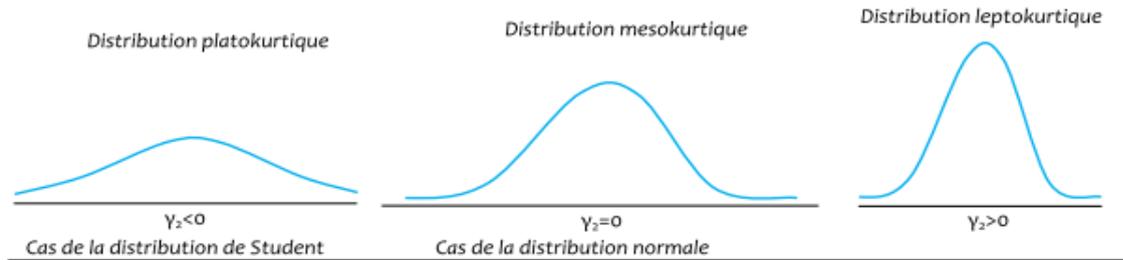
Une courbe moins aplatie que la courbe normale est dite «leptocurtique »

Le coefficient d'aplatissement de Fisher¹¹ $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$

γ_2 est nul pour une distribution mesokurtique.

¹¹ https://www.academia.edu/19786928/Econometrie_suite_lareq

Figure 4.3 : Les formes d'aplatissements d'une distribution statistique



• **Principaux coefficients d'aplatissements :**

Coefficient de Pearson :

$B_2 = \frac{M_4(x)}{M_2^2(x)} = \frac{M_4(x)}{S^4(X)}$ nous avons toujours $M_4 > M_2^2$ c'est pourquoi le coefficient B_3 est toujours > 1

Si $B_3 = 3 \Rightarrow$ courbe normale

Si $B_3 > 3 \Rightarrow$ Leptokurtique moins aplatie que la courbe normale

Si $B_3 < 3 \Rightarrow$ courbe platicurtique plus aplatie que la courbe normale

Sa valeur de passe rarement 5

Coefficient de Fisher :

$F_2 = B_3 - 3 = \frac{M_4(x)}{S^4(X)} - 3$

Si $F_2 = 0 \Rightarrow$ courbe normale

Si $F_2 > 0 \Rightarrow$ moins aplatie

Si $F_2 < 0 \Rightarrow$ plus aplatie

$F_2 > -2$ toujours (du fait l'inégalité $M_4(x) > S^4(X)$)

Tableau 4.3 : Autre méthode du calcul des moments centrés

| | | | | | | | |
|----------|---------|---------|---------|---------|---------|---------|---------|
| classe | [15,25[| [25,35[| [35,45[| [45,55[| [55,65[| [65,75[| [75,85[|
| effectif | 5 | 15 | 32 | 40 | 66 | 25 | 17 |

Calculer les moments centrés d'ordre 2,3 et4 puis la coefficient d'aplatissement de Pearson

Réponse :

x_i : centre de classe i et $z_i = (x_i - 50)/10$ en utilisant changement de variable

| | (1) | (2) | (3) | (4) | (5) | (6) |
|----------|-------|-------|------------------|--------------------|--------------------|--------------------|
| | | | (1) x(2) | (2)x(3) | (2)x(4) | (2)x(5) |
| x_i | n_i | z_i | $n_i \times z_i$ | $n_i \times z_i^2$ | $n_i \times z_i^3$ | $n_i \times z_i^4$ |
| 20 | 5 | -3 | -15 | 45 | -135 | 405 |
| 30 | 15 | -2 | -30 | 60 | -120 | 240 |
| 40 | 32 | -1 | -32 | 32 | -32 | 32 |
| 50 | 40 | 0 | 0 | 0 | 0 | 0 |
| 60 | 66 | 1 | 66 | 66 | 66 | 66 |
| 70 | 25 | 2 | 50 | 100 | 200 | 400 |
| 80 | 17 | 3 | 51 | 153 | 459 | 1377 |
| Σ | 200 | | 90 | 456 | 438 | 2520 |

$$\text{On a } m_1(z) = \bar{z} = \frac{90}{200} = 0,45$$

$$m_2(z) = \frac{456}{200} = 2,28 \text{ d'où } M_2(z) = m_2(z) - \bar{z}^2 = 2,28 - (0,45)^2 = 2,0775$$

$$\text{Donc : } \mu_2(x) = 10^2 \mu_2(z) = 207,75$$

$$m_3(z) = \frac{438}{200} = 2,19 \text{ d'où } \mu_3(z) = m_3(z) - 3 m_2(z) \bar{z} + 2 \bar{z}^3$$

$$M_3 = 2,19 - 3 \times 2,28 \times 0,45 + 2 \times 0,45^3 = 0,70575$$

$$\text{Donc } \mu_3(x) = 10^3 \mu_3(z) = -705,75$$

$$. m_4(z) = \frac{2520}{2000} = 12,6$$

$$\text{D'où } \mu_4(z) = m_4(z) - 4 m_3(z) \bar{z} + 6 m_2(z) \bar{z}^2 - 3 \bar{z}^4$$

$$12,6 - (4 \times 2,19 \times 0,45) + (6 \times 2,28 \times 0,45)^2 - (3 \times 0,45)^4$$

$$\approx 11,305182 \text{ donc } M_4(x) = 10^4 M_4(z) \approx 113051,82$$

$$F = \frac{-705,75}{29944,2} = 0,236 \text{ coefficient d'aplatissement}$$

$$B = \frac{113051,82}{43160,0625} \approx 2,62 \text{ légèrement hypo-normal}$$

Exercices résolus

Exercice 1. La série statistique suivante représente le nombre d'absences d'un échantillon de 10 étudiants, à un cours de statistiques.

2, 8, 5, 2, 12, 3, 5, 2, 8, 3.

Trouver la valeur du mode, la médiane et la moyenne.

Exercice 2. Déterminer le mode et la médiane des séries statistiques suivantes :

| | | | | |
|-------------------|--|---------------------------------|--|--------------------------------------|
| 5, 9, 1, 4, 5. | | 1, 1, 8, 4, 3, 4, 4, 11, 15, 12 | | 9, 3, 9, 3. |
| 5, 5, 5, 5, 5, 5. | | 5, 9, 8, 2, 12, 10, 6 | | 5, 5, 6, 6, 6, 7, 7, 8, 8, 8, 12, 12 |

Exercice 3. Un responsable du service marketing d'une grande marque d'ordinateurs (portables et de bureau) fait une enquête dans une université pour savoir combien d'ordinateurs (portables et de bureau) il y a dans leur foyer.

| | | | | | | |
|--------------------|-----|-----|------|-----|-----|----|
| Nombre | 0 | 2 | 4 | 5 | 6 | 8 |
| Nombre d'étudiants | 300 | 880 | 1800 | 900 | 100 | 20 |

1. Donner la valeur du mode. Interpréter.
2. Trouver la valeur de la médiane.
3. Déterminer les quartiles Q_1 et Q_3 . Interpréter.

Exercice 4. Dans un grand magasin, en fin de journée, le comptable classe les chèques encaissés d'après le montant des achats:

| | | | | | |
|------------------------|-----------|--------------|--------------|--------------|--------------|
| Montant des chèques en | [0, 1000[| [1000, 2000[| [2000, 3000[| [3000, 4000[| [4000, 5000[|
| Nombre de chèques | 18 | 54 | 45 | 36 | 27 |

1. Déterminer par le calcul la valeur du mode et de la médiane.
2. Déterminer les quartiles Q_1 et Q_3 . Commenter ces résultats.

Exercice 5. Une enquête sur la fréquentation d'un grand restaurant pendant 18 jours a conduit aux résultats suivants :

| | | | | | | | | |
|------------------------|----|----|----|----|----|----|----|----|
| Fréquentation (valeur) | 50 | 52 | 53 | 55 | 57 | 58 | 60 | 62 |
| Nombre de jours | 4 | 2 | 2 | 1 | 2 | 2 | 3 | 2 |

1. Calculer la fréquentation moyenne du restaurant pendant ces 18 jours.
2. Déterminer le mode, le premier, le deuxième et le troisième quartile de cette série.
3. Calculer la variance et l'écart type.
4. Soit Y une autre variable statistique s'écrivant en fonction de X (fréquentations) telle que : $Y = 4X + 3$. En déduire la moyenne arithmétique ainsi que la variance de la variable Y . De même pour $X + 1$.

Exercice 6. On donne la distribution statistique de 400 entreprises selon leur bénéfice en dinars.

| | | | | | |
|--------------|-----|----------|----------|----------|----------|
| Bénéfice x | [0; | [200000; | [400000; | [600000; | [800000; |
| Entreprises | 80 | 50 | 100 | 120 | 50 |

1. Identifier la variable statistique et préciser sa nature.
2. Donner le nombre d'entreprises dont le bénéfice est entre 400000 et 800000 dinars.
3. Calculer les effectifs cumulés croissants.
4. Calculer le bénéfice moyen, la valeur du mode et de la médiane.
5. Quel commentaire pouvez-vous faire sur la nature de l'asymétrie?
6. Supposons qu'une erreur s'est produite et que le nombre d'entreprises dont le bénéfice est entre 0 et 200000 dinars n'était pas 80 mais une autre valeur qu'on va noter $n > 120$.
7. Trouver n pour que le mode soit égal à 120000 dinars.
8. Que peut-on dire de la valeur de la médiane?

Exercice 7. Une étude sur le chiffre d'affaires d'une population de PME (Petites et Moyennes Entreprises) a permis d'obtenir les résultats suivants (en millions de dollars).

| | | | | | |
|----|--------|---------|-----|------------|----------|
| Le | Ecart- | Le mode | La | Le premier | L'étendu |
| 2 | 3 | 4 | 4,6 | 4 | 6 |

1. Quel est le chiffre d'affaires qui correspond au plus grand nombre de ces entreprises?
2. Quel est le chiffre d'affaires le plus grand dans cette population de PME?
3. Sachant que le coefficient de variation de l'écart interquartile est égal à 0,5.
Déterminer les quartiles Q_2 et Q_3 .
4. Donner la valeur de la variance.

Exercice 8. On dispose des résultats d'une enquête concernant le nombre d'unités vendues d'un produit donné par 40 détaillants.

| | | | | | |
|-------------------|----|----|----|----|----|
| Nombre d'unités | 30 | 40 | 55 | 70 | 90 |
| Détaillants n_i | 11 | 11 | 2 | 14 | 2 |

1. Identifier la variable statistique et préciser sa nature.
2. Retrouver la valeur du mode, le nombre moyen d'unités vendues et la médiane.
3. Déterminer les quartiles Q_1 et Q_3 .
4. Déterminer le nombre d'unités vendues ' n ' de ' n ' détaillants supplémentaires pour que la moyenne soit égale à ' n '. Donner dans ce cas la valeur du mode.

Exercice 9. Les montants des ventes réalisées par une librairie sont présentés selon le tableau ci-dessous.

| | | | | | | |
|--------------|--------|--------|--------|------------|--------|--------|
| Montant en | [1000, | [1500, | [3000, | [3500,4500 | [4500, | [5000, |
| Effectif n | 500 | 3000 | 1500 | 2000 | 1000 | 1000 |

1. Faire une représentation par histogramme.
2. Déterminer par le calcul la valeur du mode.
3. Trouver la valeur de la médiane.
4. Tracer la courbe des effectifs cumulés croissants.

Exercice 10. Le calcul du salaire moyen ainsi que l'écart-type des employés dans deux entreprises A et B , nous a permis d'avoir les résultats ci-après,

Entreprise A : le salaire moyen $X_A = 22000$ dinars et l'écart-type $U_A(X) = 2000$ dinars.

Entreprise B : le salaire moyen $X_B = 28000$ dinars et l'écart-type $U_B(X) = 6000$ dinars.

Que peut-on dire à propos de la dispersion des salaires?

Exercice 11 : Le prix de l'essence a augmenté de 20 % l'an dernier et de 10 % cette année. Quelle est la hausse moyenne de l'essence ?

Exercice 12 : Une société a vu son bénéfice augmenter ces trois dernières années : de 10 % la première année, de 21 % la deuxième et de 2% la troisième. Quelle est son augmentation annuelle moyenne ?

Exercice 13 : Si l'inflation d'un pays est de 5% la première année et de 15% la suivante, calculer l'augmentation moyenne des prix.

Exercice 14 : Yassine fait un aller - retour entre une ville A à une ville B à vélo. A l'aller, sa vitesse moyenne est de 20 km/h et au retour, elle est de 26 km/h. Quelle est sa vitesse moyenne ?

Exercice 15 : Dans une entreprise de fabrication 3 ouvriers produisent des pièces. L'ouvrier A met 10 min par pièce, l'ouvrier B met 15 min et l'apprenti C met 20 min. Quel est le temps moyen de réalisation d'une pièce ?

Solutions des exercices

Rappel.

Le mode M_0 est la valeur de variable statistique qui correspond à l'effectif le plus élevé. C'est donc la valeur la plus répétée.

Pour avoir la valeur de la médiane M_e , il faut d'abord classer les observations selon un ordre croissant.

1. Si le nombre d'observations est impair, alors, la médiane $M_e =$ la $(n+1)^{\text{ème}}$ observation.

2. Si le nombre d'observations est pair, alors,

$$M_e = \frac{n^{\text{ème}} \text{ observation} + (n+1)^{\text{ème}} \text{ observation}}{2}.$$

Corrigé de l'exercice 1.

Le mode $M_0 = 2$.

La médiane $M_e = ?$

2, 2, 2, 3, 3, 5, 5, 8, 8, 12.

10 observations (nombre pair, de la forme $2n$). Ainsi,

$$2n = 10 \Rightarrow n = 5 \Rightarrow n + 1 = 6.$$

$$\text{La médiane } M_e = \frac{5^{\text{ème}} \text{ observation} + 6^{\text{ème}} \text{ observation}}{2} = \frac{3 + 5}{2} = 4.$$

La moyenne $\bar{X} = 5$.

Corrigé de l'exercice 2.

- 1, 4, 5, 5, 9.

Le mode $M_0 = 5$.

La médiane $M_e = ?$

5 observations (nombre impair, de la forme $2n + 1$). Ainsi

$$2n + 1 = 5 \Rightarrow n = 2 \Rightarrow n + 1 = 3.$$

La médiane $M_e =$ la $3^{\text{ème}}$ observation $= 5$.

- 2, 5, 6, 8, 9, 10, 12. Le mode n'existe pas.

La médiane $M_e =$ la $4^{\text{ème}}$ observation $= 8$.

- 5, 5, 6, 6, 6, 7, 7, 8, 8, 8, 12, 12.

Deux valeurs de mode : 6 et 8. Dans ce cas la distribution statistique est dite bimodale.

La médiane $M_e = 7$.

- 5, 5, 5, 5, 5, 5, 5. Le mode n'existe pas.

La médiane $M_e = 5$.

- 1, 1, 3, 4, 4, 4, 8, 11, 12, 15.

Le mode $M_0 = 4$.

La médiane $M_c = ?$

10 observations (nombre pair, de la forme $2n$). Ainsi,

$$2n = 10 \Rightarrow n = 5 \Rightarrow n + 1 = 6.$$

$$\text{La médiane } M_c = \frac{5^{\text{ème}} \text{ observation} + 6^{\text{ème}} \text{ observation}}{2} = \frac{4 + 4}{2} = 4.$$

- 3, 3, 9, 9. Le mode n'existe pas.

$$\text{La médiane } M_c = \frac{3 + 9}{2} = 6.$$

Corrigé de l'exercice 3.

| x_i | n_i | $N(x_i) \nearrow$ |
|----------|-------|-------------------|
| 0 | 300 | 0 |
| 2 | 880 | 300 |
| 4 | 1800 | 1180 |
| 5 | 900 | 2980 |
| 6 | 100 | 3880 |
| 8 | 20 | 3980 |
| Total | 4000 | 4000 |

$N/4 = 4000/4 = 1000$
 $N/2 = 4000/2 = 2000$
 $3N/4 = 3 \cdot 4000/4 = 3000$

1. Le mode $M_0 = 4$ (Le nombre d'ordinateurs auxquels correspond le plus grand nombre d'étudiants).
2. La médiane $M_c = 4$.
3. Les quartiles : $Q_1 = 2$ et $Q_3 = 5$.

Ainsi,

25% des étudiants possèdent moins de 2 ordinateurs.

50% des étudiants possèdent entre 2 et 5 ordinateurs.

Les 25% restants possèdent plus de 5 ordinateurs.

Corrigé de l'exercice n°4 :

| ec_i | n_i | $N(ec_i) \uparrow$ | |
|-----------------|-----------|--------------------|--|
| $[0 - 1000[$ | 18 | 18 | |
| $[1000 - 2000[$ | 54 | 72 | $\frac{N}{4} = \frac{180}{4} = 45$ |
| $[2000 - 3000[$ | 45 | 117 | $\frac{N}{2} = \frac{180}{2} = 90$ |
| $[3000 - 4000[$ | 36 | 153 | $\frac{3N}{4} = \frac{3 \cdot 180}{4} = 135$ |
| $[4000 - 5000[$ | 27 | 180 | |
| Total | $N = 180$ | | |

La classe médiane est $[2000, 3000[$

$$\text{La médiane } M_e = e_i + \left[n_i \left(\frac{\frac{N}{2} - N(ec_i) \uparrow}{n_i} \right) \right]$$

$$\left\{ \begin{array}{l} e_i = 2000 \\ n_i = 3000 - 2000 = 1000 \\ N(ec_i) \uparrow = N(2000) \uparrow = 72 \\ m_i = 45 \end{array} \right.$$

$$\Rightarrow M_e = 2000 + \left[1000 \left(\frac{90 - 72}{45} \right) \right] = 2000 + \left[1000 \cdot \frac{18}{45} \right] = 2400 \text{ DA.}$$

2) Les quartiles :

$$q_1 = e_i + \left[n_i \left(\frac{\frac{N}{4} - N(ec_i) \uparrow}{n_i} \right) \right]$$

$$\text{Donc } q_1 : [1000, 2000[\leftarrow \frac{N}{4} = \frac{180}{4} = 45$$

$$\text{Ainsi } \left\{ \begin{array}{l} e_i = 1000 \\ n_i = 2000 - 1000 = 1000 \\ N(ec_i) \uparrow = N(1000) \uparrow = 18 \text{ et } m_i = 54 \end{array} \right.$$

$$q_1 = 1000 + \left[1000 \left(\frac{45-18}{54} \right) \right] = 1500 \text{ DA.}$$

$$q_3 = e_i + \left[\alpha_i \left(\frac{3N}{4} - N(e_i) \right) \right]$$

$$[3000-4000] \leftarrow \frac{3N}{4} = \frac{3 \cdot 180}{4} = 135$$

$$\begin{cases} e_i = 3000 \\ \alpha_i = 4000 - 3000 \\ N(e_i) \uparrow = N(3000) \uparrow = 117 \\ n_i = 36 \end{cases}$$

$$\Rightarrow q_3 = 3000 + \left[1000 \left(\frac{135 - 117}{36} \right) \right] = 3500 \text{ DA.}$$

Corrigé de l'exercice n°5 :

$$\begin{aligned}
 1/ \bar{x} &= \frac{1}{N} \sum m_i x_i \\
 &= \frac{1}{18} (4 \cdot 50 + 2 \cdot 52 + 2 \cdot 53 + 1 \cdot 55 + 2 \cdot 57 + 2 \cdot 58 + 3 \cdot 60 + 2 \cdot 62) \\
 &= \frac{999}{18} = 55,5
 \end{aligned}$$

2/

| x_i | m_i | $N(x_i) \uparrow$ |
|-------|----------|-------------------|
| 50 | 4 | 4 |
| 52 | 2 | 6 |
| 53 | 2 | 8 |
| 55 | 1 | 9 |
| 57 | 2 | 11 |
| 58 | 2 | 13 |
| 60 | 3 | 16 |
| 62 | 2 | 18 |
| Total | $N = 18$ | |

Annotations:

- Between 52 and 53: $\frac{2}{4} = \frac{18}{4} = 4,5$
- Between 55 and 57: $\frac{2}{2}$
- Between 60 and 62: $\frac{3N}{4} = \frac{3 \cdot 18}{4} = 13,5$

Le mode $M_0 = 50$

$$Q_1 = 52, \quad Q_2 = M_e = \frac{55 + 57}{2} = 56, \quad Q_3 = 60.$$

3/ \bar{x} est déjà calculée (fréquentation moyenne).

$$\begin{aligned}
 \bullet V(x) &= \left[\frac{1}{N} \sum m_i x_i^2 \right] - (\bar{x})^2 \\
 &= \left[\frac{1}{18} (4 \cdot 50^2 + 2 \cdot 52^2 + 2 \cdot 53^2 + 1 \cdot 55^2 + 2 \cdot 57^2 + 2 \cdot 58^2 + 3 \cdot 60^2 + 2 \cdot 62^2) \right] \\
 &\quad - (55,5)^2 \\
 &= \frac{55765}{18} - (55,5)^2 = 3098,05 - 3080,25 = 17,8
 \end{aligned}$$

$$\bullet \text{L'écart-type } \sigma(x) = \sqrt{V(x)} = \sqrt{17,8} = 4,21$$

4/ Si $Y = \alpha X + \beta$. Alors,

$$\bar{Y} = \alpha \bar{X} + \beta \text{ et } V(Y) = \alpha^2 V(X).$$

Ainsi, pour $Y = 4X + 3$, on a, $\bar{Y} = 4\bar{X} + 3 = (4 \times 55.5) + 3$ et $V(Y) = 4^2 V(X) = 16 \times 17.8$

De même, en posant $Z = X + 1$.

On a, $\bar{Z} = \bar{X} + 1 = 55.5 + 1 = 56.5$ et $V(Z) = V(X) = 17.8$.

Corrigé de l'exercice 6.

1. Salaire moyen :

$$\begin{aligned} \bar{X} &= \frac{1}{N} \sum_i n_i c_i \\ &= \frac{1}{200} [(42 \times 8500) + (60 \times 9500) + (69 \times 10500) + (16 \times 11500) + (13 \times 12500)] \\ &= 9990 \text{ dinars.} \end{aligned}$$

2. • Le mode M_0 :

La classe modale est [10000; 11000[. Ainsi,

$$M_0 = e_i + \left[\frac{a_i \times \alpha_1}{\alpha_1 + \alpha_2} \right]$$

$$\begin{cases} e_i = 10000 \\ a_i = 11000 - 10000 = 1000 \\ \alpha_1 = 69 - 60 = 9 \\ \alpha_2 = 69 - 16 = 53 \end{cases}$$

$$\Rightarrow M_0 = 10000 + \left[\frac{1000 \times 9}{9 + 53} \right] = 10145,16 \text{ dinars.}$$

• La médiane M_c :

| Salaire | n_i | $N(x_i) \nearrow$ |
|----------------|-------|-------------------|
| [8000; 9000[| 42 | 42 |
| [9000; 10000[| 60 | 102 |
| [10000; 11000[| 69 | 171 |
| [11000; 12000[| 16 | 187 |
| [12000; 13000[| 13 | 200 |
| Total | 200 | |

----- $N/2=200/2=100$

La classe médiane est donc [9000; 10000[.

$$M_c = e_i + \left[a_i \left(\frac{\frac{N}{2} - N(e_i) \nearrow}{n_i} \right) \right]$$

$$\begin{cases} e_i = 9000 \\ a_i = 10000 - 9000 = 1000 \\ N(e_i) \nearrow = N(9000) \nearrow = 42 \\ n_i = 60. \end{cases}$$

$$\Rightarrow M_c = 9000 + \left[1000 \left(\frac{100 - 42}{60} \right) \right] = 9966,66 \text{ dinars.}$$

3. Notons par "n" l'effectif et "c" le centre de la nouvelle classe ajoutée [13000, 14000[.

Dans ce cas la nouvelle moyenne est donnée par

$$11062,5 = \frac{1}{200 + n} \left(\underbrace{\sum_i n_i c_i}_{1998000} + n.c \right)$$

De plus

$$\left\{ \begin{array}{l} 9990 = \frac{1}{200} \sum_i n_i c_i \implies \sum_i n_i c_i = 200 \times 9990 \\ \text{et le centre de la classe } [13000, 14000[, c = \frac{1400 + 13000}{2} = 13500. \\ \implies 11062,5 = \frac{1}{200 + n} ((200 \times 9990) + (n \times 13500)) \\ \implies n = \frac{(11062,5 \times 200) - 1998000}{13500 - 11062,5} = \frac{214500}{2437,5} = 88. \end{array} \right.$$

Donc, l'effectif de la nouvelle classe ajoutée est 88, qui devient dans ce cas l'effectif le plus élevé. Comme ça, la classe modale devient $[13000, 14000[$ et le mode sera donné par,

$$M_0 = e_i + \left[\frac{a_i \alpha_1}{\alpha_1 + \alpha_2} \right], \quad \text{avec } \begin{cases} e_i = 13000 \\ a_i = 1000 \\ \alpha_1 = 88 - 13 = 75 \\ \alpha_2 = 88 \end{cases}$$

$$\implies M_0 = 13000 + \left[\frac{1000 \times 75}{75 + 88} \right] = 13460,12 \text{ dinars.}$$

Corrigé de l'exercice 7.

$$\begin{cases} x_i : \text{Le chiffre d'affaires en millions de dollars} \\ n_i : \text{Les PME (entreprises)} \end{cases}$$

1. Quel est le chiffre d'affaires qui correspond au plus grand nombre de ces entreprises ?

C'est la valeurs de variable x_i (chiffre d'affaires) qui correspond à l'effectif n_i (le nombre des entreprises) le plus élevé.

C'est donc la valeur du mode $M_0 = 4$ millions de dollars.

2. le chiffre d'affaires le plus grand = ?

$$\text{L'étendue } E = (\text{le chiffre d'affaires le plus grand}) - (\text{le plus petit})$$

$$\implies \text{L'étendue } E = (\text{le chiffre d'affaires le plus grand}) - (\text{le minimum})$$

$$\implies \text{le chiffre d'affaires le plus grand} = E + \text{le minimum} = 6 + 2 = 8 \text{ millions de dollars.}$$

3. • $Q_2 = ?$

On sait que $Q_2 =$ La médiane $M_e = 4.6$ millions de dollars.

4. • $Q_3 = ?$

Le coefficient de variation de l'écart interquartile est donné par : $\alpha = \frac{Q_3 - Q_1}{E}$
 $\Rightarrow Q_3 = [(\alpha \times E) + Q_1] = [(0.5 \times 6) + 4] = 3 + 4 = 7$ millions de dollars.

5. L'écart-type $\sigma(x) = \sqrt{V(x)}$

\Rightarrow La variance $V(x) = \sigma^2(x) = (3 \text{ millions de dollars})^2$

Corrigé de l'exercice 8.

| Nombre d'unités vendues | n_i | $N(x_i) \nearrow$ | |
|-------------------------|-------|-------------------|----------------|
| 30 | 11 | 0 | |
| 40 | 11 | 11 | N/4=40/4=10 |
| 55 | 2 | 22 | N/2=40/2=20 |
| 70 | 14 | 24 | |
| 90 | 2 | 38 | |
| Total | 40 | 40 | 3N/4=3.40/4=30 |

1. Il s'agit d'une variable discrète (nombre d'unités vendues).

2. • Le mode $M_0 = 70$.

• Le nombre moyen d'unités vendues :

$$\bar{X} = \frac{1}{N} \sum_i n_i x_i = \frac{1}{40} (11.30 + 11.40 + 2.55 + 14.70 + 2.90) = \frac{2040}{40} = 51.$$

• La médiane $M_e = 40$.

3. Les quartiles : $Q_1 = 30$ et $Q_3 = 70$.

4. On a,

$$\bar{X} = \frac{1}{40} \sum_i n_i x_i = 51 \Rightarrow \sum_i n_i x_i = 40 \times 51 = 2040,$$

$$\text{et } \bar{X}_{\text{nouvelle}} = n = \frac{1}{40+n} \left(\left(\sum_i n_i x_i \right) + n.n \right)$$

$$\Rightarrow n = \frac{1}{40+n} (2040 + n^2)$$

$$\Rightarrow n^2 + 40n = 2040 + n^2 \Rightarrow n = 51.$$

L'effectif le plus élevé est dans ce cas 51, ainsi le nouveau mode $M_0 = 51$.

Corrigé de l'exercice 9

Indication :

Les classes sont d'amplitudes inégales. Pour le traçage de l'histogramme et le calcul du mode, il va falloir corriger les effectifs n_i .

| | | | | | | |
|----------------------------|--------------|--------------|--------------|-------------|--------------|--------------|
| Montant en dinars | [1000, 1500[| [1500, 3000[| [3000, 3500[| [3500,4500[| [4500, 5000[| [5000, 6000[|
| Effectif n | 500 | 3000 | 1500 | 2000 | 1000 | 1000 |
| Effectif corrigé n_i/a_i | 500/500=1 | 3000/1500=2 | 1500/500=3 | 2000/1000=2 | 1000/500=2 | 1000/1000=1 |

Ainsi, la classe modale est [3000, 3500[et le mode se calcule classiquement par sa formule connue.

Pour le calcul de la médiane et la représentation par courbe cumulative, on doit utiliser les effectifs m et non pas les effectifs corrigés.

Les effectifs corrigés ne sont utilisés que pour le traçage de l'histogramme, le polygone et le calcul du mode.

Corrigé de l'exercice 10

Indication :

Il est clair que les deux moyennes (les salaires moyens) sont différentes. Dans ce cas il est impossible de comparer directement les deux écarts-types. Ainsi, afin de bien interpréter la dispersion, il va falloir calculer le coefficient de variation (CV) pour chaque cas.

$$CV_A = \frac{\sigma_A(X)}{\bar{X}} = \frac{2000}{22000} = 0,09 = 9\%$$

$$CV_B = \frac{\sigma_B(X)}{\bar{X}} = \frac{6000}{28000} = 0,21 = 21\%$$

Ainsi, les employés de l'entreprise B sont en moyenne mieux payés que ceux de A . Mais d'un autre côté, leurs salaires restent beaucoup plus dispersés autour de la moyenne (une différence assez importante entre ces salaires en les comparant à ceux de l'entreprise A).

Corrigé de l'exercice 11 :

Le prix de l'essence a augmenté de 20 % l'an dernier et de 10 % cette année. Quelle est la hausse moyenne de l'essence ?

Soit p le prix de l'essence. Au bout de la première année, le prix de l'essence est de

$1,20 \times p$. Au bout de la deuxième année, son prix est de $1,10 \times 1,20 \times p$ soit $1,32 \times p$.

Appelons à présent m le pourcentage moyen d'augmentation par an. Au bout de la première année, le prix de l'essence est de $(1 + m)p$. Au bout de la deuxième année, son prix est de

$(1 + m)^2 = 1,32$. Nous avons donc $(1 + m)^2 = 1,32$ soit $1 + m = \sqrt{1,32} \approx 1,1489$. Le pourcentage d'augmentation est donc environ de 14,89 %.

Cette moyenne est la moyenne géométrique des augmentations annuelles.

Corrigé de l'exercice 12 :

Une société a vu son bénéfice augmenter ces trois dernières années : de 10 % la première année, de 21 % la deuxième et de 2% la troisième. Quelle est son augmentation annuelle moyenne ?

Soit C son chiffre d'affaire initial. A l'issue de la première année, son chiffre d'affaire est $C \times 1,1$. A l'issue de la seconde année, son chiffre d'affaire est $C \times 1,1 \times 1,21$. A l'issue de la troisième année, son chiffre d'affaire est $C \times 1,1 \times 1,21 \times 1,02$.

Soit m le pourcentage moyen d'augmentation annuelle. A l'issue de la première année, son chiffre d'affaire serait $C \times (1+m)$. A l'issue de la seconde année, son chiffre d'affaire est $C \times (1+m) \times (1+m)$. A l'issue de la troisième année, son chiffre d'affaire est $C \times (1 + m) \times (1 + m) \times (1 + m)$.

On veut donc que $(1 + m)^3 = 1,1 \times 1,21 \times 1,02$ ie $1 + m \approx 1,107$. Son augmentation annuelle moyenne est de 10,7 %.

Corrigé de l'exercice 13 :

Si l'inflation d'un pays est de 5% la première année et de 15% la suivante, calculer l'augmentation moyenne des prix.

L'augmentation moyenne des prix se calcule grâce à la moyenne géométrique des coefficients multiplicateurs 1,05 et 1,15 soit une augmentation moyenne de $\sqrt{1,05 \times 1,15} \approx 1,0988$ et donc un pourcentage moyen d'augmentation annuelle de 9,88%.

Corrigé de l'exercice 14 :

Soit d la distance entre la ville A et la ville B, t le temps mis à l'aller et T le temps mis au retour. Alors $d = 20 \times t$ soit $t = d/20$ pour l'aller et $d = 26 \times T$ soit $T = d/26$ pour le retour. Durant l'aller-retour : la distance parcourue est égale à $2d$; la durée du parcours est égale à $t + T$; on note V est la vitesse moyenne parcourue durant l'ensemble du trajet. On a alors : $2d = V \times (t+T) = V (d/20+d/26)$ Soit $2 = V (1/20+1/26)$ en divisant chaque membre de l'équation par d On en déduit que $V = 2/(1/20+1/26) \approx 22,6$ km/h. Ce résultat est la moyenne harmonique des vitesses.

Corrigé de l'exercice 15 :

En 60mn, l'ouvrier A a fait 6 pièces, l'ouvrier B en a fait 4 et l'ouvrier C en a fait 3 donc au total 13 pièces ont été fabriquées en 180 minutes, soit une moyenne de 13,85 mn par pièce.

On pouvait trouver directement ce résultat en utilisant la moyenne harmonique :

$$m = \frac{3}{\frac{1}{10} + \frac{1}{15} + \frac{1}{20}} = \frac{180}{13} \approx 13,85$$

Exercices proposées

Exercice 01 :

Dans une petite localité, on a relevé de nombre de pièces par appartement.

Directive: le «nombre de pièces par appartement» est à considérer comme une variable statistique discrète à valeurs entières. (A l'interprétation, il faudra préciser que les «demi pièces» ne sont pas comptabilisées).

| Modalité | Effectif |
|----------|----------|
| 1 | 48 |
| 2 | 72 |
| 3 | 96 |
| 4 | 64 |
| 5 | 39 |
| 6 | 25 |
| 7 | 3 |
| Total | 347 |

- 1- Déterminer la population statistique.
- 2- Dessiner le diagramme en bâtons.
- 3- Donner le mode et interpréter sa valeur.
- 4- Calculer la moyenne arithmétique, la médiane.
- 5- Calculer la variance, l'écart-type et l'étendue.

Exercice n°02 :

Dans le tableau ci-dessous, les valeurs d'une variable X ont été groupées par classes de valeurs d'amplitudes égales.

| Classes | n_i | N_i |
|----------|-------|-------|
| [0, 5[| 2 | 2 |
| [5, 10[| 7 | 9 |
| [10, 15[| 18 | 27 |
| [15, 20[| 3 | 30 |

1. Déterminer par le calcul la valeur du mode et de la médiane.
2. Déterminer les quartiles Q_1 et Q_3 . Commenter ces résultats.
3. Déterminer la moyenne arithmétique.
4. Déterminer l'écart-type.

Exercice n°03 :

| Classes (nombre de voitures louées) | [0, 10[| [10,20[| [20,30[| [30,40[| [40,50[| [50,60[| [60,70[| Total |
|-------------------------------------|---------|---------|---------|---------|---------|---------|---------|-------|
| Effectifs n_i | 2 | 19 | 28 | 54 | 31 | 21 | 11 | 166 |

1. La variable étudiée « Nombre de voitures louées » est-elle discrète ou continue?
2. Calculer les centres de classes, les fréquences et les fréquences cumulées.
3. Quel type de graphique peut-on utiliser pour représenter la série étudiée?
4. Calculer le mode.
5. Quel est le nombre moyen de véhicules loués par jour ?
6. Calculer la médiane.
7. Déterminer la valeur de l'écart-type de l'échantillon considéré.

Exercice n°04 :

D'un échantillon d'étudiants de sexe masculin, on a mesuré la masse de chacun. Les masses ont été arrondies à l'entier. Les données ont été groupées en 7 classes :

| Masses en kg | 45-54 | 55-59 | 60-64 | 65-69 | 70-74 | 75-79 | 80-89 |
|--------------------|-------|-------|-------|-------|-------|-------|-------|
| Nombre d'étudiants | 5 | 14 | 33 | 47 | 26 | 13 | 2 |

1. Faire une représentation par histogramme.
2. Déterminer par le calcul la valeur du mode.
3. Trouver la valeur de la médiane et interpréter.
4. Tracer la courbe des effectifs cumulés croissants.

Exercice n°05 :

Après avoir constaté que plusieurs chèques n'ont pas été comptabilisés. Le comptable a donc décidé de refaire les calculs. Il obtient le tableau ci-après,

| | | | | | | |
|------------------------------|----------------|----------------|-------------|-------------|----------------|-------------|
| Montant des chèques x_i en | $[1550; 1590[$ | $[1590; 1630[$ | $[1630; y[$ | $[y; 1670[$ | $[1670; 1710[$ | $[1710; x[$ |
| Nombre de chèques n_i | 120 | 40 | $2y-3260$ | $10020-6y$ | 40 | 80 |

1. Calculer la borne manquante x sachant que l'étendue de cette série est égale à 240 dinars.
2. Calculer la borne manquante x sachant que le mode est égal à 1665 dinars.

Exercice n°05 :

On a les données de 200 employés répartis selon le salaire annuel en milliers de dinars.

| | | | | | |
|--------------------------|-----------|-----------|-----------|-----------|-----------|
| salaire annuel 10^3 DA | $[10,20[$ | $[20,30[$ | $[30,50[$ | $[50,60[$ | $[60,90[$ |
| effectif n_i | 20 | 60 | 50 | 40 | 30 |

1. Identifier la variable statistique X et préciser sa nature.
2. Dessiner l'histogramme.
3. Calculer le salaire moyen et la médiane.
4. Calculer un coefficient qui permet de conclure sur la forme de la distribution de X .

Conférence n°10

Distribution à deux dimensions

- Présentation en tableau
- Distribution marginal et indépendance statistique

Indépendance statistique

Paramètres d'une distribution à deux dimensions

VIII. Distribution à deux dimensions

Dans beaucoup de recherches statistiques, on ne s'intéresse pas qu'à un seul caractère mais à plusieurs en même temps. Quand on étudie deux caractères X et Y sur une population donnée, c'est en général parce qu'on cherche à savoir s'il existe un lien entre eux et qu'elle est l'intensité du lien.

L'objectif de l'analyse bi-variée est d'étudier les éventuelles relations entre deux variables statistiques.

Définition:

Soit deux caractères X et Y définis sur une même population d'effectif total (X et Y) peuvent être tous les deux qualitatifs, les deux quantitatifs ou l'un qualitatif et l'autre quantitatif).

Tableau 5.1 : Méthodes d'analyse selon la nature du variable

| | | Variable X explicative | |
|------------------------|-------------|---|-------------------------------------|
| | | Quantitatif | Qualitatif |
| Variable Y à expliquer | Quantitatif | Régression simple Corrélation simple | Analyse de la variance à un facteur |
| | Qualitatif | Régression logistique | Test du khi-deux d'indépendance |

Nous entamerons les variables quantitatives dans notre cours.

La distribution statistique à deux dimensions relatives au couple (X,Y) est définie par la donnée :

Des p valeurs possibles $n_1 \dots\dots\dots n_p$ de X.

Des q valeurs possibles $y_1 \dots\dots\dots y_p$ de Y.

Des pxq effectifs correspondants aux observations (X= n_i et Y= y_i) notée n_{ij}

VIII.1. Présentation en tableau

Tableau 5.2 : Tableau de données ponctuelles

| Observation n° | Valeur de X | Valeur de Y |
|----------------|-------------|-------------|
| 1 | x_1 | y_1 |
| ... | ... | ... |
| i | x_i | y_i |
| ... | ... | ... |
| n | x_n | y_n |

Tableau 5.3 : Tableau à double entrée (ou de contingence)

| Valeur de Y | Y_1 | _____ | Y_j | _____ | Y_q |
|-------------|----------|-------|----------|-------|----------|
| Valeur de X | | | | | |
| N_1 | n_{11} | ----- | n_{1j} | ---- | n_{1q} |
| ---- | ---- | ----- | ----- | ----- | ----- |
| n_i | n_{i1} | ----- | n_{ij} | ----- | n_{iq} |
| ----- | ----- | ----- | ----- | ----- | ----- |
| n_p | n_{p1} | ----- | n_{pj} | ---- | n_{pq} |

VIII.2. Distribution marginal et indépendance statistique

De la distribution statique du couple (X,Y) on peut déduire les distribution de X seul et y seul

En indiquant par un point une sommation effectuée suivant l'indice i

$n_{i.}$: le total des effectifs de la ligne i :

$$n_{i.} = n_{i1} + n_{i2} + \dots + n_{iq} = \sum_{j=1}^q n_{ij}$$

$n_{.j}$: total des effectifs de la ligne j

$$n_{.j} = n_{j1} + n_{j2} + \dots + n_{jp} = \sum_{i=1}^p n_{ij}$$

$n_{..}$: effectif total

$$n_{..} = \sum_{i=1}^p n_{i.} = \sum_{j=1}^q n_{.j} = \sum_{i=1}^p \sum_{j=1}^q n_{ij} = n$$

$(x_i, n_{i.})_{i=1, \dots, p}$ est la distribution marginal de X

$(x_i, n_{.j})_{j=1, \dots, q}$ est la distribution marginal de Y

VIII.3. Indépendance statistique

Deux caractères X et Y sont indépendants ssi :

$$\forall_i = (1, \dots, p), \forall_j = (1, \dots, q)$$

$$f_{ij} = f_i \cdot f_j \text{ ou } n_{ij} = \frac{n_i \cdot n_j}{n}$$

$$\left| \begin{aligned} f_{ij} &= \frac{n_{ij}}{n} \\ f_i &= \frac{n_i}{n} = \sum_{j=1}^q f_{ij} \\ f_{.j} &= \frac{n_{.j}}{n} = \sum_{i=1}^p f_{ij} \\ \sum_{i=1}^p \sum_{j=1}^q f_{ij} &= \sum_{i=1}^p f_i = \sum_{j=1}^q f_{.j} = 1 \end{aligned} \right.$$

VIII.4. Paramètres d'une distribution à deux dimensions :

X et Y sont des caractères quantitatifs

VIII.4.1. Moyennes marginales

$$\bar{X} = \frac{1}{n_{..}} \sum_{i=1}^p n_i \cdot x_i \text{ et } \bar{Y} = \frac{1}{n_{..}} \sum_{j=1}^q n_{.j} \cdot y_j$$

VIII.4.2. Variance marginales

$$V(X) = \delta^2(X) = \frac{1}{n_{..}} \sum_{i=1}^p n_i \cdot (x_i - \bar{X})^2 = \frac{1}{n_{..}} \sum_{i=1}^p n_i \cdot x_i^2 - \bar{X}^2$$

$$V(Y) = \delta^2(Y) = \frac{1}{n_{..}} \sum_{j=1}^q n_{.j} (y_j - \bar{Y})^2 = \frac{1}{n_{..}} \sum_{j=1}^q n_{.j} \cdot y_j^2 - \bar{Y}^2$$

VIII.4.3. Covariance

$$\text{Cov}(x, y) = \frac{1}{n_{..}} \sum_{i=1}^p \sum_{j=1}^q n_{ij} \cdot (x_i - \bar{X}) \cdot (y_j - \bar{Y})$$

$$= \frac{1}{n_{..}} \sum_{i=1}^p \sum_{j=1}^q n_{ij} \cdot x_i \cdot y_j - \bar{X} \bar{Y}$$

Tableau 5.4 : Tableau à double entrée (ou de contingence) de X et Y

| x \ y | 15 | 25 | 35 | 45 | n _i |
|-----------------|-----|-----|-----|-----|-----------------------|
| -1 | 40 | 168 | 320 | 68 | 596 |
| 0 | 84 | 28 | 48 | 16 | 176 |
| 1 | 128 | 32 | 24 | 44 | 228 |
| n _{.j} | 252 | 228 | 392 | 128 | n _{..} =1000 |

Tableau 5.5 : Distribution marginale de X

| x_i | -1 | 0 | 1 | Total Σ | $\frac{\Sigma}{n}$ |
|-------------------|------|-----|-----|-------------------|--------------------|
| n_i | 596 | 176 | 228 | 1000 | 1 |
| $n_i \cdot x_i$ | -596 | 0 | 228 | -368 | -0,368 |
| $n_i \cdot x_i^2$ | 596 | 0 | 228 | 824 | 0.824 |

D' où $\bar{X} = -0.368$ et $V(X) = 0,884 - (-0,368)^2 = 0,689$

Tableau 5.6 : Distribution marginale de Y

| y_j | 15 | 25 | 35 | 45 | Total Σ | Σ / n |
|-------------------|-------|--------|--------|--------|----------------|--------------|
| n_j | 252 | 228 | 392 | 128 | 1000 | 1 |
| $n_j \cdot y_j$ | 3780 | 5700 | 13720 | 5760 | 28960 | 28,96 |
| $n_j \cdot y_j^2$ | 56700 | 142500 | 482000 | 259200 | 938600 | 938,6 |

Dou $\bar{Y} = 28,96$ et $V(X) = 938,6 - (28,96)^2 = 99,9$

Conférence n°11

**Corrélation et
régression
linéaire**

**La théorie de
la corrélation**

**Régression
linéaire simple**

- Détermination de la droite de régression
- Détermination de la corrélation

IX. Corrélation et régression linéaire

IX.1. La théorie de la corrélation¹²

Lorsque deux phénomènes ont une évolution commune, nous disons qu'ils sont « corrélés ».

La corrélation simple mesure le degré de liaison existant entre ces deux phénomènes

représentés par des variables. Si nous cherchons une relation entre trois variables ou plus,

nous ferons appel alors à la notion de corrélation multiple. Nous pouvons distinguer la

corrélation linéaire, lorsque tous les points du couple de valeurs (x , y) des deux variables

semblent alignés sur une droite, de la corrélation non linéaire lorsque le couple de valeurs se

trouve sur une même courbe d'allure quelconque. Deux variables peuvent être :

- ✓ en corrélation positive ; on constate alors une augmentation (ou diminution, ou constance) simultanée des valeurs des deux variables ;
- ✓ en corrélation négative, lorsque les valeurs de l'une augmentent, les valeurs de l'autre diminuent ;
- ✓ non corrélées, il n'y a aucune relation entre les variations des valeurs de l'une des variables et les valeurs de l'autre.

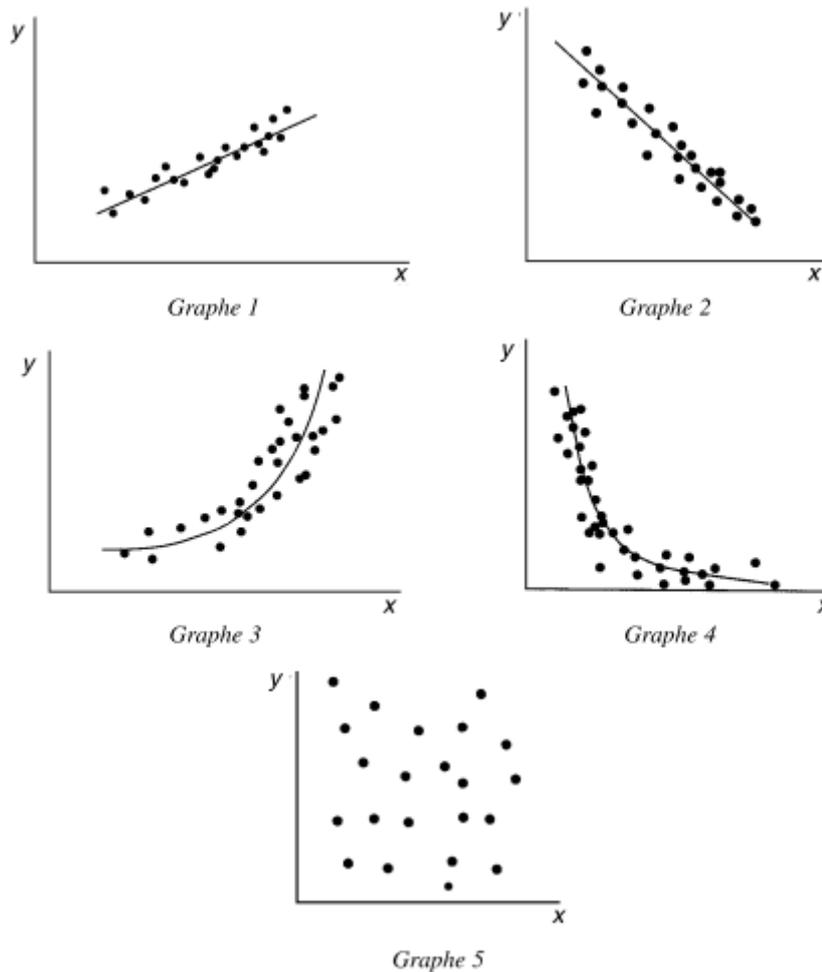
Le tableau 5.7, en croisant les critères de linéarité et de corrélation, renvoie à une représentation graphique.

Tableau 5.7 : les critères de linéarité et de corrélation

| | Corrélation positive | corrélation négative | Absence de corrélation |
|-----------------------|----------------------|----------------------|------------------------|
| Relation linéaire | Graphe 1 | Graphe 2 | Graphe 5 |
| Relation non linéaire | Graphe 3 | Graphe 4 | Graphe 5 |

¹² Bourbonnais, R (2015), « Econométrie, Cours et exercices corrigés », 9^{ème} édition, © Dunod, Paris, pp17-19

Figure 6.1 : Linéarité et corrélation



- **Mesure et limite du coefficient de corrélation**

Le coefficient de corrélation linéaire

La représentation graphique ne donne qu'une « impression » de la corrélation entre deux variables sans donner une idée précise de l'intensité de la liaison, c'est pourquoi nous calculons une statistique appelée coefficient de corrélation linéaire simple, noté $r_{x,y}$. Il est

$$\text{égal à : } r_{x,y} = \frac{\text{cov}(xy)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{Y})^2}}$$

avec :

$\text{Cov}(x,y)$ = covariance entre x et y ;

σ_x et σ_y = écart type de x et écart type de y ;

n = nombre d'observations.

On peut démontrer que, par construction ce coefficient reste compris entre - 1 et 1 :

- proche de 1, les variables sont corrélées positivement ;
- proche de - 1, les variables sont corrélées négativement ;
- proche de 0, les variables ne sont pas corrélées.

Dans la pratique, ce coefficient est rarement très proche de l'une de ces trois bornes et il est donc difficile de proposer une interprétation fiable à la simple lecture de ce coefficient. Ceci est surtout vrai en économie où les variables sont toutes plus au moins liées entre elles. De plus, il n'est calculé qu'à partir d'un échantillon d'observations et non pas sur l'ensemble des valeurs.

Nous allons maintenant aborder, la statistique descriptive en retenant les variables les liaisons pouvant exister entre elles.

Pour saisir l'objet de l'étude de la régression et de la corrélation, nous sommes amené a poser un certain nombre de questions jugés nécessaires.

- Existe-t-il une relation ou une dépendance entre les variables statistiques
- Si elle existe est elle linéaire ou non ?
- Si une relation existe comment peut-on la traduire par une équation mathématique ?

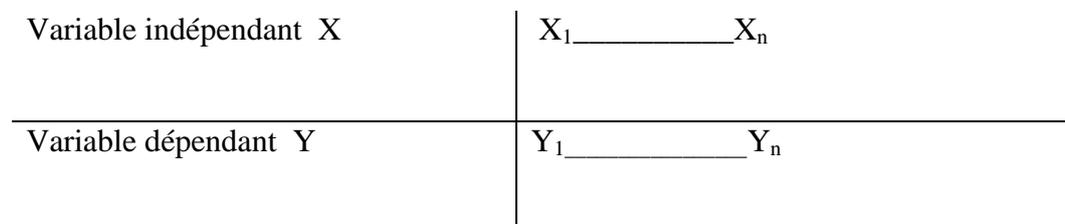
Pour répondre à tous ces questions, nous ferons appel à l'analyse de la régression.

IX.2.Régression linéaire simple :

Définition : nous appelons reg linéaire l'ajustement d'une droite au nuage statistique d'une série de couple de données.

Cette régression linéaire va nous permettre de résumer d'interpréter et de prévoir les fluctuations d'un caractère dit dépendant en fonction d'un autre caractère dit indépendant.

IX.2.1.Détermination de la droite de régression

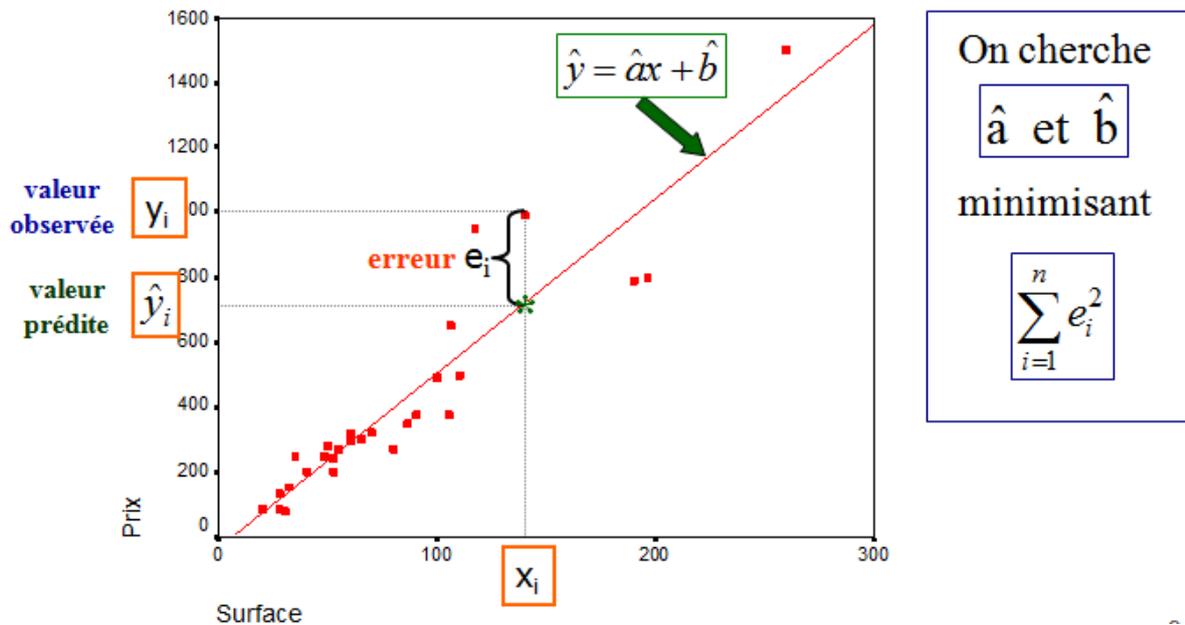


Le problème consiste maintenant a trouver une droite qui trait avec le plus de fidélité, la relation ou le lien entre X et Y cette droite est notée :

$$\hat{Y} = a + bx \text{ (équations linéaire simple)}$$

La technique utilisée pour déterminer la droite \hat{Y} est connue sous le nom (moindres carrés) le dit, son principe est de rendre minimale la somme du carré de la distance verticale (e_i) qui sépare chaque point (X_i, Y_i) de la droite elle-même

Figure 6.2 : La détermination de la droite \hat{Y}



Ce qui conduit à mini G , en calculant les dérivées partielles de l'expression par rapport à a et b .

$$G = \sum_{i=1}^n e_i^2 = e_1^2 + e_2^2 + \dots + e_n^2 \quad \text{avec} \quad e_i = y_i - \hat{y}_i = y_i - (a + bx)$$

La droite cherchée est encore celle qui minimise

$$G = \sum_{i=1}^n (Y_i - a - bX_i)^2$$

$\frac{\delta G}{\delta a} = 0 = -2 \sum (y_i - a - bX_i) = 0 \Rightarrow a = \bar{y} - b\bar{x}$ (L'ordonnée à l'origine de l'équation de la régression empirique (estimée):

$$\frac{\delta G}{\delta b} = 0 = -2 \sum x_i (y_i - a - bX_i) = 0 \Rightarrow b = \frac{\sum x_i y_i - \bar{X}\bar{Y}}{\sum x_i^2 - \bar{X}^2} = \frac{\text{cov}(X, Y)}{V(x)}$$
 (La pente de l'équation de la régression empirique (estimée))

IX.2.2. Détermination de la corrélation :

Après avoir souligné l'intérêt de l'ajustement d'une droite de régression à une série de paires de données (x_i, y_i) , nous abandonnons maintenant le problème de corrélation, ou le problème du degré de dépendance entre deux variables statistiques, qui se mesure par le degré de dispersion des points autour de la droite de régression (le coefficient de corrélation noté r).

Il existe trois mesures possibles pour quantifier l'intensité de la relation entre X et Y:

IX.2.2.1. Le coefficient de détermination de Y en fonction de X (R^2)

Le coefficient de détermination théorique de Y en fonction de X, noté R^2 mesure la proportion de la variation de Y qui est expliquée par la régression ou qui est expliquée par la variable X au niveau de toute la population. (Fournit une indication de la force de la liaison possible pouvant exister entre Y et X au niveau de la population. De plus, c'est un indice de la qualité de l'ajustement de la droite aux points expérimentaux.

Le coefficient de détermination indique si le modèle linéaire défini colle aux données $R^2 = 100 \cdot r^2$ représente le pourcentage de variation totale de y s'expliquant par la liaison de y par rapport X.

Exemple : $r = 0.9 \Rightarrow R^2 = 100, r^2 = 81\%$

Dans quelle mesure l'équation estimée de la régression s'ajuste-t-elle aux données?

Le coefficient de détermination permet de mesurer le degré d'adéquation

Formule pour le coefficient de détermination

$$R^2 = \text{SCreg} / \text{SCT} = (\text{Variation expliquée} / \text{variation totale})$$

$$\text{où: } \text{SCT} = \text{SCreg} + \text{SCres} = \sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

SCT = Somme des carrés totale (variation totale)

SCreg = Somme des carrés de la régression (variation expliquée)

SCres = Somme des carrés des résidus.

La régression linéaire simple se préoccupe essentiellement de **la forme de la relation** linéaire qui existe entre X et Y, alors que La corrélation essaye de mesurer **l'intensité ou la force** de la relation linéaire qui existe entre X et Y.

IX.2.2.2. Le coefficient de corrélation entre X et Y (r)

Le coefficient de corrélation peut être déterminé de la manière suivante (ou encore en prenant la racine carrée du coefficient de détermination):

$$r = \frac{\text{cov}(xy)}{\delta_x \delta_y}$$

Avec δ_x : écart type de X δ_y : écart type de y. On a toujours:

Si $r_{xy} = \pm 1$ alors il existe une relation linéaire exacte entre X et Y

Si $r_{xy} = 0$ alors soit que X et Y sont indépendantes, soit qu'il y a une dépendance non linéaire entre les deux variables

Si $r_{xy} \neq \pm 1$ ou $r_{xy} \neq 0$ alors il existe une relation linéaire plus ou moins forte entre X et Y.

Le coefficient de corrélation permet de voir s'il est facile d'approcher les données par une droite.

$$-1 \leq r \leq 1$$

$r < 0 \Rightarrow$ corrélation négative c'est-à-dire $x \uparrow$ et $Y \downarrow$
 $r > 0 \Rightarrow$ corrélation positive c'est-à-dire $x \uparrow$ et $Y \uparrow$

IX.2.3. La covariance entre X et Y (cov(xy)) :

La covariance est légèrement différente. Si la variance permet d'étudier les variations d'une variable par rapport à elle-même, la covariance va permettre d'étudier les variations simultanées de deux variables par rapport à leur moyenne respective.

En finance, cette notion permet de mesurer le degré de liaison des fluctuations de deux titres entre eux, ou encore d'un titre avec un indice.

La covariance peut être vue comme le produit des valeurs de deux variables moins le produit des deux moyennes. Mathématiquement, la formule est la suivante :

$$\text{cov}(x,y) = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X}) \times (Y_i - \bar{Y}) = \frac{1}{N} \sum_{i=1}^N X_i Y_i - \bar{X} \bar{Y}$$

Du résultat obtenu par cette mesure on en déduit que plus la covariance est faible et plus les variables sont indépendantes et inversement plus elle est élevée et plus les variables sont liées.

Une covariance nulle correspondant à deux variables totalement indépendantes.

Tableau 6.1 : Les ventes et les dépenses publicitaires pendant 6ans

| | | | | | | |
|--|-----|-----|-----|-----|-----|-----|
| X :dépenses publicitaires en maillons de DA | 70 | 80 | 30 | 50 | 35 | 45 |
| Y :volume des ventes en million de DA | 580 | 380 | 200 | 310 | 400 | 450 |

1. Déterminer l'équation de la droite de régression linéaire de Y en fonction de X.
2. Calculer le coefficient de détermination. Commenter le résultat obtenu.

Réponse :

Figure 6.3 : la droite de régression linéaire des ventes Y en fonction des dépenses X

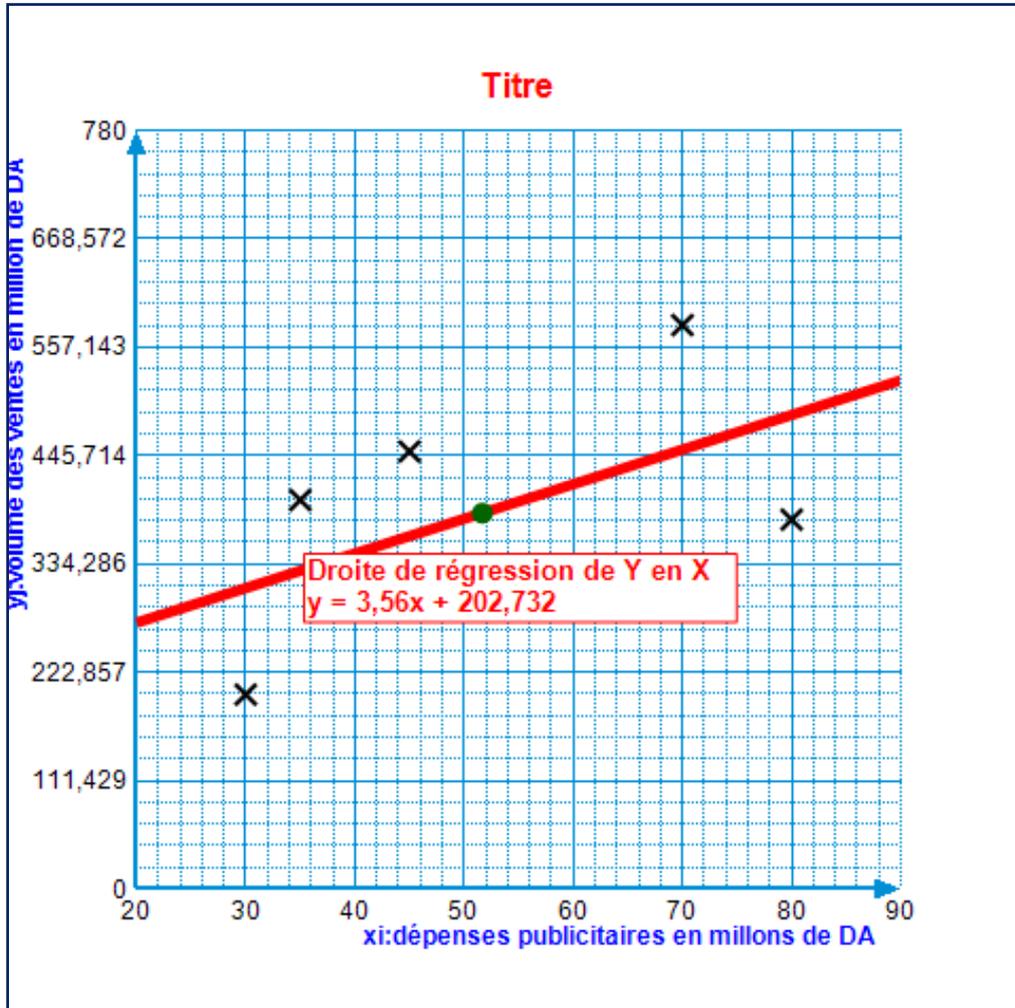


Tableau 6.2 : Calcul intermédiaire pour les coefficients de la droite d'ajustement

| | 70 | 80 | 30 | 50 | 35 | 45 | Total |
|---|--------|--------|-------|-------|--------|--------|---------------|
| x :dépenses publicitaires en millions de DA | 70 | 80 | 30 | 50 | 35 | 45 | 310 |
| y :volume des ventes en million de DA | 580 | 380 | 200 | 310 | 400 | 450 | 2320 |
| x*y | 40600 | 30400 | 6000 | 15500 | 14000 | 20250 | 126750 |
| x ² | 4900 | 6400 | 900 | 2500 | 1225 | 2025 | 17950 |
| y ² | 336400 | 144400 | 40000 | 96100 | 160000 | 202500 | 979400 |

Droite de régression de Y en fonction de X :

$D_{Y(X)} : Y = bX + a$, avec $a = \bar{Y} - b\bar{X}$ et $b = \frac{\text{cov}(X,Y)}{V(x)}$

Calculons d'abord, \bar{X} , \bar{Y} et $V(X)$, $V(Y)$ et $\text{COV}(X, Y)$.

Moyenne arithmétique de X ($\bar{X} = \frac{\sum_{i=1}^6 x_i}{N}$): 51,667 ; Variance de X $V(X) = \frac{\sum_{i=1}^6 x_i^2}{N} - \bar{X}^2$: 322,188 ;

Ecart type de x ($\sigma_x = \sqrt{V(x)}$) = 17,95 ; Moyenne arithmétique de Y ($\bar{Y} = \frac{\sum_{i=1}^6 y_i}{N}$) = 386,667 ;

Variance de Y ($V(y) = \frac{\sum_{i=1}^6 y_i^2}{N} - \bar{Y}^2$) = 13721,97 ; Ecart type de Y ($\sigma_y = \sqrt{V(y)}$) = 117,141 ;

Covariance de (X,Y) ($\text{COV}(x,y) = \frac{\sum_{i=1}^6 x_i \times y_i}{N} - \bar{X}\bar{Y}$) = 1147,077 ;

$b = \frac{\text{cov}(X,Y)}{V(x)} = 3,56$; $a = \bar{y} - b\bar{x} = 202,732$

| | |
|--|---|
| <p>$D_{Y(X)} : Y = 3,56 X + 202,732.$</p> <p>La droite de régression qui permet de déterminer le volume moyen des ventes pour un coût publicitaire donné x.</p> | <p>X : Dépenses en publicité</p> <ul style="list-style-type: none"> • Ordonnée à l'origine (volume de vente moyen sans dépenser un sous en publicité) • C'est l'augmentation du volume des ventes (Y) pour une augmentation unitaire du coût en publicité (X) |
|--|---|

Le coefficient de corrélation linéaire : ($r = \frac{\text{cov}(xy)}{\delta_x \delta_y}$): 0,546 ≈ 0.51 ; le coefficient de détermination

$R^2 = 100 \times (0,51)^2 = 26\%$

Il est clair qu'il s'agit d'une corrélation positive, donc une relation positive entre X et Y. De plus 26 % de la variation totale de Y se trouve expliquée par le lien entre X et Y.

Exercices résolues

Exercice 1. Une étude sur le rapport entre le revenu mensuel X et l'épargne Y, pour un échantillon de 20 ménages, donne la série double suivante (données exprimées en Dinars).

| X \ Y | 3000 | 4000 | 5000 | $n_{i.}$ |
|----------|------|------|------|------------|
| 20000 | 5 | 1 | 5 | |
| 40000 | 2 | 4 | 1 | 7 |
| 60000 | 0 | 1 | 1 | |
| $n_{.j}$ | | 6 | | $n_{..}=?$ |

1. Trouver les valeurs manquantes.
2. Que représentent les nombres grisés 5, 6 et 7 du tableau ?
3. Quel est le pourcentage des ménages dont le revenu mensuel est de 20000 dinars ?
4. Quel est le pourcentage des ménages ayant 5000 dinars d'épargne ?
5. Calculer les moyennes marginales et la covariance. Que peut-on conclure ?

Exercice 2. On étudie dans une population le nombre X d'arrêts de travail pour raison de santé, par an, en fonction de l'âge Y des individus à partir d'un échantillon de 800 personnes

| X \ Y | [20, 30[| [30, 40[| [40, 50[| [50, 60[|
|-------|----------|----------|----------|----------|
| 0 | 21 | 154 | 23 | 0 |
| 1 | 102 | 103 | 140 | 12 |
| 2 | 61 | 52 | 100 | 32 |

Calculer les moyennes marginales et la covariance de cet échantillon. Commenter le résultat.

Exercice 3.

Une entreprise a présenté ses dépenses publicitaires et son chiffre d'affaires pour les 6 dernières années suivant le tableau ci-après :

| | | | | | | |
|---|----|----|----|-----|-----|-----|
| Dépenses publicitaires "X" en dinars x 10^6 | 2 | 5 | 9 | 14 | 18 | 24 |
| Chiffre d'affaires "Y" en dinars x 10^6 | 11 | 15 | 50 | 120 | 140 | 210 |

L'entreprise pense qu'il y a une relation entre les dépenses publicitaires et le chiffre d'affaires. En supposant que cette relation peut être représentée par une droite d'équation $Y = aX + b$. On donne les variances $V(X) = 57.10^{12}$ et $V(Y) = 5210.10^{12}$.

- 1- Déterminer les valeurs a et b.
- 2- Calculer le coefficient de corrélation linéaire et le coefficient de détermination puis commenter le résultat obtenu.
- 3- Combien l'entreprise peut-elle espérer réaliser comme chiffre d'affaires avec des dépenses de publicité de 30 millions de dinars ?

Exercice 4. Le tableau ci-dessous donne l'observation des prix Y et des quantités disponibles X d'un produit sur le marché (Y est en dinars et X est un nombre de produits).

| | | | | | |
|---|------|------|------|------|------|
| X | 1 | 2 | 4 | 8 | 10 |
| Y | 2000 | 4000 | 5000 | 6000 | 8000 |

1. Déterminer l'équation de la droite de régression linéaire de Y en fonction de X.
2. Calculer le coefficient de détermination. Commenter le résultat obtenu.

Solution des exercices

Corrigé de l'exercice 01 :

1.

| X \ Y | 3000 | 4000 | 5000 | $n_{i.}$ | $\sum_j n_{ij} x_i y_j$ |
|-------------------------|------------------------------|------------------------------|------------------------------|----------------|----------------------------------|
| 20000 | 5 (30.10^7) | 1 (8.10^7) | 5 (50.10^7) | 5+1+5 = 11 | (88.10^7) |
| 40000 | 2 (24.10^7) | 4 (64.10^7) | 1 (20.10^7) | 2+4+1 = 7 | (108.10^7) |
| 60000 | 0 (0) | 1 (24.10^7) | 1 (30.10^7) | 0+1+1 = 2 | (54.10^7) |
| $n_{.j}$ | 5 + 2 + 0 = 7 | 1 + 4 + 1 = 6 | 5 + 1 + 1 = 7 | $n_{..} = 20$ | (250.10^7) |
| $\sum_i n_{ij} x_i y_j$ | (54.10^7) | (96.10^7) | (100.10^7) | (250.10^7) | $= \sum_i \sum_j n_{ij} x_i y_j$ |

2. '5' est le nombre des ménages ayant 5000 dinars d'épargne et un revenu mensuel de 20000 dinars.

'6' est le nombre des ménages ayant 4000 dinars d'épargne.

'7' est le nombre des ménages dont le revenu mensuel est 40000 dinars.

3. Le pourcentage des ménages dont le revenu mensuel est de 20000 dinars est $11/20 = 0.55 = 55\%$.

4. Le pourcentage des ménages ayant 5000 dinars d'épargne est $7/20 = 0.35 = 35\%$.

5. Moyenne marginale de X :

$$\bar{X} = \frac{1}{n_{..}} \sum_i n_{i.} x_i = \frac{1}{20} (11 \times 20000 + 7 \times 40000 + 2 \times 60000) = \frac{620000}{20} = 31000 \text{ dinar}$$

Moyenne marginale de Y :

$$\bar{Y} = \frac{1}{n_{..}} \sum_j n_{.j} y_j = \frac{1}{20} (7 \times 3000 + 6 \times 4000 + 7 \times 5000) = \frac{80000}{20} = 4000 \text{ dinar}$$

La covariance : $COV(X, Y) = \left[\frac{1}{n_{..}} \sum_i \sum_j n_{ij} x_i y_j \right] - \bar{X} \bar{Y}$

A partir du tableau :

$$\frac{1}{n_{..}} \sum_i \sum_j n_{ij} x_i y_j = 250 \cdot 10^7 \text{ Ainsi } COV(X, Y) = \frac{250 \cdot 10^7}{20} - (31000 \cdot 4000) = 12,5 \times 10^7 - 12,4 \times 10^7 = 1000000$$

Covariance positive, donc corrélation positive entre X et Y (si X↑ alors Y↑ et si X↓ Y↓)

Corrigé de l'exercice 3.

1. Droite de régression de Y en fonction de X :

$$D_{Y(X)} : Y = aX + b, \text{ avec } \begin{cases} a = \frac{COV(X, Y)}{V(X)} \\ b = \bar{Y} - a\bar{X} \end{cases}$$

Calculons d'abord, \bar{X} , \bar{Y} et $COV(X, Y)$.

$$\bar{X} = \frac{1}{N} \sum_i x_i = \frac{2 + 5 + 9 + 14 + 18 + 24}{6} = \frac{72}{6} = 12 \text{ millions de dinars.}$$

$$\bar{Y} = \frac{1}{N} \sum_i y_i = \frac{11 + 15 + 50 + 120 + 140 + 210}{6} = \frac{546}{6} = 91 \text{ millions de dinars.}$$

On a déjà, $V(X) = 57$ et $V(Y) = 5210$.

La covariance :

$COV(X, Y)$

$$\begin{aligned} &= \left[\frac{1}{N} \sum_i x_i y_i \right] - \bar{X} \bar{Y} \\ &= \left[\frac{1}{6} ((2 \times 11) + (5 \times 15) + (9 \times 50) + (14 \times 120) + (18 \times 140) + (24 \times 210)) \right] - (12 \times 91) \\ &= \frac{9787}{6} - 1092 = 539,16. \end{aligned}$$

De là,

$$\begin{cases} a = \frac{COV(X, Y)}{V(X)} = \frac{539,16}{57} = 9,45 \\ b = \bar{Y} - a\bar{X} = 91 - (9,45 \times 12) = 91 - 113,4 = -22,4 \end{cases}$$

et $D_{Y(X)} : Y = 9,45 X - 22,4$.

Le coefficient de corrélation linéaire :

$$r = \frac{COV(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{539,16}{\sqrt{57}\sqrt{5210}} = \frac{539,16}{544,23} = 0,99.$$

Le coefficient de détermination :

$$100.r^2 = 100 \times (0,99)^2 = 98,01 \%$$

Il est clair qu'il s'agit d'une corrélation positive, donc une relation positive entre les dépenses publicitaires X et le chiffre d'affaires Y .

De plus 98.01 % de la variation totale de Y se trouve expliquée par le lien entre X et Y , soit une forte liaison entre X et Y .

2. Du moment que X représente les dépenses publicitaires et Y le chiffre d'affaires, il suffit donc de remplacer la valeur de $X = 30$ millions de dinars dans l'équation de droite trouvée, comme ça on pourra avoir la valeur de Y . On a donc,

$$\begin{cases} X = 30 \text{ millions de dinars} \\ \text{et } Y = 9,45X - 22,4 \Rightarrow Y = (9,45 \times 30) - 22,4 = 261,1 \text{ millions de dinars.} \end{cases}$$

Corrigé de l'exercice 4.

Droite de régression de Y en fonction de X :

$$D_{Y(X)} : Y = aX + b, \text{ avec } \begin{cases} a = \frac{COV(X, Y)}{V(X)} \\ b = \bar{Y} - a\bar{X} \end{cases}$$

Calculons d'abord, \bar{X} , $V(X)$, \bar{Y} , $V(Y)$ et $COV(X, Y)$.

$$\bar{X} = \frac{1}{N} \sum_i x_i = \frac{1 + 2 + 4 + 8 + 10}{5} = 5.$$

$$\bar{Y} = \frac{1}{N} \sum_i y_i = \frac{2000 + 4000 + 5000 + 6000 + 8000}{5} = 5000.$$

$$V(X) = \left(\frac{1}{N} \sum_i x_i^2 \right) - \bar{X}^2 = \frac{1^2 + 2^2 + 4^2 + 8^2 + 10^2}{5} - 5^2 = \frac{185}{5} - 25 = 12.$$

$$V(Y) = \left(\frac{1}{N} \sum_i y_i^2 \right) - \bar{Y}^2 = \frac{2000^2 + 4000^2 + 5000^2 + 6000^2 + 8000^2}{5} - 5000^2 = 4000000.$$

La covariance :

$$\begin{aligned} COV(X, Y) &= \left[\frac{1}{N} \sum_i x_i y_i \right] - \bar{X} \bar{Y} \\ &= \left[\frac{1}{5} (2000 + 8000 + 20000 + 48000 + 80000) \right] - (5 \cdot 5000) = 6600. \end{aligned}$$

Ainsi,

$$\begin{cases} a = \frac{COV(X, Y)}{V(X)} = \frac{6600}{12} = 550 \\ b = \bar{Y} - a\bar{X} = 5000 - (550 \cdot 5) = 2250 \end{cases}$$

et $D_{Y(X)} : Y = 550X + 2250$.

Le coefficient de corrélation linéaire :

$$r = \frac{COV(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}} = \frac{6600}{\sqrt{12}\sqrt{4000000}} = 0,95.$$

Le coefficient de détermination :

$$100 \cdot r^2 = 100 \cdot (0,95)^2 = 90,25 \%$$

Il est clair qu'il s'agit d'une corrélation positive, donc une relation positive entre X et Y . De plus 90,25% de la variation totale de Y se trouve expliquée par le lien entre X et Y , soit une forte liaison entre X et Y .

Exercices proposées

Exercice 01 :

Soit la série statistique bidimensionnelle du couple (X, Y) suivante :

| | | | | | |
|-----|----|----|---|----|-----|
| X/Y | -2 | 0 | 2 | 3 | ni. |
| 2 | 3 | 4 | 0 | 6 | 13 |
| 3 | 4 | 3 | 3 | 2 | 12 |
| 4 | 2 | 3 | 3 | 2 | 10 |
| n.j | 9 | 10 | 6 | 10 | 35 |

Calculer les moyennes marginales et la covariance. Que peut-on conclure ?

Exercice 02 :

Soit la série des 12 mesures suivantes :

| | | | | | | | | | | | | |
|-------|-----|-----|---|---|---|---|---|-----|-----|----|----|-----|
| X_i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Y_i | 1.5 | 2.5 | 2 | 2 | 3 | 4 | 4 | 3.5 | 4.5 | 4 | 5 | 5.5 |

- 1) Trouver l'équation d'ajustement de Y en X (D(Y)).
- 2) Trouver l'équation d'ajustement de X en Y (D(X)).
- 3) Calculer le coefficient de corrélation linéaire entre X et Y.

Exercice 03 :

Dans une maternité, on a mesuré les poids et tailles des 10 premiers nouveaux nés d'une journée donnée. On a consigné ces données dans le tableau suivant :

| n° i du nouveau né | Taille X_i en cm | Poids Y_i en g |
|--------------------|--------------------|------------------|
| 1 | 50 | 3100 |
| 2 | 52 | 3230 |
| 3 | 47 | 2950 |
| 4 | 49 | 3050 |
| 5 | 50 | 3100 |
| 6 | 55 | 3350 |
| 7 | 54 | 3480 |
| 8 | 51 | 3250 |
| 9 | 45 | 3000 |
| 10 | 49 | 3150 |

1. Construire le nuage de points et commenter ce graphique.
2. Déterminer le coefficient de corrélation linéaire.
3. Déterminer la droite de régression de Y en X expliquant le poids d'un nouveau né en fonction de sa taille.
4. Calculer la variation qui est expliquée par la droite des moindres carrés.
5. Quelle proportion de la variation totale est expliquée par la droite des moindres carrés ?
6. Après une nouvelle naissance, on a mesuré la taille d'un nouveau né et obtenu 53cm. Quel poids peut-on s'attendre à mesurer pour ce nouveau né ?

Exercice 04 :

Nous avons 100 salariés qui sont observés à l'aide de deux variables "âge" et "salaire" et on a le tableau à double entrée des deux variables comme suit :

| Age \ Salairex10 ³ | [6,7[| [7,8[| [8,9[| [9,10[| [10,11[| ni• | fi• |
|-------------------------------|-------|-------|-------|--------|---------|-----|-----|
| [15, 25[| 07 | 02 | 01 | 02 | 0 | | |
| [25, 35[| 05 | 05 | 05 | 10 | 15 | | |
| [35, 45[| 03 | 05 | 02 | 04 | 04 | | |
| [45, 55[| 07 | 03 | 10 | 02 | 08 | | |
| n•j | | | | | | 100 | 1 |
| f•j | | | | | | 1 | |

- 1- Compléter le tableau en calculant les effectifs marginaux.
- 2- Donner le nombre des salaires ayant moins de 45 ans et donne le nombre de qui sont plus de 45 ans
- 3- Calculer les moyennes et variances marginales.
- 4- Calculer la covariance, que peut-on conclure.

Conférence n°12

**Caractéristique
de
concentration**

Valeur globales

- Valeurs globales et valeurs globales relatives
- Valeurs globales relatives cumulées croissantes

Médiale

- Détermination graphique
- Détermination algébrique

X. Caractéristique de concentration :

Pour les variables quantitatives, l'étude des caractéristiques de tendance centrale et de dispersion permet d'obtenir un résumé de l'information. Mais ce résumé ne permet pas de prendre en compte de certains aspects de la distribution. En particulier, dans le cas où la variable étudiée concerne une grandeur positive cumulable (cas fréquent en économie), il ne permet pas d'étudier les inégalités de répartition.¹³

L'étude de la concentration a pour objectif de mettre en évidence et de mesurer ces éventuelles inégalités. Elle doit permettre en particulier de compléter l'analyse de la « dispersion relative » d'une distribution.

Les domaines d'application sont nombreux en économie : concentration des salaires, concentration des revenus ou du patrimoine, de l'emploi, concentration économique d'un secteur.¹⁴

X.1.Valeur globales :

Le tableau 7.1 précise la distribution des salaires nets annuels des agents de l'état dans une période donnée.

Le tableau 7.1 : la distribution des salaires nets annuels des agents de l'état

| Salaires annuels | Effectif n_i |
|------------------|----------------|
| Moins de 50 | 9246 |
| [50 ; 70[| 111107 |
| [70 ; 80[| 129457 |
| [80 ; 100[| 476716 |
| [100 ; 120[| 543670 |
| [120 ; 150[| 666398 |
| [150 ; 180[| 301428 |
| [180 ; 300[| 294687 |
| [300 ; 500[| 42172 |
| Plus de 500 | 3871 |
| Total | 2578752 |

La variable statistique est le salaire net annuel. Les observations (valeur observée de la variable salaire) ont été groupées en classes statistiques et exprimées en milliers d'Unité Monétaire).

La population étudiée est celle des agents de l'état.

On peut s'interroger, à partir de ce tableau, sur la répartition de la masse totale des salaires entre les divers groupes des salariés. Quelle part de cette masse totale se partagent les agents de qui gagnent moins de 100 K (U M).

¹³ Pascal Chareille, Yves Pinault (1998), **Statistique descriptive**, Edition Montchrestien, Paris, pp125-128

¹⁴ On rappelle que la concentration économique d'un secteur représente le poids, dans le secteur, des plus grandes entreprises classés dans l'ordre décroissant des variables économiques étudiées.

Quelle part se partagent ceux qui sont le mieux payés ?

Ce type d'interrogation conduit à s'intéresser davantage à la structure du budget annuel consacré par l'état à la rémunération de ses agents.

L'étude de la concentration passe nécessairement par le calcul de quantités particulières.

Définitions :

Soit X la variable associée à un caractère quantitatif continu.

Soit $\{(x_i, n_i)\}_{i=1,2,\dots,k}$ la distribution statistique correspondante.

X.1.1 Valeurs globales et valeurs globales relatives

On appelle valeur globale associée au couple (x_i, n_i) le produit $n_i \cdot x_i$

On appelle valeurs globales relatives associées au couple (x_i, n_i) , et on note q_i le rapport défini par :

$$q_i = \frac{n_i \times x_i}{\sum_{i=1}^k n_i \times x_i}$$

En pratique, on exprime les valeurs globales relatives q_i en %, soit :

$$q_i(\%) = \frac{n_i \times x_i}{\sum_{i=1}^k n_i \times x_i} \times 100$$

Dans le cas des données groupées en classes, les valeurs x_i sont les centres de classes.

Si on note S la (masse totale de la variable X) (soit $S = \sum_{i=1}^k n_i \times x_i$), la valeur q_i est la fréquence associée à la valeur globale $n_i \cdot x_i$. Elle exprime la proportion que représente la valeur globale correspondant à la modalité x_i par rapport à la masse S .

X.1.2. Valeurs globales relatives cumulées croissantes :

Définitions :

On appelle -pour données groupées en classes statistiques- valeur globale relative cumulée croissante à l'extrémité de classe e_i et on note Q_i la somme des valeurs globales relatives associées à la classe $[e_{i-1}, e_i[$ et aux classes précédentes.

On a $Q_i = q_1 + q_2 + \dots + q_i = \sum_{j=1}^i q_j$

Présentation des données en tableau :

En pratique, les valeurs globales, les valeurs globales relatives et les valeurs globales relatives cumulées croissantes sont présentées en tableau.

Tableau 7.2 : les valeurs globales, les valeurs globales relatives et les valeurs globales relatives cumulées croissantes

| extrémité de classe e_i | Salaires annuel en ($\times 10^3$ U.M) | centre de classe | Effectif n_i | fréquence f_i | fréquence cumulée croissante F_i (%) | valeur globale $n_i \cdot x_i$ | valeur globale relative q_i (%) | V G R cumulée croissante Q_i (%) |
|---------------------------|---|------------------|----------------|-----------------|--|--------------------------------|-----------------------------------|------------------------------------|
| 0 | Moins de 50 | 25 | 9246 | 0,4 | 0 | 231150 | 0,1 | 0 |
| 50 | [50 ; 70[| 60 | 111107 | 4,3 | 0,4 | 6666420 | 1,9 | 0,1 |
| 70 | [70 ; 80[| 75 | 129457 | 5 | 4,7 | 9709275 | 2,8 | 2,0 |
| 80 | [80 ; 100[| 90 | 476716 | 18,5 | 9,7 | 42904440 | 12,3 | 4,8 |
| 100 | [100 ; 120[| 110 | 543670 | 21,1 | 28,2 | 59803700 | 17,1 | 17,1 |
| 120 | [120 ; 150[| 135 | 666398 | 25,8 | 49,3 | 89963730 | 25,8 | 34,2 |
| 150 | [150 ; 180[| 165 | 301428 | 11,7 | 75,1 | 49735620 | 14,2 | 60,0 |
| 180 | [180 ; 300[| 240 | 294687 | 11,4 | 86,8 | 70724880 | 20,3 | 74,2 |
| 300 | [300 ; 500[| 400 | 42172 | 1,6 | 98,2 | 16868800 | 4,8 | 94,5 |
| 500 | Plus de 500 | 650 | 3871 | 0,2 | 99,8 | 2516150 | 0,7 | 99,3 |
| 800 | Total | | 2578752 | 100 | 100 | 349124165 | 100,000 | 100 |

On a par exemple :

- $f_1 = \frac{9246}{2578752} \times 100 = 0,4$
- $n_1 \times x_1 = 9246 \times 25 = 231150$
- $q_1(\%) = \frac{n_1 \times x_1}{\sum_{i=1}^{10} n_i \times x_i} = \frac{231150}{349124165} \times 100 = 0,1$
- $F_4(\%) = f_1 + f_2 + f_3 + f_4 = 0,4 + 4,3 + 5,0 + 18,5 = 28,2$
- $n_9 \times x_9 = 42172 \times 400 = 16868800 \times K$ (UM)
- $S = \sum_{i=1}^{10} n_i \times x_i = 231150 + 6666420 + \dots + 2516150 = 349124165 \times K$ (UM)

La masse totale des salaires (S) est égale à $349124165 \times K$ (UM) cette somme que doit consacrer l'état tous les mois à la rémunération de ses agents.

- Les 28,2% d'agents de l'état qui gagnent moins de $100 \times K$ (UM) par an partagent ensemble 17,1% de la masse totale des salaires ;
- Les 42172 agents de l'état qui gagnent entre 300 et 500000 (UM) par an coutent ensemble à l'état 16868800000 (UM) ;
- Les 231150 agents de l'état (soit 0,4% des effectifs) qui gagnent moins de 50000 (UM) par an se partagent 0,1% de la masse totale des salaires.

X.2.Médiale

Définition :

On appelle médiale d'une distribution statistique, et on note Mle la valeur de la variable associée à $Q_1=50\%$

Sa détermination (comme pour la médiane) peut se faire graphiquement (en construisant par exemple une courbe des valeurs globales relatives cumulées croissantes) ou algébriquement (par interpolation linéaire).

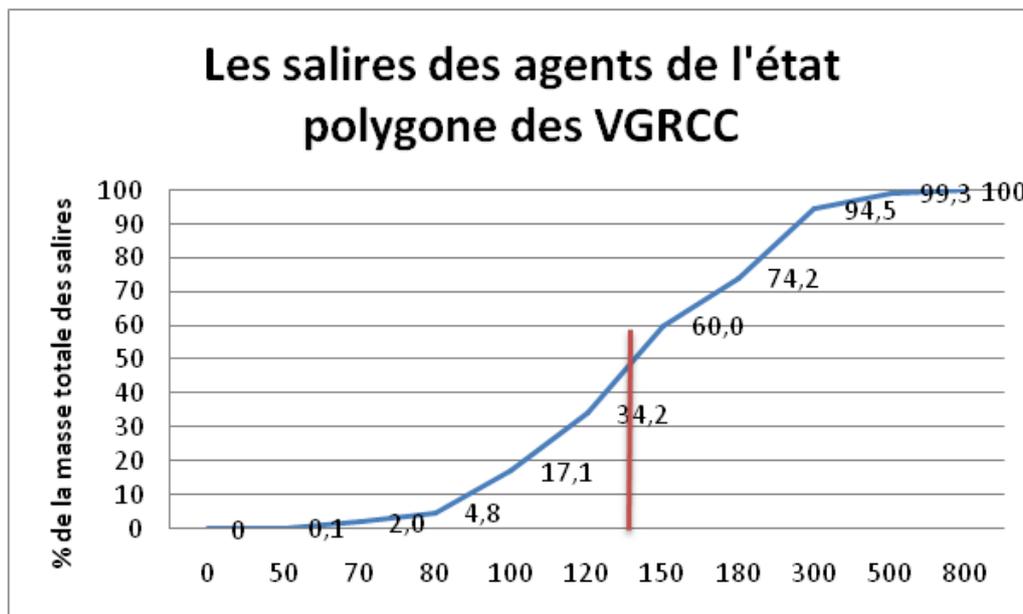
la Mle est donc la valeur du caractère qui partage la somme des $n_i \times x_i$ en 2 parties égales, 50% de cette masse ($\sum n_i \times x_i$) est $<$ à cette valeur et 50% est \geq à cette valeur.

X.2.1. Détermination graphique :

On appelle polygone des valeurs globales relatives cumulées croissantes la ligne polygonale joignant ; dans un système d'axes orthogonaux, les points d'abscisses (extrémité de classe) et d'ordonnée Q (valeur globale relative cumulée croissante).

La médiale est estimée graphiquement comme l'abscisse du point d'ordonnée 50% du polygone des valeurs globales relatives cumulées croissantes(VGRCC)¹⁵.

Figure 7.1 : L'estimation graphique de la médiale



L'estimation graphique de la médiale se fait selon le schéma proposé est de $Mle=140000$ (UM)

¹⁵ Pascal Chareille, Yves Pinault (1998), **Statistique descriptive**, Edition Montchrestien, Paris.

Cette courbe permet également d'obtenir d'autres informations sur la répartition de la masse totale des salaires, par exemple les agents de l'état dont le salaire annuel est inférieur à 250000 (UM) se partagent plus de 85% de la masse totale des salaires. Un peu moins de 15% de cette masse totale est donc consacré à la rémunération des agents dont le salaire annuel dépasse 250000 (UM) .

X.2.2. Détermination algébrique :

Les étapes conduisant au calcul du médiale sont les suivants :

1. Calcul des valeurs globales.
2. Calcul des valeurs globales relatives q_i .
3. Calcul des valeurs globales relatives cumulées croissantes Q_i .
4. Détermination de la classe qui contient la médiale (celle qui correspond à $Q_i=50\%$).
5. Calcul de la médiale.

S'il n'y a aucune extrémité de classe correspond à la valeur $Q_i=50\%$, on peut utiliser (comme pour le calcul de la médiane) le tableau suivant :

| | | | | |
|-------------------------|-----------|--------|-----|--------|
| Valeur de la variable | x | a | Mle | b |
| Val.Glob.Rel.Cum.Croiss | $Q_i(\%)$ | $Q(a)$ | 50 | $Q(b)$ |

$$\text{On a : } \frac{Q(b)-Q(a)}{b-a} = \frac{50-Q(a)}{Mle-a} \text{ soit : } Me = a + (b-a) \times \frac{50-Q(a)}{Q(b)-Q(a)}$$

Revenant à notre exemple, on a le tableau suivant :

| | | | |
|-------------------------|-------------|-----|-----------|
| Valeur de la variable | a=120 | Mle | b=150 |
| Val.Glob.Rel.Cum.Croiss | $Q(a)=34,2$ | 50 | $Q(b)=60$ |

$$\text{On a alors : } \frac{60-34,2}{150-120} = \frac{50-34,2}{Me-120} \text{ soit : } Me = 120 + (150 - 120) \times \frac{50-34,2}{60-34,2} \approx 138,7.$$

Interprétation :

La moitié de la masse totale des salaires distribués par l'état à ses agents concerne les salariés dont le traitement annuel est inférieur à 138370 (UM).

L 'écart Médiale – Médiane : ΔM

La comparaison de la valeur de la médiale et de la médiane constitue une mesure de la concentration .

La mesure de l'écart entre médiale et médiane renseigne sur le degré de concentration, plus l'écart est grand et plus la concentration est forte et il est nul en cas de parfaite égalité.

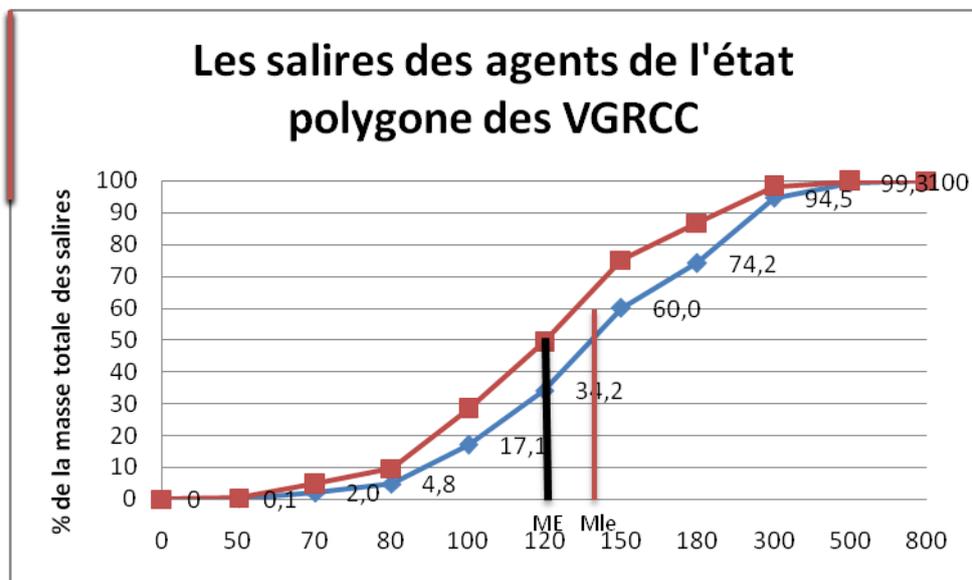
$$\Delta M = Mle - Me$$

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

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

- Si ΔM est grand par rapport à l'étendue : la concentration est forte
- Si ΔM est petit par rapport à l'étendue : la concentration est faible
- Si ΔM est nul : la concentration est nulle ($Mle = Me$) → égalité parfaite

Figure 7.2 : Les salaires des agents : courbes des fréquences cumulées croissantes et des valeurs globales relatives cumulées croissantes



Pour la distribution des salaires des agents, l'étendue est de 800 000 (UM).

La valeur de la médiale est de 138370 (UM)

| | | | |
|---------------------------------|------|----|------|
| Valeur de la variable (salaire) | 120 | Me | 150 |
| Fréquence .Cum.Croiss F_i (%) | 49,3 | 50 | 75,1 |

La valeur de la médiane est calculée par interpolation linéaire :

On a :

$$\frac{75,1-49,3}{150-120} = \frac{50-49,3}{M_{le}-120} \text{ soit : } M_e = 120 + (150 - 120) \times \frac{50-49,3}{75,1-49,3} \approx 120,81$$

La moitié des agents de l'état ont perçu un salaire annuel inférieur à 120810 (UM) ;

L'écart $\Delta M = M_{le} - M_e = 137,37 - 120,81 = 17,56$.

Comme ΔM est petit par rapport à l'étendue, la concentration est faible

ΔM représente dans cet exemple moins de 2,2% ($17,56/800 = 0,022$) de l'étendue.

Conférence n°13

La courbe de Lorentz (Courbe de Concentration)

- Construction de la courbe de Lorentz
- L'Indice de Concentration ou Coefficient de GINI

X.3. LA COURBE DE LORENTZ (COURBE DE CONCENTRATION) :

La courbe de concentration permet de représenter graphiquement la répartition des valeurs d'un caractère quantitatif, discret ou continu, entre les individus de la population. La courbe de concentration permet de comparer à une répartition strictement égalitaire, la répartition d'une série statistique donnée.

La comparaison des fréquences cumulées croissantes et des valeurs globales relatives cumulées croissantes conduit à la construction d'une courbe dite courbe de concentration (ou courbe de Lorenz).

On appelle courbe de concentration (ou courbe de Lorenz).

La courbe obtenue en associant à chaque valeur x_i de la variable X un point de coordonnées $F(x)$ en abscisse (proportion des individus de la population pour lesquels la variable est inférieure à x) et $Q(x)$ en ordonnée (proportion de la masse totale de la variable correspondante).

X.3.1 Construction de la courbe de Lorenz¹⁶ :

Sur les 2 côtés d'un carré, on porte (à la même échelle) :

- Sur les abscisses (axe OX), les fréquences cumulées croissantes en % des

individus de chaque classe soit f_i % cumulés croissantes (fréquence cumulée croissante F_i (%)) ou F_i

- Sur les ordonnées (axe OY), les fréquences cumulées croissantes en % des

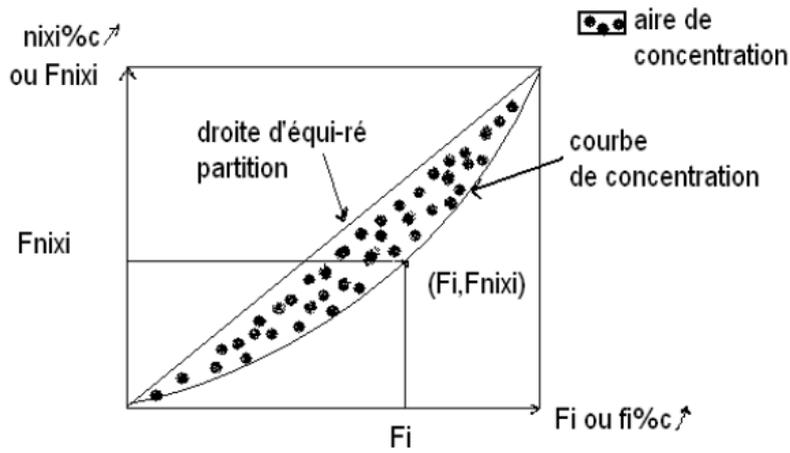
masses soit les $n_i x_i$ en % cumulés croissantes $Q(x)$ ou $F_n i x i$

Remarque : $F_n i x_i = q_i(\%) = \frac{n_i \times x_i}{\sum_{i=1}^k n_i \times x_i} \times 100$ cumulées croissantes (cum \uparrow) ou (V G R cumulée croissante Q_i (%))

- On porte dans le carré les différents points coordonnées (F_i , $F_n i x i$) ou (f_i % cum \uparrow , $n_i x_i$ en % cum \uparrow). La courbe de concentration est obtenue en joignant les points représentés.

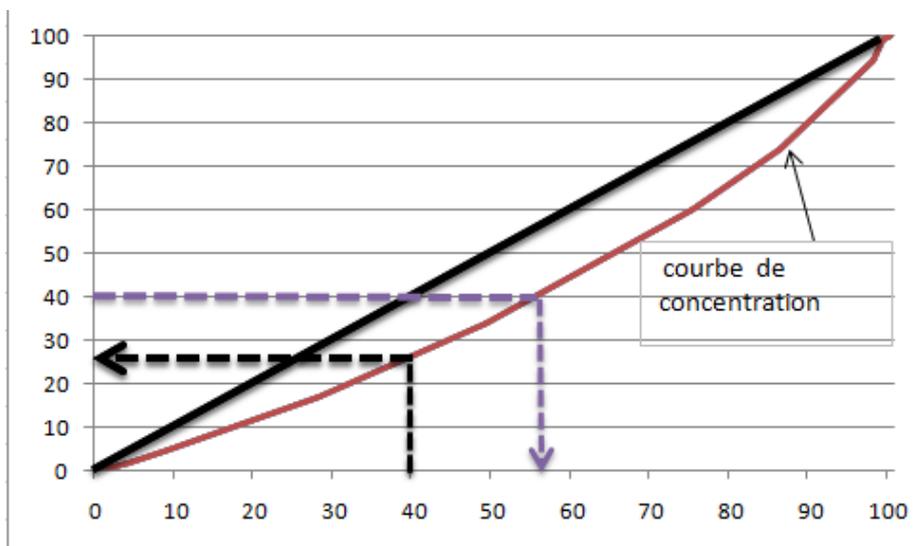
¹⁶ <https://p9.storage.canalblog.com/98/61/1177787/100938518.pdf>, consulté le 2021-08-26, pp3,6

Figure 7.3 : La courbe de concentration



| fréquence cumulée croissante F_i (%) | V G R cumulée croissante Q_i (%) |
|--|------------------------------------|
| 0 | 0 |
| 0,4 | 0,1 |
| 4,7 | 2 |
| 9,7 | 4,8 |
| 28,2 | 17,1 |
| 49,3 | 34,2 |
| 75,1 | 60 |
| 86,8 | 74,2 |
| 98,2 | 94,5 |
| 99,8 | 99,3 |
| 100 | 100 |

Figure 7.4 : Les salaires des agents de l'état : Courbe de concentration :



Par exemple :

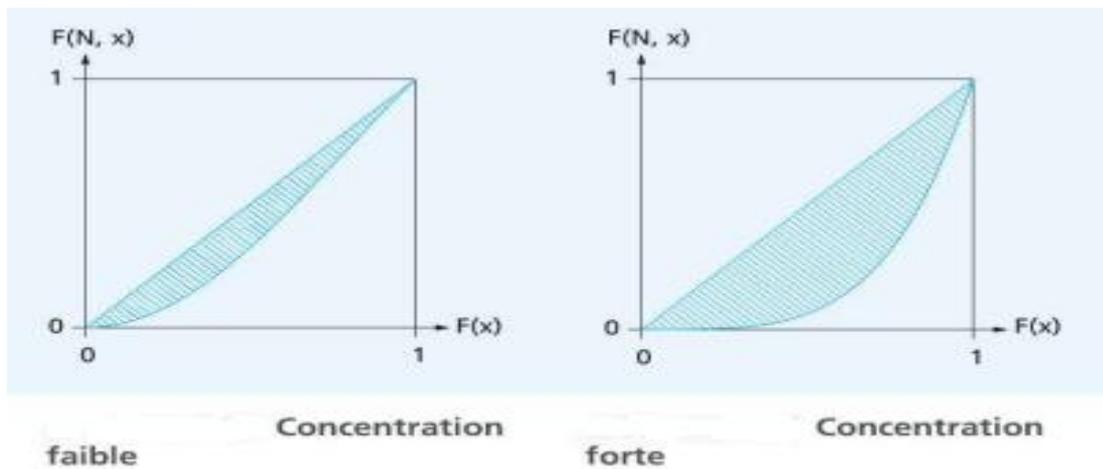
Les 40% d'agents de l'état les moins bien payés se partagent 27% de la masse salariale.

L'état consacre 40% de la masse totale salariale à la rémunération des 55% d'agents les moins bien payés.

La diagonale du carré correspond à la ligne d'équirépartition parfaite (cas théorique où tous les agents ont le même salaire : 10% des agents reçoivent 10% de la masse salariale, 20% des agents reçoivent 20% de la masse salariale ; etc.). Cette diagonale principale est la ligne de concentration nulle.

La courbe de concentration s'écarte peu de la diagonale du carré ; la concentration est ici relativement faible.

Figure 7.5 : type de concentration



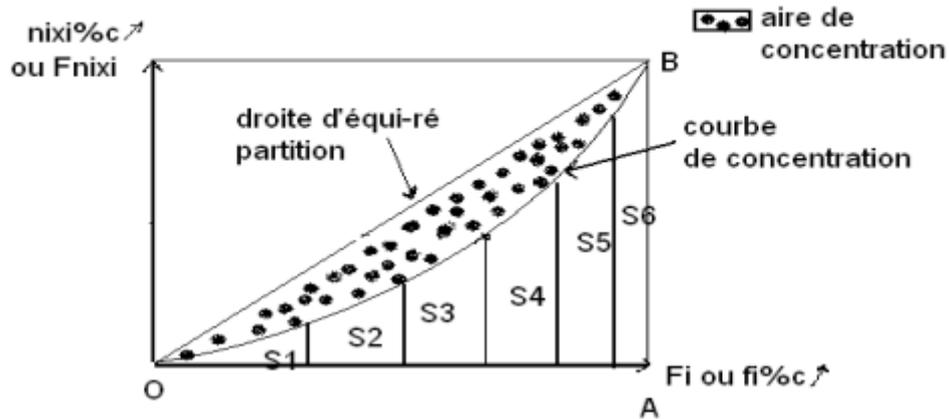
X.3.2.L'INDICE DE CONCENTRATION OU COEFFICIENT DE GINI¹⁷

Définition

L'indice de Gini est le rapport de l'aire située entre la bissectrice et la courbe de Lorentz à l'aire totale de demi-carré.

¹⁷ <https://www.fichier-pdf.fr/2019/10/31/chapitre-4-les-caracteristiques-de-concentration/preview/page/2/?>

Figure 7.6 : la courbe de Lorentz



$$I_g = \frac{\text{Aire de concentration}}{\text{Aire de triangle OAB}}$$

Calcul pratique : Méthode des trapèzes

L'aire sous la courbe est décomposée en triangle et trapèzes dont on calcule la surface ($\sum S_i$).

On obtient un triangle et des trapèzes

L'aire A d'un trapèze dont les bases sont a et b et dont la hauteur est h est : $A = ((a+b) \times h) / 2$

$$\text{Aire de triangle OAB} = \frac{100 \times 100}{2} = 5000$$

$$\text{Aire de concentration} = \text{Aire de triangle OAB} - \sum S_i = 5000 - \sum S_i$$

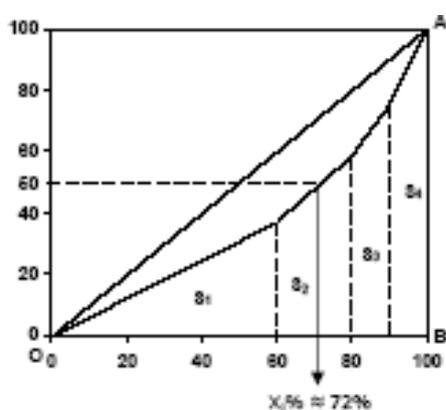
$$I_g = \frac{\text{Aire de concentration}}{\text{Aire de triangle OAB}} \Rightarrow I_g = \frac{5000 - \sum S_i}{5000}$$

On prend l'exemple Suivant :

Tableau 7.3 : Les salariés d'une entreprise sont répartis selon le salaire mensuel

| Salaire en 10 ³ UM | Salariés fi% | xi | fi%xi | qi% (fi%xi %) | fi% cum ↑ Fi% | (fi%xi %)cum ↑ Qi% |
|-------------------------------|--------------|----|-------|---------------|------------------|-----------------------|
| [2-4[| 60 | 3 | 180 | 37,5 | 60 | 37,5 |
| [4-6[| 20 | 5 | 100 | 20,8 | 80 | 58,3 |
| [6-10[| 10 | 8 | 80 | 16,7 | 90 | 75 |
| [10-14[| 10 | 12 | 120 | 25 | 100 | 100 |
| Total | 100 | | 480 | 100 | | |

Figure 7.7 : la courbe pour calculer l'indice de Gini



Calcul de l'indice I_G :

$$\sum S_i = S_1 + S_2 + S_3 + S_4$$

$$S_1 = \frac{a \times b}{2} = \frac{37,5 \times 60}{2} = 1125 ; S_2 = \frac{(a+b) \times h}{2} = \frac{(37,5+58,3)}{2} \times 20 = 958$$

$$S_3 = \frac{(58,3+75)}{2} \times 10 = 666,5 ; S_4 = \frac{(75+100)}{2} \times 10 = 875$$

$$\Rightarrow \sum S_i = 3624,5 \Rightarrow I_G = \frac{5000 - \sum S_i}{5000} = \frac{5000 - 3624,5}{5000} \times 100 = 27,51\%$$

- **Interprétation de l'indice de Gini**

L'indice de Gini est toujours compris entre 0 et 1. Plus il est proche de 1 plus la concentration est forte et plus il est proche de 0 plus la concentration est faible.

- Si $I_G = 0$ → cela signifie que l'aire de concentration est nul, la courbe est confondu avec la bissectrice, on dit qu'on a une situation d'égalité parfaite.
- Si $I_G = 1$ → cela signifie que l'aire de concentration est égal à l'aire de triangle OAB, on dit qu'on a une situation d'inégalité parfaite.

Conférence n°14

**Les indices
élémentaires**

**Les indices
synthétiques**

- Indices de LASPEYRES
et de PAASCHE
- L'indice de Fisher

XI. Les indices élémentaires :

Les économistes, les sociologues et les gestionnaires étudient l'évolution dans le temps de grandeurs représentatives de phénomènes économiques et sociaux tels que les prix, la production, le chômage, les salaires. Ils comparent également ces grandeurs dans différents lieux ou pour différents groupes (entreprises, branches, départements, régions, pays,...). Pour évaluer ces évolutions ou effectuer ces comparaisons, ils utilisent principalement des indices.

Il existe deux catégories d'indices : les indices élémentaires et les indices synthétiques. Les indices élémentaires font l'objet de ce chapitre ; les indices synthétiques sont traités dans le chapitre suivant.

Introduction

Les indices sont des nombres sans unité qui traduisent les variations relatives de grandeurs dans le temps ou dans l'espace. Ces grandeurs peuvent être « simples » ou « complexes ».

Une grandeur est dite simple lorsque, dans chacune des situations dans lesquelles elle est observée, sa valeur est exprimée par un nombre unique. Le prix ou la quantité d'un produit sont des grandeurs simples, mais aussi la valeur d'un produit ou d'un ensemble de produits, tels que la dépense générée par l'achat d'un ou de plusieurs biens, le chiffre d'affaires ou les coûts d'une entreprise. Un nombre de personnes observées à un moment donné ou dans un lieu donné (bacheliers, personnes pauvres ou salariés par exemple), un nombre d'objets, une superficie d'un territoire sont aussi des grandeurs simples.

Un indice calculé sur une grandeur simple est appelé indice élémentaire.

Un indice calculé sur une grandeur complexe est appelé indice synthétique.

Indice élémentaire¹⁸ :

Un indice élémentaire est un indice qui renseigne sur l'évolution temporelle ou situationnelle (spatiale) d'une seule valeur. Il a pour définition :

¹⁸ M.BOUKELLA-BOUZOUANE, (2001), « Statistique descriptive, rappels de cours avec exercices corrigés », CASBAH Edition, Alger.pp 115-116

$$I_{\frac{t}{0}} = \frac{V_t}{V_0} \times 100$$

Où V_0 représente la valeur de référence et V_1 la valeur qui est comparée à la valeur de référence.

Dans le cas d'un indice temporel, « 0 » représente la période référence (la base) et « t » la période que l'on compare à la période de référence.

Dans le cas d'un indice de situation ou indice spatial, « 0 » représente la situation de référence (la base) et « t » la situation que l'on compare à la situation de référence.

Condition que doivent remplir les indices :

Parmi les condition que doivent remplir les indices, nous en retenons deux qui se prêtent, aisément, à une vérification mathématique.

XI.1. Condition de réversibilité

Un indice est réversible si : $I_{1/0} = \frac{V_1}{V_0} = \frac{1}{\frac{V_0}{V_1}} = \frac{1}{I_{0/1}}$

XI.2. Condition de transférabilité (ou de circularité)

$I_{2/0} = I_{2/1} \times I_{1/0}$; En effet:

$$I_{2/0} = \frac{V_2}{V_0} = \frac{V_2}{V_1} \times \frac{V_1}{V_0} = I_{2/1} \times I_{1/0} *$$

Cette formule permet changement de base.

Exemple :

Le prix de litre est passé de 20 DA en 2000 à 25 DA en 2001. Calculer l'indice simple du prix du lait, 2000/2001.

$$I_{2001/2000} = \frac{I_{2001}}{I_{2000}} \times 100 = \frac{25}{20} \times 100 = 125\%$$

Nous interprétons cet indice en disant : l'indice du prix du lait en 2001 base 100 en 2000, est de 125%. Le prix donc augmenté de 25%, soit $125\% - 100\% = 25\%$.

XII. Les indices synthétiques

Un indice synthétique est un indice calculé sur une grandeur « complexe ». Une grandeur complexe est soit un ensemble de prix, soit un ensemble de quantités de deux ou plusieurs produits.

Introduction :

Les économistes et les gestionnaires s'intéressent principalement aux variations des prix, des quantités ou des valeurs monétaires globales. Les indices calculés sont donc essentiellement des indices de prix, des indices de quantités et des indices de valeurs monétaires. Les situations d'observations sont presque toujours des périodes ; les indices seront désormais considérés à une date ou une période courante par rapport à une date ou une période de référence.

L'indice de valeur d'un ensemble de n produits est un indice élémentaire puisque la valeur des n produits est unique à chacune des dates où elle est évaluée.

En économie, on étudie souvent l'évolution des prix, des quantités et de leur produit, appelé valeur. Trois types d'indices peuvent alors être calculés : indice des prix, indice des quantités (ou indice de volume) et indice de valeur. Considérons une grandeur complexe g composée de k éléments: g_1, g_2, \dots, g_k . Ainsi, si g est l'indice des prix à la production, les éléments g_i , $i = 1, \dots, k$, désignent le nombre produits utilisés dans le calcul de l'indice. Soient par ailleurs les notations suivantes : p_0^i et p_t^i ; le prix respectivement aux dates 0 et t d'un élément g^i , et q_0^i et q_t^i ; les quantités respectivement aux dates 0 et t de ce même élément g^i ;

L'indice de valeur $I_{t/0}^V$ est donné par le rapport entre la somme des valeurs des k éléments g^i ; $i = 1, \dots, k$, de la grandeur g considérée à la date t et cette même

somme à la date de référence 0, soit :
$$I_{t/0}^V = \frac{\sum_i p_t^i \times q_t^i}{\sum_i p_0^i \times q_0^i}$$

Cet indice de valeur est relativement peu informatif au sens où, s'il augmente, il n'est pas possible de distinguer si cette hausse provient d'une augmentation des prix accompagnée d'une baisse des quantités ou de toute autre combinaison. Pour pallier cette difficulté, on suppose que l'une des deux variables (prix ou quantité) est fixe alors que l'autre varie en calculant des indices de prix et de quantités :

- On calcule un indice des prix en neutralisant l'influence des quantités, c'est-à-dire que l'on considère que les quantités sont fixes sur la période considérée.

- De façon réciproque, un indice des quantités (ou indice de volume) se calcule en neutralisant l'influence des prix, c'est-à-dire que l'on considère que les prix sont constants sur la période considérée.

Les indices des prix et des quantités les plus fréquemment utilisés sont les indices de Laspeyres et de Paasche que nous présentons ci-après.¹⁹

XII.1. INDICES DE LASPEYRES ET DE PAASCHE

Ce sont des indices synthétiques qui sont des résumés numériques des indices élémentaires lorsqu'on cherche à mesurer l'évolution d'un ensemble de plusieurs produits.

coefficient de pondération ou budgétaire du produit j par rapport à la date t :

$$\alpha_{j,t} = \frac{P_{j,t} \times Q_{j,t}}{\sum_{j=1}^n P_{j,t} \times Q_{j,t}}$$

XII.1.1 Indice de Laspeyres des prix

$$L(P)_{t/t_0} = \sum_j \alpha_{j,t_0} I(P_j)_{t/t_0} = \frac{\sum_j P_{j,t} Q_{j,t_0}}{\sum_j P_{j,t_0} Q_{j,t_0}} \times 100$$

$$L(P)_{t/t_0} = \frac{\text{Dépense de la date courante exprimé en quantités de la date de référence}}{\text{Dépense de la date de référence}} \times 100$$

XII.1.2 Indice de Laspeyres des quantités

$$L(Q)_{t/t_0} = \sum_j \alpha_{j,t_0} I(Q_j)_{t/t_0} = \frac{\sum_j P_{j,t_0} Q_{j,t}}{\sum_j P_{j,t_0} Q_{j,t_0}} \times 100$$

$$L(Q)_{t/t_0} = \frac{\text{Dépense de la date courante exprimé en prix de la date de référence}}{\text{Dépense de la date de référence}} \times 100$$

XII.1.3 Indice de Paasche des prix :

$$P(P)_{t/t_0} = \frac{1}{\sum_j \frac{\alpha_{j,t}}{I(P_j)_{t/t_0}}} = \frac{\sum_j P_{j,t} Q_{j,t}}{\sum_j P_{j,t_0} Q_{j,t}} \times 100$$

¹⁹ Elisabeth Olivier (2008), **L'essentiel de statistique descriptive** Edition Gualino, collection Les carrés

$$P(P)_{t/t_0} = \frac{\text{Dépense de la date courante}}{\text{Dépense de la date de référence exprimé en quantité de la date courante}} \times 100$$

XII.1.4 Indice de Paasche des quantités

$$P(Q)_{t/t_0} = \frac{1}{\sum_j \frac{\alpha_{j,t}}{I(Q_j)_{t/t_0}}} = \frac{\sum_j P_{j,t} Q_{j,t}}{\sum_j P_{j,t} Q_{j,t_0}} \times 100$$

$$P(Q)_{t/t_0} = \frac{\text{Dépense de la date courante}}{\text{Dépense de la date de référence exprimé en prix de la date courante}} \times 100$$

Exemple : Soit le tableau, qui donne les prix et les quantités de deux produits 1 et 2, aux périodes t_0 et t .

Tableau 7.4 : les prix et les quantités de deux produits 1 et 2

| Année | t_0 | | t | |
|------------|-------|----------|------|----------|
| | Prix | Quantité | Prix | Quantité |
| produits 1 | 10 | 4 | 14 | 8 |
| produits 2 | 6 | 12 | 5 | 9 |

Calculer les indices de Laspeyres t/t_0 et de Paasche t/t_0 (pour les prix et pour les quantités).

Réponse :

Appelons 0 l'année t_0 (année de base) et l'année t (année en cours). Faisons un tableau qui facilite les calculs :

Tableau 7.5 : Calcul des indices de Laspeyres t/t_0 et de Paasche t/t_0

| produits | P_0 | Q_0 | P_1 | Q_1 | $P_0 Q_0$ | $Q_1 P_0$ | $P_1 Q_0$ | $P_1 Q_1$ |
|--------------|-------|-------|-------|-------|-----------|-----------|-----------|-----------|
| produits 1 | 10 | 4 | 14 | 8 | 40 | 80 | 56 | 112 |
| produits 2 | 6 | 12 | 5 | 9 | 72 | 54 | 60 | 45 |
| Total | | | | | 112 | 134 | 116 | 157 |

On calcul l'indice de valeur $I_{t/0}^V$

$$I_{t/0}^V = \frac{\sum_i p_t^i \times q_t^i}{\sum_i p_0^i \times q_0^i} = \frac{157}{112} \times 100 = 0,956 \times 100 = 140,17\%$$

Les recettes ont donc augmenté de 40.17% de t_0 à t .

Mais cette augmentation est-elle plus due à une augmentation des prix ou des quantités ?

Pour pouvoir déterminer l'influence des prix ou des quantités, il faut définir des scénarios permettant d'isoler soit les prix soit les quantités.

C'est l'idée des indices de Laspeyres et de Paasche :

- Les indices de Laspeyres t/t_0 (pour les prix et pour les quantités).

Indice de Laspeyres t/t_0 des prix

$$L(P)_{t/t_0} = \frac{\sum_{j=1} P_{j,t} Q_{j,t_0}}{\sum_{j=1} P_{j,t_0} Q_{j,t_0}} \times 100 = \frac{116}{112} \times 100 = 1,0357 \times 100 = 103,57\%$$

Dans notre exemple, le prix du bien 1 a augmenté (de 10 à 14) et le prix du bien 2 a baissé.

L'indice, qui synthétise ces deux variations contraires, nous permet de conclure à une

« inflation », c'est-à-dire une augmentation du niveau général des prix égale à 3,57%.

Indice de Laspeyres t/t_0 des quantités

$$L(Q)_{t/t_0} = \frac{\sum_{j=1} P_{j,t_0} Q_{j,t}}{\sum_{j=1} P_{j,t_0} Q_{j,t_0}} \times 100 = \frac{134}{112} \times 100 = 1,1964 \times 100 = 119,64\%.$$

Dans notre exemple, la quantité du bien 1 a augmenté (de 4 à 8) et la quantité du bien

2 a baissé. L'indice, qui synthétise ces deux variations contraires, nous permet de conclure à une augmentation des volumes égale à 19,64%.

Comment interpréter les valeurs $L(P) = 103,57\%$ et $L(Q) = 119,64\%$.

A quantités fixées en année t_0 , les prix ont augmenté de 03,57% de l'année t_0 à l'année t .

A prix fixés en t_0 , les quantités ont augmenté de 19,64% de l'année t_0 à l'année t .

- Les indices de Paasche t/t_0 (pour les prix et pour les quantités).

Indices de Paasche t/t_0 des prix :

$$P(P)_{t/t_0} = \frac{\sum_{j=1} P_{j,t} Q_{j,t}}{\sum_{j=1} P_{j,t_0} Q_{j,t}} \times 100 = \frac{157}{134} \times 100 = 1,1716 \times 100 = 117,16\%$$

Dans notre exemple, le prix du bien 1 a augmenté (de 10 à 14) et le prix du bien 2 a baissé.

L'indice, qui synthétise ces deux variations contraires, nous permet de conclure à une

« inflation », c'est-à-dire une augmentation du niveau général des prix égale à 17,6% (contre 3,57% quand on utilise la formule de LASPEYRES).

Indices de Paasche t/t_0 des quantités).

$$P(Q)_{t/t_0} = \frac{\sum_{j=1} P_{j,t} Q_{j,t}}{\sum_{j=1} P_{j,t} Q_{j,t_0}} \times 100 = \frac{157}{116} \times 100 = 1,3534 \times 100 = 135,34\%$$

Dans notre exemple, la quantité du bien 1 a augmenté (de 4 à 8) et la quantité du bien 2 a baissé. L'indice, qui synthétise ces deux variations contraires, nous permet de conclure à une augmentation des volumes égale à 35,34% (contre 19,64% quand on utilise la formule de LASPEYRES).

Comment interpréter les valeurs $P(P) = 117,16\%$ et $P(Q) = 135,34\%$.

A quantités fixées en année t, les prix ont augmenté de 19.4% de l'année t_0 à l'année t.

A prix fixés en t, les quantités ont augmenté de 35.34% de l'année t_0 à l'année t.

| | | | | |
|------------------------|-------------------|-------------------|-------------------|---------------------|
| $I_{t/0}^V = 140,17\%$ | $L(P) = 103,57\%$ | $P(P) = 117,16\%$ | $L(Q) = 119,64\%$ | $P(Q) = 135,34\%$. |
|------------------------|-------------------|-------------------|-------------------|---------------------|

Il est possible que les indices de Laspeyres et de Paasche soient très déférents.

Pour nuancer cette possible situation, Fisher a défini des indices moyens des indices précédents.

XII.2. L'indice de Fisher

Définition

L'indice de Fisher des prix (resp. des quantités) est défini comme la moyenne géométrique des indices de Laspeyres et de Paasche des prix (resp. des quantités), c-à-d

$$F_{t/0}^P = \sqrt{L_{t/0}^P \times P_{t/0}^P} \quad \text{et} \quad F_{t/0}^Q = \sqrt{L_{t/0}^Q \times P_{t/0}^Q}$$

| | | | | | | |
|------------------------|-------------------|-------------------|------------------------|---------------------|-------------------|------------------------|
| $I_{t/0}^V = 140,17\%$ | $L(P) = 103,57\%$ | $P(P) = 117,16\%$ | $F_{t/0}^P = 110,15\%$ | $L(Q) = 119,64\%$. | $P(Q) = 135,34\%$ | $F_{t/0}^Q = 127,24\%$ |
|------------------------|-------------------|-------------------|------------------------|---------------------|-------------------|------------------------|

Interprétation :

D'un point de vue global, l'augmentation de 40,17% des recettes globales est davantage due à l'augmentation des quantités (de l'ordre de 27,24% de l'année t_0 à l'année t) qu'à l'augmentation des prix (de l'ordre de 10,15%).

Exercices résolues

Exercice 01 :

Les salariés d'une entreprise sont répartis selon le salaire mensuel en 1000DA

| Salaires | xi | ni |
|----------|----|-----|
| 10-14 | 12 | 20 |
| 14-20 | 17 | 60 |
| 20-24 | 22 | 10 |
| 24-30 | 27 | 10 |
| Total | | 100 |

- 1- Calculer la médiane Me et la Médiale Mle de cette distribution
- 2- Analyser la concentration correspondant à cette distribution en précisant les valeurs globales, puis les valeurs globales relatives ainsi les valeurs globales relatives cumulées croissantes en %.
- 3- Traçer la courbe de concentration et calculer l'indice de Gini

Exercice 02 :

Pour trois produits A, B et C, les prix en euros par kg et les quantités consommées en kg ont été les suivants aux dates 0 et t :

| Produit | Prix (date 0) | Quantité (date 0) | Prix (date t) | Quantité (date t) |
|---------|---------------|-------------------|---------------|-------------------|
| A | 10 | 5 | 15 | 4 |
| B | 8 | 3 | 9 | 4 |
| C | 10 | 2 | 9 | 2 |

- 1- Pour chaque produit, calculer l'indice élémentaire du prix à la date t, base 100 à la date 0, puis de l'indice élémentaire de quantité et l'indice élémentaire de dépense.
- 2- Calcul du coefficient budgétaire de chaque produit en 0, puis en t.
- 3- Calcul de l'indice de la dépense globale à la date t, base 100 à la date 0.
- 4- Calcul des indices des prix et des quantités de Laspeyres et de Paasche.
- 5- Calcul de l'indice des quantités et l'indice des prix de Fisher.

Solution des exercices

Corrigé de l'exercice 01 :

1- D'abord on calcul les fréquences puis les fréquences cumulés croissants

| Salaires | xi | ni | fi=ni/N | Fi cum | ni*xi | qi% | Qi% |
|----------|----|-----|---------|--------|-------|-----|------|
| 10-14 | 12 | 10 | 10% | 10% | 120 | 6% | 6% |
| 14-20 | 17 | 30 | 30% | 40% | 510 | 25% | 31% |
| 20-24 | 22 | 40 | 40% | 80% | 880 | 43% | 74% |
| 24-30 | 27 | 20 | 20% | 100% | 540 | 26% | 100% |
| Total | | 100 | 1 | | 2050 | 1 | |

$$F(\text{Me})=1/2=0,5 \Rightarrow 0,4 < 0,5 < 0,8 \Rightarrow \text{Me} \in [20-24[\text{donc } \text{Me} = e_i + a_i \frac{0,5 - F(x_i)^\uparrow}{f_i} = 20 + 4 \times \frac{0,5 - 0,4}{0,4} =$$

21 donc 50% des salariés reçoivent moins de 21000 DA.

Calcul de la médiale :

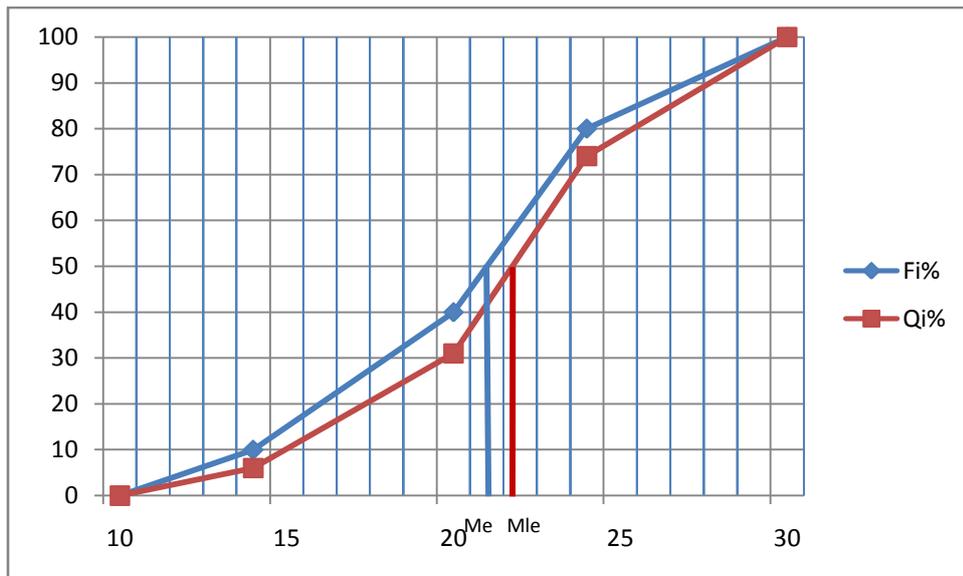
$$\text{Mle} = a + (b-a) \times \frac{50 - Q(a)}{Q(b) - Q(a)}$$

Revenant à notre exemple, on a le tableau suivant :

| | | | |
|-------------------------|---------|-----|---------|
| Valeur de la variable | a=20 | Mle | b=24 |
| Val.Glob.Rel.Cum.Croiss | Q(a)=31 | 50 | Q(b)=74 |

$$\text{Mle} = 20 + (24-20) * [(50-31)/(74-31)] = 21,767$$

Les salariés de l'entreprise dont le salaire mensuel est inférieur à 21767DA se partagent (ensemble) la moitié de la masse salariale ou la moitié de la masse totale des salaires distribuée par l'entreprise à ses salariés concerne les salariés dont le traitement mensuel est inférieur à 21767DA.



L 'écart Médiale – Médiane : ΔM

La comparaison de la valeur de la médiale et de la médiane constitue une mesure de la concentration. La mesure de l'écart entre médiale et médiane renseigne sur le degré de concentration. Plus l'écart est grand et plus la concentration est forte et il est nul en cas de parfaite égalité.

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

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

- Si ΔM est grand par rapport à l'étendue : la concentration est forte
- Si ΔM est petit par rapport à l'étendue : la concentration est faible
- Si ΔM est nul : la concentration est nulle ($Mle = Me$) → égalité parfaite

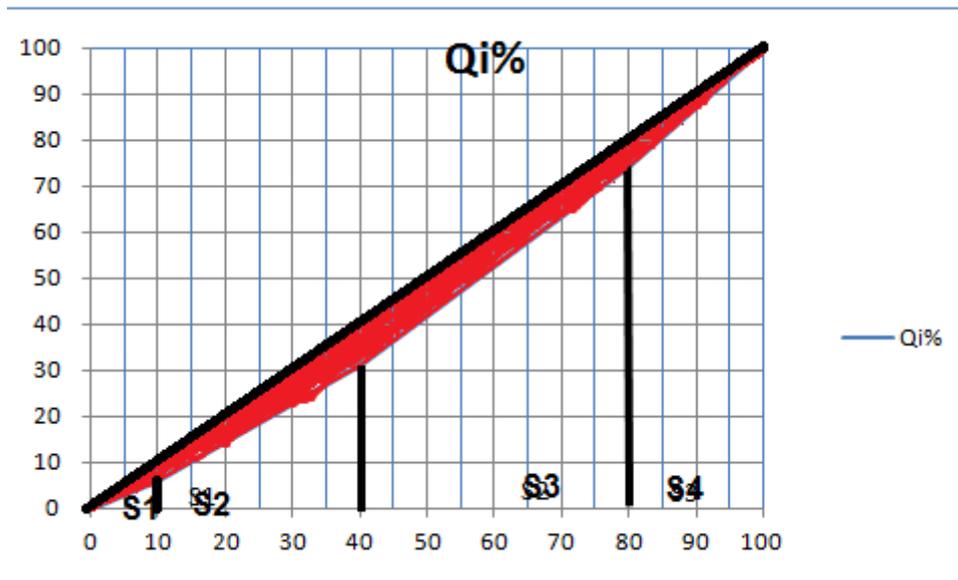
$$\Delta M = Mle - Me = 21767 - 21000 = 767$$

$$E = X_{\max} - X_{\min} = 30 - 10 = 20$$

$$\frac{\Delta M}{E} \times 100 = \frac{767}{20000} \times 100 = 03,38\%$$

La concentration est relativement faible car elle représente 03.38% de l'étendue.

3. La courbe de concentration :



Calcul de l'indice I_G :

$$\sum S_i = S_1 + S_2 + S_3 + S_4$$

$$S_1 = \frac{a \times b}{2} = \frac{10 \times 6}{2} = 30 ; S_2 = \frac{(a+b) \times h}{2} = \frac{(6+31)}{2} \times 30 = 555$$

$$S_3 = \frac{(31+74)}{2} \times 40 = 2100 ; S_4 = \frac{(74+100)}{2} \times 20 = 1740$$

$$\Rightarrow \sum S_i = 4425 \Rightarrow I_G = \frac{5000 - \sum S_i}{5000} = \frac{5000 - 4425}{5000} \times 100 = 11,5\%$$

- **Interprétation de l'indice de Gini**

La concentration est faible car elle représente 11.5% du triangle.

Corrigé de l'exercice 02 :

| Produit | Prix (date 0) | Quantité (date 0) | Prix (date t) | Quantité (date t) | $P_0 \times Q_0$ | $P_t \times Q_t$ | $P_0 \times Q_t$ | $P_t \times Q_0$ |
|---------|---------------|-------------------|---------------|-------------------|------------------|------------------|------------------|------------------|
| A | 10 | 5 | 15 | 4 | 50 | 60 | 40 | 75 |
| B | 8 | 3 | 9 | 4 | 24 | 36 | 32 | 27 |
| C | 10 | 2 | 9 | 2 | 20 | 18 | 20 | 18 |
| Tota | | | | | 94 | 114 | 92 | 120 |

- 1- Pour chaque produit, calcul de l'indice élémentaire du prix à la date t, base 100 à la date 0, puis de l'indice élémentaire de quantité et l'indice élémentaire de dépense :

$$I(p_A)_{t/0} = \frac{15}{10} \times 100 = 150; I(p_B)_{t/0} = \frac{09}{08} \times 100 = 112,5; I(p_C)_{t/0} = \frac{09}{10} \times 100 = 90$$

$$I(Q_A)_{t/0} = \frac{04}{05} \times 100 = 80; I(Q_B)_{t/0} = \frac{04}{03} \times 100 = 133,33; I(Q_C)_{t/0} = \frac{02}{02} \times 100 = 100$$

$$I(D_A)_{t/0} = \frac{60}{50} \times 100 = 120; I(D_B)_{t/0} = \frac{36}{24} \times 100 = 150; I(D_C)_{t/0} = \frac{18}{20} \times 100 = 90$$

Bien A : le prix a augmenté de 50 %, la quantité a diminué de 20 %, donc la dépense a augmenté de 20 %.

Bien B : le prix a augmenté de 12,5 %, la quantité a augmenté de 33,33 %, donc la dépense a augmenté de 50 %.

Bien C : le prix a diminué de 10 %, la quantité est restée constante, donc la dépense a diminué de 10 %.

- 2- Calcul du coefficient budgétaire de chaque produit en 0, puis en t :

$$C_0^A = \frac{50}{50+24+20} = \frac{50}{94} = 0,5319; C_0^B = \frac{24}{94} = 0,2593; C_0^C = \frac{20}{94} = 0,2128$$

$$C_t^A = \frac{60}{60+36+18} = \frac{60}{114} = 0,5263; C_t^B = \frac{36}{114} = 0,3158; C_t^C = \frac{18}{114} = 0,1579$$

- 3- Calcul de l'indice de la dépense globale à la date t, base 100 à la date 0 :

$$I_{t/0}^V = \frac{\sum_i p_t^i \times q_t^i}{\sum_i p_0^i \times q_0^i} = \frac{114}{94} \times 100 = 121,27$$

La dépense globale a donc augmenté de 21,27 %.

- 4- Calcul des indices des prix et des quantités de Laspeyres et de Paasche, en utilisant les deux formules pour chacun des indices :

$$L(P)_{t/t_0} = \frac{\sum_{j=1} P_{j,t} Q_{j,t_0}}{\sum_{j=1} P_{j,t_0} Q_{j,t_0}} \times 100 = \frac{120}{94} \times 100 = 127,66$$

$$P(P)_{t/t_0} = \frac{\sum_{j=1} P_{j,t} Q_{j,t}}{\sum_{j=1} P_{j,t_0} Q_{j,t}} \times 100 = \frac{114}{92} \times 100 = 123,91$$

$$L(P)_{t/t_0} = \sum_{i=1}^n C_0^i I(p_i)_{t/0} = 0,5319 \times 150 + 0,2553 \times 112,5 + 0,2128 \times 90$$

$$= 127,66$$

$$P(P)_{t/t_0} = \frac{1}{\sum_{i=1}^n C_t^i / I(p_i)_{t/0}} = \frac{1}{\frac{0,5263}{150} + \frac{0,3158}{112,5} + \frac{0,1579}{90}} = 123,91$$

D'après l'indice de Laspeyres, les prix des trois produits ont augmenté de 27,67 %, tandis que d'après l'indice de Paasche, ils ont augmenté de 23,91 %. Cette différence dans les résultats est due, d'une part, à l'évolution des coefficients budgétaires entre les dates 0 et t et, d'autre part, à l'utilisation pour les calculs de deux types de moyennes distinctes. Il n'y a aucun critère objectif qui permette de prétendre qu'un des résultats est « meilleur » que l'autre.

$$L(Q)_{t/t_0} = \frac{\sum_{j=1}^n P_{j,t_0} Q_{j,t}}{\sum_{j=1}^n P_{j,t_0} Q_{j,t_0}} \times 100 = \frac{92}{94} \times 100 = 97,87$$

$$P(Q)_{t/t_0} = \frac{\sum_{j=1}^n P_{j,t} Q_{j,t}}{\sum_{j=1}^n P_{j,t} Q_{j,t_0}} \times 100 = \frac{114}{120} \times 100 = 95$$

$$\begin{aligned} L(Q)_{t/t_0} &= \sum_{i=1}^n C_0^i I(Q_i)_{t/t_0} = 0,5319 \times 80 + 0,2553 \times 133,33 + 0,2128 \times 100 \\ &= 97,87 \end{aligned}$$

$$P(Q)_{t/t_0} = \frac{1}{\sum_{i=1}^n C_t^i / I(Q_i)_{t/t_0}} = \frac{1}{\frac{0,5263}{80} + \frac{0,3158}{133,33} + \frac{0,1579}{100}} = 95$$

D'après l'indice de Laspeyres, les quantités consommées des trois produits ont diminué de 2,13 %, tandis que d'après l'indice de Paasche, elles ont diminué de 5 %. Ces deux résultats sont sensiblement différents, et, comme pour les indices de prix, il n'est pas possible de considérer que l'un est plus satisfaisant que l'autre.

- 5- Calcul de l'indice des quantités et l'indice des prix de Fisher, à la date t, base 100 à la date 0 :

$$F_{t/t_0}^P = \sqrt{L_{t/t_0}^P \times P_{t/t_0}^P} = \sqrt{127,66 \times 123,91} = 125,77$$

$$F_{t/t_0}^Q = \sqrt{L_{t/t_0}^Q \times P_{t/t_0}^Q} = \sqrt{97,87 \times 95} = 96,42$$

D'après ce calcul, globalement, les prix des trois produits ont augmenté de 25,77 % et les quantités consommées ont diminué de 3,58 %.

Ces calculs résultant de moyennes, ils donnent des valeurs intermédiaires entre les valeurs des indices de Laspeyres et de Paasche : 25,77 % est compris entre 27,67 % et 23,91 %, 3,58 % est compris entre 2,13 % et 5 %.

Exercice proposées

Exercice 01 :

Les salariés d'une entreprise sont répartis comme suit en fonction de leur salaire brut mensuel (en dix milliers de DA) :

| Salaires x10 ⁴ DA | [3 ; 4[| [4 ;5[| [5 ;6[| [6 ;7[| [7 ;8[|
|------------------------------|---------|--------|--------|--------|--------|
| Effectif | 26 | 33 | 64 | 07 | 10 |

- 1- Calculer la médiane Me et la Médiale Mle de cette distribution
- 2- Analyser la concentration correspondant à cette distribution en précisant les valeurs globales, puis les valeurs globales relatives ainsi les valeurs globales relatives cumulées croissantes en %.
- 3- Tracer la courbe de concentration et calculer l'indice de Gini

Exercice 02 :

Soit le tableau ci-après, qui donne les prix et les quantités de deux produits 1 et 2, aux dates 0 et t. On peut supposer que le produit 1 est un pantalon et le produit 2 un tee shirt

| | P_0 | Q_0 | P_1 | Q_1 |
|-----------|-------|-------|-------|-------|
| Produit 1 | 15 | 03 | 22 | 10 |
| Produit 2 | 07 | 09 | 05 | 08 |

- 1- Pour chaque produit, calculer l'indice élémentaire du prix à la date t, base 100 à la date 0, puis de l'indice élémentaire de quantité et l'indice élémentaire de dépense.
- 2- Calcul du coefficient budgétaire de chaque produit en 0, puis en t.
- 3- Calcul de l'indice de la dépense globale à la date t, base 100 à la date 0.
- 4- Calcul des indices des prix et des quantités de Laspeyres et de Paasche.
- 5- Calcul de l'indice des quantités et l'indice des prix de Fisher.

Bibliographie

- Alain Baccini (2010) ; **Statistique Descriptive Élémentaire** ; *Publications de l'Institut de Mathématiques de Toulouse* ;p 13.
- Bourbonnais.R, (2015), « Econométrie, Cours et exercices corrigés », 9^{ème} édition, *Edition Dunod, Paris*,pp17-19
- BOUKELLA-BOUZOUANE.M, (2001), « Statistique descriptive, rappels de cours avec exercices corrigés », *CASBAH Edition, Alger*.pp 115-116
- Chareille.P, Pinault.Y, (1998), **Statistique descriptive**, *Edition Montchrestien, Paris*, pp125-128
- Dreyfus.G, Martinez.J.-M, M. Samuelides,M. B. Gordon, F. Badran, S. Thiria, (2008), “apprentissage statistique”, *ÉDITIONS EYROLLES, Paris*.
- Goldfarb.B et Pardoux.C, 2013; « Introduction à la méthode statistique, Statistique et probabilités, Cours et exercices corrigés » *Edition Dunod, Paris*, pp 13-14
- HURLIN.C, MIGNON.V, (2015), « STATISTIQUE ET PROBABILITES EN ECONOMIE-GESTION », *Edition Dunod, Paris*.
- LEBOUCHER.L, VOISIN. M.-J., (2011), « Introduction à la statistique descriptive, Cours et exercices avec tableur », *CÉPADUÈS-ÉDITIONS, Toulouse – France*.
- MAZEROLLE.F, (2005), « Statistique Descriptive », *Edition Gualino éditeur, EJA – Paris – 2006, ISBN 2 - 84200 - 891 – X, p80*
- Olivier.E, (2008), « L'essentiel de statistique descriptive », *Edition Gualino, collection Les carrés*.

Ressources électroniques :

- http://www.astro.ulg.ac.be/cours/magain/STAT/Stat_Main_Fr/Download/StatChap1.pdf
- <http://mathsv-ressources.univ-lyon1.fr/cours/stats/chap5/c5p4/c5p4.html>
- <http://serge.mehl.free.fr/exos/histog.html>
- <http://www.ofarcy.net/documentation/Statistique-descriptive.pdf>
- https://www.academia.edu/19786928/Econometrie_suite_lareq
- <https://p9.storage.canalblog.com/98/61/1177787/100938518.pdf>
- <https://www.fichier-pdf.fr/2019/10/31/chapitre-4-les-caracteristiques-de-concentration/preview/page/2/?>