République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

ABOUD SOUMIA

Titre :

Statistique descriptive des données financière

Membres du Comité d'Examen :

Dr.	**SAYAH ABDALLAH**	UMKB	Président
Dr.	**BRAHIM BRAHIMI**	UMKB	Encadreur
Dr.	**SOURAYA KHEREDDINE**	UMKB	Examinateur

September 2020

Dédicace

A vant tout je remarcie mon "Dieu" qui me donne la patience et la force pour réaliser ce modeste travail,qui est toujours avec moi le long de ma vie.

Je dédie ce travail à :

A l'âm pure ma mère Sarhouda.

A mon très cher père Mouhammed.

A mon chère frères : Thabet, Issam, Sami.

A mes très chère amis : Fayza, Nadia, Mouna, Fatima, khawla, samira, khadidja, Manal, Sabrina.

Tous mes amis qui vivent à **BISKRA** ou ailleurs.

. Soumia.

REMERCIEMENTS

Au nom de Dieu le Miséricordieux

"Si vous êtes reconnaissants, très certainement j'augmenterai pour vous..."

Ibrahim-7

Mercie à notre "**Dieu**", notre guide, notre force, et la raison de notre existence, c'est lui qui nous a fait comprendre le but de cette vie,

et qui nous a donné le povoir d'apprécier les choses, et qui nous a donné le courage et la voloné pour terminer ce travail.

Tout d'abord, je tiens à exprimer toute ma reconnaissance à mon encadreur

Monsieur le **Dr. BRAHIM BRAHIMI**.

Je le remercie pour sa patience, ses orientations et conseils.

J'adresse aussi mes remerciements aux membres du Jury

Monsieur le **Dr.SAYAH ABDALLAH**

et

Madame le **Dr. SOURAYA KHEREDDINE**.

Nous remercions aussi toute personne ayant contribué, de prés ou de loin, à la réalisation de ce travail.

Mercie à tous.

Table des matières

Table des figures

Liste des tableaux

Introduction

Les statistiques sont l'une des importante branche en mathématiques, avec diverses applications et ce sont de élément essentiel dans chaque thèse scientifique.

L'action de dénombrer qui renvoie aux statistiques fût mentionnée, dans le saint Coran, Le Très Haut dit " Alors que nous avons dénombré toutes choses écrit" Al Nabaa-29.

Les statistiques sont connus comme étant le savoir qui s'intéresse ou qui thèse dans un recueil de données ; il les organise et les expose, puis il les analyse et donnée des résultats, Et c'est sur cette base que les décision sont prises.

Cela comprend l'organisation et l'exposition des données, qui concernent une forme quelconque, en les simplifiant dans des tableaux et diagrammes. Cette méthode est la première utilisée dans les statistiques, les statistiques se divisent en deux parties :

1- Statistiques descriptives : Est un ensemble de méthodes permettant de décrire, présenter, résumer des données souvent trés nombreuses.

2- Statistiques inférentielle : Est d'effectuer des estimations et des prévisions à partir d'un sous-ensemble de population.

Dans notre étude, nous nous intéresserons aux statistiques descriptives, qui à leur tour, sont composées en deux catégories.

1- La statistique descriptive univariée : Correspond à l'analyse d'un seul caractère, c'est l'étude de la population selon une seule variable.

2- La statistique descriptive multivariées : Les analyses multivariées, c'est l'étude de la population à plusieurs variables.Les statistique descriptive bivariées sont des cas particuliers à

deux variables.

Les questions qui se posent à travers l'étude du sujet des statistiques descriptives, sont nombreuses, parmi elles, par exemple :

Qu'est-ce que les statistiques descriptives ?

Quel est l'intérêt des statistiques descriptives dans notre vie quotidienne ?

Quelles sont les étapes des statistiques descriptives ?

Les statistiques ont fait leur apparition, dans les anciennes époques, comme il a été cité dans la Sunnah du prophète, que la paix d'Allah et le salut soit sur lui : "dénombrez moi combien l'islam est prononcé" Recueilli par Muslim.

Et jusqu'à la fin du 19 ème siècle, les statistiques sont restées, de façon essentielle, parmi les techniques du calcul (démographie, calcul des nombres des soldats, les armes...etc).

Vers la fin du 19 ème siècle, et en 1960, les statistiques ont été développées, et ont suivi le développement général des sciences, mais surtout des maths et de la physique.

Une étudiante m'a précédé, des département de mathématiques en 2018 l'étude statistique descriptive des reposait sur un seul variable et j'ajouterai dans ma dans cette thèse les statistique descriptive deux variables.

Basé dans cette thèse sur quatre chapitres principaux :

Chapiter1 : Généralités sur la statistique.

Qu'est-ce que la statistique ? Quelques définitions.

Concept de base statistique et notation standard. Graphique. Exercices.

Chapiter2 : Statistique descriptive univariée.

Les caractéristiques de tendance centrale (mode, moyenne, médiane, quantiles, etc.),les caractéristiques de dispersion (variance, écart-type, coefficient de variation, etc.), mesures de forme (l'asymétrie, l'aplatissement), Le test du Chi-deux. Diagrammes. Exercices.

Chapiter3 : Statistique descriptive bivariée.

Les caractéristiques de tendance centrale et dispersion (moyenne marginale, variance margi-

nale, etc), deux variables qualitatives (le test du khi-deux, test d'indépendance, etc), deux variables quantitatives (cofficient de corrélation, droite de régression, etc). Exercices.

Chapiter4 : Logiciel R. Exercices.

Chapitre 1

Généralités sur la statistique

La statistique est une méthode d'analyse des ensembles comportant un grand nombre d'éléments. C'est une science qui permet de traiter et d'analyser les résultats des mesures effectuées sur les individus d'une population relativement un certain nombre de caractères. Les résultats des mesures sont, en général, appelés observations.[Meghlaoui (2011)]

Les statistiques jouent un rôle de plus important dans tous les aspects de l'activité humaine, il sert d'autres sciences et les aide à développer et à étendre des recherches scientifiques précises et solides, comme il est courmment utilisé dans, agriculture, administration des affaires, physique et chimie....[Hachmi]

1.1 Définitions fondamentales

Définition 1.1.1 ***La statistique :*** *Est une branche des mathématiques appliquées qui a pour objet l'étude des phénoménes mettant en jeu un grand nombre d'éléments, les statistiques consistent en diverses méthodes de classement des données tels que les tableaux, les histogrammes et les graphiques ensuite[Chekroun (2018)], analysez-le et interprétez les résultats.*

Définition 1.1.2 ***La statistique descriptive[Goldfarb, Catherine (2011)]*** *: Est un ensemble de méthodes permettant de décrire, présenter, résumer des données souvent trés nom-*

breuses. Ces méthodes peuvent être numériques (tris, élaboration de tableaux, calcul de moyenne, ...) ou mener à des représentations graphique. Elle statistique descriptive se compose de deux domaines distincts :

-La statistique descriptive univariée[Mémoire (2018)] : Correspond à l'analyse d'un seul caractère, c'est l'étude de la population selon une seule variable.

-La statistique descriptive bivariée[Mémoire (2018)] : Est l'étude de la relation qui peut exister entre deux variable, que l'on traite avec des méthode comme l'analyse.

Définition 1.1.3 ***La statistique théorique ou mathématique[Lethielleux (2016)] :*** *Est de formuler des lois coprotement à partir d'obsarvation souvent incomplétes qui prend la suite de la statistique descriptive lorsque l'on peut énoncer ou élaborer des loi : Loi khi deux, loi normale, ...etc.[Fabrice (2006)]*

1.2 Concept de base statistique

Population[Goldfarb, Catherine (2011)] : Une population est l'ensemble des éléments aux quels se rapportent les données étudiées. Cet ensemble est noté Ω.[Chekroun (2018)] En statistique, le terme ‹ population › s'applique à des ensembles de toute nature : étudiants d'une académie, production d'une usine, poissons d'une riviére, entreprises d'un secteure donné...

Unité statistique[Fabrice (2006)] : Les unités d'une population, que le critére soit qualitatif ou quantitatif. Peuvent étre présentées individuellement (c'est généralement le cas lorsque les données sont saisies) ou regroupées. On appelle individu tout élément de la population Ω, il est noté ω (ω dans Ω)[Chekroun (2018)].

Caractère (Variable statistique)[Chekroun (2018)] : On appelle caractère (ou variable statistique, dé notée V.S) toute application $\mathbf{X} : \Omega \to C$.

Modalités[Chekroun (2018)] : Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci une variable.

Echantillon[Hammed (2012)] : Est un sous ensemble de la population statistique, il n'est généralement pas possible de collecter des données sur tous les éléments d'une population alors on se contente d'extraire une partie de la population appelée échantillon et restreindre l'étude à cet échantillon.Le nombre d'éléments dans l'échantillon s'appelle taille de l'échantillon et sera noté par N.[Mémoire (2018)]

Série statistique :[Dagnelie (2006)] Est une ensemble de valeure obtenues à l'observation d'un phénoméne La forme la plus élémentaire de présentation des données statistiques relatives à une seul variable consiste en une simple énumération des observations par ordre croissant :

$$x_1 \leq x_2 \leq ... \leq x_i \leq ... \leq x_k$$

Tableau statistique :[Dagnelie (2006)]Est une méthode permettent de présenter les données sous la forme numérique**.** Il peut être aussi bien utilisé pour représenter des données brutes que des résultats statistiques.

1.3 Type du caractère

1.3.1 Les caractères qualitatives

Une variable est dite qualitative si ses différentes modalités ne sont pas numérique [Goldfarb, Catherine (2
Ainsi : La situation matrimoniale, le nationalité, la profession,..., sont des variables les éléments de **C** [Chekroun (2018)] sont représentés par autre chose que des chiffres. Le a deux catégories :

Une variable qualitative ordinale

Elle dite ordinale quand les modalités peuventêtre naturellement ordonnées [Mémoire (2018)], par exemple : grade, classe sociale,..., etc.

Une variable qualitative nominale

Elle est dite nominale lorsque ses modalitésne peuvent être classées de façon naturelle [Mémoire (2018)] par exemple : la variable couleur des yeux,..., etc.

1.3.2 Les caractères quantitatives

Une variable est dite quantitative lorsqu'elle est intrinséquement numérique, une variable quantitative peut être une variable statistique discréte ou continue **[Goldfarb, Catherine (2011)]**.

Les variables statistiques discrètes

Une variable statistique est dite discrète lorsqu'elle ne peut prendre que des valeure isolées dans son intervalle de variation(représenté par nombres naturels $\mathbb{N}$) comme "nombre de maisons vendues par ville" [Goldfarb, Catherine (2011)].

Les variables statistiques continues

Une variable statistique est dite continue lorsqu'elle peut prendre toutes les valeure à l'intérieur de son intervalle de variation(représenté par nombres décimaux $\mathbb{Q}$) comme" revenu brut" [Goldfarb, Catherine (2011)].

1.4 Effectif partiel -Effectif total

On suppose un caractère statistique numérique admet chaque modalité x_i avec i varie de 1 à K.

$$X : \begin{array}{c} \Omega \to \{x_1, x_2, ..., x_n\} \\ Card\,(\Omega) := N \end{array}$$

1.4.1 Effectif partiel

Définition 1.4.1 *[Chekroun (2018)] Pour chaque valeure x_i, on pose par définition.On a :*

$$n_i = Card\{\omega \in \Omega : X(\omega) = x_i\} \tag{1.1}$$

n_i : S'appelle effectif partiel de x_i.

1.4.2 Effectif total

Définition 1.4.2 *[Fabrice (2006)] L'éffectif total N d'une valeure est la somme de l'effectif de cette valeure et de tous les effectifs des valeure qui précédent. Pour chaque valeure x_i, on pose par définition : [Chekroun (2018)]*

$$N = n_1 + n_2 + ... + n_k; N = \sum_{i=1}^{k} n_i = card(\Omega). \tag{1.2}$$

1.5 Fréquence partielle (fréquence relative)

Définition 1.5.1 *[Chekroun (2018)] La fréquence relative est égale à la effectifs divisée par l'éffectif total. Pour chaque valeur x_i, on pose par définition :*

$$f_i = \frac{n_i}{N} \tag{1.3}$$

Remarque 1.5.1 *[Chekroun (2018)]On peut remplacer f_i par $f_i * 100$ qui représente alors un pourcentage où $f_i =$ est le pourcentage des ω tel que $X(\omega) = x_i$.*

-La valeur de la fréquence relative est toujour entre 0 et 1.On peut multiplier la fréquence par 100, ainsi on obtient une fréquence exprimée en %, entre 0% et 100%.

Propriété 1.5.1 *[Chekroun (2018)] Soit f_i défini comme $f_i = \frac{n_i}{N}$, alors $\sum_{i=1}^{k} f_i = 1$, ou $\sum_{i=1}^{k} f_i(\%) = 100$, le cas des fréquences en pourcentage.*

Preuve. [Chekroun (2018)] Rappelons que $\sum_{i=1}^{k} n_i = N$. Ce qui implique que

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

Méme pour cas pourcentage. ■

1.6 Cumulée croissante

Quand les valeurs du caractère sont rangées dans l'ordre croissant [Chekroun (2018)], la fréquence cumulée croissante FCC (ou effectif cumulé croissant ECC) d'une valeure est la somme des fréquences (ou effectif) de cette valeur et de celles qui la précédent.

1.7 Cumulée décroissante

Quand les valeurs du caractère sont rangées dans l'ordre croissant [Chekroun (2018)], la fréquence cumulée décroissante FCD (ou effectif cumulé décroissante ECD) d'une valeur est la somme des fréquence (ou effectif) de cette valeur et de celles qui la suivent.

1.8 Représentation graphique des séries statistiques

1.8.1 Représentation graphique des caractères qualitatives

Les modalités d'un caractère qualitatif n'étant pas des ordonnées, on les représente généralement par des graphiques qui utilisent des surfaces il existe deux types de représentations fréquemment utilisées : représentation en cercle et rectangle [Leboucher, Marie (2013)].

Représentation par secteur

Dans cette représentation les aires et par conséquent les angles au centre sont proportionnels aux effectifs (ou aux fréquences) des différentes modalités [Leboucher, Marie (2013)]. En effet

Modalités du caractère X	Effectifs n_i	ECC N_i	ECD $N_i^{'}$
x_1	n_1	$N_1 = n_1$	$N_1^{'} = n_k + ... + n_1$
x_2	n_2	$N_2 = n_1 + n_2$	$N_2^{'} = n_k + ... + n_2$
...	...	...	...
x_i	n_i	$N_i = n_1 + ... + n_i$	$N_i^{'} = n_k + ... + n_i$
...	...	...	...
x_k	n_k	$N_k = n_1 + ... + n_k$	$N_k^{'} = n_k$
Total	N		

TAB. 1.1 – Table statistique d'un caractère qualitatifs

Fréquence f_i	FCC F_i	FCD $F_i^{'}$
f_1	$F_1 = f_1$	$F_1^{'} = f_k + ... + f_1 = 1$
f_2	$F_2 = f_1 + f_2$	$F_2^{'} = f_k + ... + f_2$
...	...	...
f_i	$F_i = f_1 + ... + f_i$	$F_i^{'} = f_k + ... + f_i$
...	...	...
f_k	$F_K = f_1 + ... + f_i + ... + f_k = 1$	$F_k^{'} = f_k$
1		

TAB. 1.2 – Table statistique d'un caractère qualitatifs

$$\alpha_i = 360^\circ \frac{n_i}{N} = 360^\circ f_i \tag{1.4}$$

Représentation par rectangle

Cette représentation fait figurer les différentes modalités du caractère sous forme de rectangle dont la base est constante et dont la hauteur est proportionnelle à l'effectif (ou à la fréquence) [Leboucher, Marie (2013)].

Exemple 1.8.1 *[Lethielleux, Chevalier (2017)]On s'intrésse à une série statistique du variable "état civil" sur* 22 *personnes. On obtient le tableau suivant :*

état civil X	Effectif personnes n_i	Fréquence f_i
C : célibataire	10	0.454
M : marié(e)	4	0.182
V : veuf(ve)	6	0.273
D : divorcée	2	0.091
Total	22	1

TAB. 1.3 – Table de stasti pour l'état civil

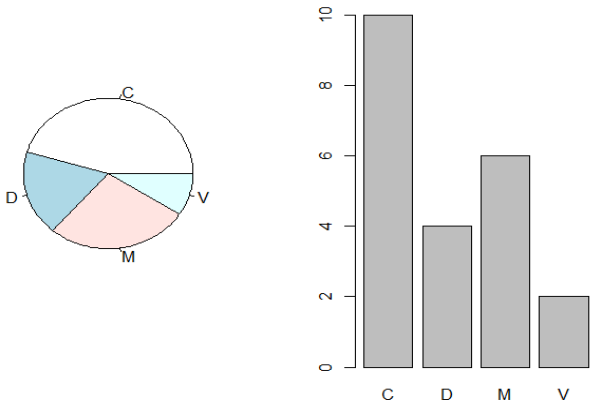

FIG. 1.1 – Diagramme en secteurs et barres des effectifs d'une variable qualitative

1.8.2 Représentation graphique des caractères quantitatives

Caractère discrète[Meghlaoui (2011)]

Dans le cas des séries statistiques discrètes il existe deux types de représentations graphiques.

Tableau statistique d'un caractère quantitatif discrète

La représentation en diagramme en bâtons :[Meghlaoui (2011)] La représentation en diagramme en bâtons est la représentation de la distribution des fréquences ou des effectifs d'une caractère discrète. A chaque valeur x_i portée en abscisse on fait correspondre un segment vertical de longueur proportionnelle à l'effectif n_i ou à la fréquence f_i de cette valeur.

Modalités du caractère X	Effectif	ECC	ECD	Fréquencef_i	FCC	FCD
x_1	n_1	N_1	$N_1^{'}$	f_1	F_1	$F_1^{'}$
x_2	...	...	...	...	...	...
...	...	...	...	...	...	...
x_i	...	...	...	...	...	...
...	...	...	...	...	...	...
x_k	n_k	N_k	$N_k^{'}$	f_k	F_k	$F_k^{'}$
Total	N			1		

TAB. 1.4 – Table statistique d'un caractère quantitatif discrète

La représentation en diagramme en cumulatifs La courbe cumulative est la représontation graphique des effectifs cumulés ou des fréquences cumulées [Meghlaoui (2011)]. C'est un graphique en escalier dont les paliers horizontaux ont pour ordonnées respectivement F_i ou N_i.

Exemple 1.8.2 *[Lethielleux, Chevalier (2017)]Le tableau suivant donne la distribution de* 90 *familes selon le nombre d'enfants.*

Nombre d'enfants	Nombre familes n_i	Fréquences f_i	FCC
0	15	0.167	0.167
1	20	0.222	0.389
2	20	0.222	0.611
3	15	0.167	0.778
4	8	0.089	0.867
5	5	0.055	0.922
6	3	0.033	0.955
7	2	0.022	0.977
8	2	0.022	$\simeq 1$
Total	90	$0.999 \simeq 1$	

TAB. 1.5 – Table la distribution de 90 familes selon le nombre d'enfants.

Tracer le diagramme en bâtons de cette effectif.

Déterminer la fonction de répartition de cette distribution.

Tracer la courbe des fréquence cumulées croissantes. La fonction F pour une variable discrète est constante par moroeaux, c'est une fonction en escalier.

Si :

$$0 \leq x < 1 F(x) = 0.167$$

$$1 \leq x < 2 F(x) = 0.389$$

$$2 \leq x < 3 F(x) = 0.611$$

$$3 \leq x < 4 F(x) = 0.778$$

$$4 \leq x < 5 F(x) = 0.867$$

$$5 \leq x < 6 F(x) = 0.922$$

$$6 \leq x < 7 F(x) = 0.955$$

$$7 \leq x < 8 F(x) = 0.977$$

$$x \geq 8 F(x) = 1.$$

Trace le diagramme en bâtons et la courbe des FCC.

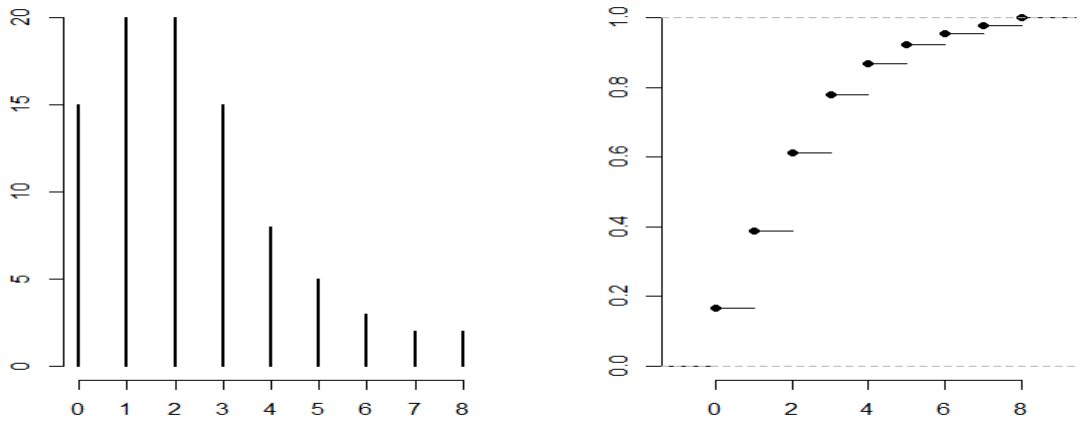

FIG. 1.2 – Diagramme en bâtonnets des effectifs et fonction de répartition d'une variable quantitative discrète

Caractère continue[Meghlaoui (2011)]

Comme pour les caractère discrètes il existe pour les variable statistiques contines deux types de représentation graphique.

Tableau statistique d'un caractère quantitative continue

Les classes	CentresC_i	L'amplitudea_i	Effectif	ECC	ECD	Fréquencef_i	FCC	FCD
$]b_1; b_2[$	C_1	a_1	n_1	N_1	$N_1^{'}$	f_1	F_1	$F_1^{'}$
$]b_2; b_3[$	C_2	a_2	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
$]b_i; b_{i+1}[$	C_i	a_i	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
$]b_k; b_{k+1}[$	C_k	a_k	n_k	N_k	$N_k^{'}$	f_k	F_k	$F_k^{'}$
Total			N			1		

TAB. 1.6 – Table statistique d'un caractère quantitative continue

La représentation en histogramme La courbe des fréquence [Meghlaoui (2011)] est la fonction en escalier dont les paliers sont constitués par les bases supérieures des rectangles formant l'histogramme des fréquence.

Courbe cumulative :[Meghlaoui (2011)] Comme pour les variables discrètes, la courbe cumulative ou histogramme des fréquences cumulées, est la représentation graphique de la fonction cumulative ou fonction de répartition $F(x)$.

Exemple 1.8.3 *[Lethielleux, Chevalier (2017)]On mesure la taille en centimetres de* 54 *élève d'une classe :*

La taille	Nombre d'étudiants n_i	Fréquences f_i	FCC
$[159, 163[$	8	0.148	0.148
$[163, 167[$	18	0.333	0.481
$[167, 171[$	11	0.204	0.685
$[171, 175[$	7	0.130	0.815
$[175, 179[$	10	0.185	1
Total	54	1	

TAB. 1.7 – Table de stasti pour mesure la taille

Tracer l'histogramme des fréquences et tracer la courbe des FCC

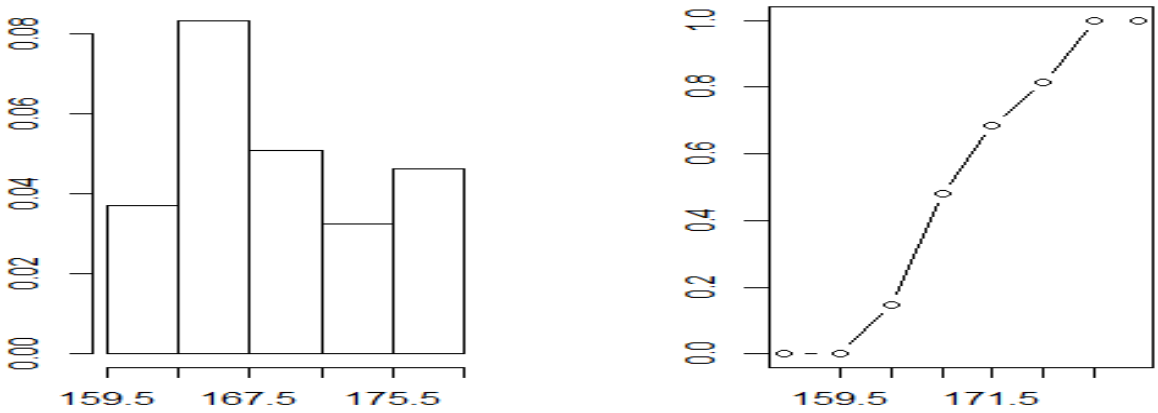

FIG. 1.3 – Histogramme des fréquences et fonction de répartition d'une variable quantitative continue

Chapitre 2

Statistique descriptive univariée

Nous prenons un ensemble de données qui ont été fournies et prenons plusieurs mesures pour analyser ces données. Nous devons connaître la raison de la collecte des données et quelles échelles de mesure.

Il nous faut à présent traiter cet ensemble de données. Tout naturellment, cela commence par les organiser, les regarder, les représenter graphiquement, regrouper celles qui se ressemblent, élaborer les moyens de rassembler l'information sous une forme aisée à manipuler et à communiquer ... bref, faire appel aux outils et méthodes de la statistique univariée l'étude d'une seule variable, que celle-ci soit quantitative ou qualitative. La statistique univariée fait partie de la statistique descriptive. [Université (2010)].

2.1 Mesures descriptives dans les statistiques

Il existe trois mesures descriptives [Monino et al (2010)] :

–Mesures de tendance centrale

–Mesures de dispresion

–Mesures de forme

2.1.1 Les caractéristiques de tendance centrale

Les caractéristiques de tendance centrale [Bahouayila (2016)] ou «mesures de tendance centrale» : Les données ont généralement tendance à être centrées autour d'une valeur spécifique qui peut êtr appelée la valeur centrale. Dans ce cas, les échelles sont utilisées pour reconnaître cette valeur centrale de la représentation des données. Parmi les mesures les plus importantes de la tendance centrale :

Le mode

Définition 2.1.1 *[Hamdani (1988)]On définit le mode comme étant la valeur de la variable statistique à laquelle correspond le plus grand effectif (ou fréquence) de la distribution statistique. On l'appelle encore valeur dominante est la valeur la plus représentée d'une variable quelconque dans une population donnée est noté* M_o.

a) Cas d'une caractère statistique quantitative discrète La valeur modale est exacte lorsque la variable statistique est discrète [Hamdani (1988)].

Exemple 2.1.1 *[Fabrice (2006)] Calcul du mode.*

Série statistique simple

Soit la série de chiffres $\{8, 5, 9, 13, 25\}$

Il n'y a pas de mode car chaque valeur n'est répétée qu'une fois.

Série statistique à valeurs répétitives

Soit la série de chiffres $\{8, 8, 8, 7, 4, 4, 4, 4, 4, 5, 5, 5, 6\}$:

La valeur la plus fréquente est le 4.

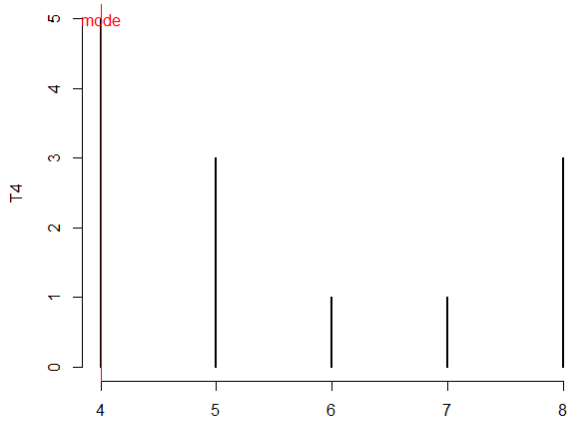

FIG. 2.1 – Mode d'une caractère statistique quantitative discrète

b) Cas d'une caractère statistique quantitative continue :[Hamdani (1988)] On parle dans ce cas d'ne classe modale, elle correspond au maximun de la fréquence moyenne par unité d'amplitude[1].

Exemple 2.1.2 *[Fabrice (2006)] :Le tableau suivant représente la distribution despoints de mathématiques pour* 30 *éléves.*

Effectifs groupés par classes d'amplitudes égales

Note (x_i)	d'amplitude a_i	Nombre d'éléves (n_i)	Effective croissant $N(x_i)$
$[0-5[$	5	2	2
$[5-10[$	5	7	9
$[10-15[$	5	18	27
$[15-20[$	5	3	30

TAB. 2.1 – Table Notes d'examen de mathématiques par classes d'amplitudes égales

Dans ce cas, pour calculer le mode , il faut appliquer la formule suivante :

$$M_o = b_{i-1} + (b_i - b_{i-1}) * \frac{(n_i - n_{i-1})}{(n_i - n_{i-1}) + (n_i - n_{i+1})} \tag{2.1}$$

[1]Amplitude de classe : $a_i = (b_i - b_{i-1})$
b_{i-1} : Borné inférieure de la classe modale.

Alors :

$$M_o = 10 + 5 * \frac{(18-7)}{(18-7)+(18-3)} = 12.115$$

Effectifs groupés par classes d'amplitudes inégales

Note (x_i)	Nombre d'éléves n_i	d'amplitude a_i	$h_i = \frac{n_i}{a_i}$
$[0-10[$	9	10	0.9
$[10-12[$	17	2	8.5
$[12-20[$	4	8	0.5

TAB. 2.2 – Table Notes d'examen de mathématiques par classes d'amplitudes inégales

Dans ce cas, pour calculer le mode, il faut appliquer la formule(1.5), mais la définition de $(n_i - n_{i-1})$ et de $(n_i - n_{i+1})$ change, car il faut remplacer les effectifs n_i, par les amplitudes corrigées $h_i = \frac{n_i}{a_i}$. Donc

$$M_o = 10 + (12-10) * \frac{(8.5-0.9)}{(8.5-0.9)+(8.5-0.5)} = 10.974$$

Remarque 2.1.1 *[Hammed (2012)] Une distribution peut avoir un seul mode et on dit qu'elle est unimodale, ou plusieurs modes et on dit qu'elle est multimodale.*

La médiane

Définition 2.1.2 *[Hamdani (1988)] La médiane est définie comme la valeur de la variable statistique qui divise l'effectif total en deux effectifs égaux, est noté Me.*

a) Cas d'une caractère statistique quantitatif discrète :

Exemple 2.1.3 *[Fabrice (2006)] Calcul la médiane.*

Série statistique simple

1-Soit la série impair de chiffres suivants :$\{8, 4, 5, 13, 11, 25, 9\}$.

–Classer la série impair par ordre croissante de valeurs $\{4, 5, 8, 9, 11, 13, 25\}$

–Localiser la valeur qui partage l'effectif tatal en deux sous effectifs égaux en appliquant la formule

$$Me = x_{(\frac{N+1}{2})} \tag{2.2}$$

C'est-à-dire ici $Me = 9$.

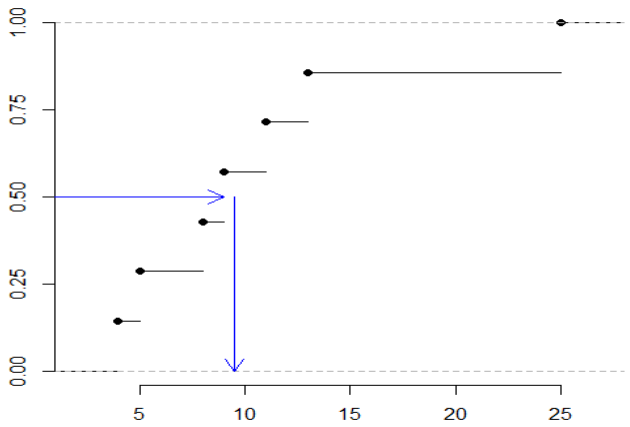

FIG. 2.2 – Médiane quand N est impair

2-Soit la série pair de chiffres des valeurs $\{1, 2, 3, 4, 6, 8, 12, 15, 25, 30, 36, 41\}$

Appliquer la formule $\frac{N+1}{2}$,c'est-à-dire ici $\frac{(12+1)}{2} = 6.5$.Ceci nous indique que l'intervalle médian est constitué par les $6^{ème}$ et la $7^{ème}$ valeurs. La médiane est donc

$$Me = \frac{1}{2} * \left\{ x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)} \right\} \tag{2.3}$$

$$Me = \frac{(8+12)}{2} = 10$$

Série statistique à valeurs répétitives

Pour déterminer la médiane, on repère 0.5 dans la colonne des fréquences cumulées $F(x)$ ou bien $\frac{N}{2}$ dans la colonne des effectifs cumulés $N(x)$. Donc la médiane égale 11.

x_i	n_i	f_i	$F(x)$	$N(x)$
2	2	0.066	0.066	2
8	3	0.1	0.166	5
9	4	0.133	0.3	9
10	4	0.133	0.433	13
11	5	0.167	0.6	18
12	3	0.1	0.7	21
13	6	0.2	0.9	27
15	1	0.033	0.933	28
18	2	0.067	1	30

TAB. 2.3 – Table Calcul de la médiane quand les données sont groupées par valeurs.

x_i	n_i	$N(x_i)$
$[0-5[$	2	2
$[5-10[$	7	9
$[10-15[$	18	27
$[15-20[$	3	30

TAB. 2.4 – Table Calcul de la médiane par classes

b) Cas d'une caractère statistique quantitative continue Dans ce cas , le calcul de la médiane nécessite d'appliquer la formule suivante :

$$Me = b_i + a_i \left[\frac{\frac{N}{2} - N(x_{i-1})}{n_i} \right] \tag{2.4}$$

[2]

Exemple 2.1.4 *[Fabrice (2006)] Calcul de la médiane*

La médiane

$Me = 10 + 5 * \left[\frac{15-9}{18}\right] = 11.67$

Remarque 2.1.2 *[Hammed (2012)] Le calcul de la médiane est basé sur l'ordre des observations et non sur leur valeur. la médiane est insensible aux données extrêmes. Dans le cas où les données sont très différentes, la médiane est une meilleure mesure de tendance centrale.*

[2] b_{i-1} = borne inférieure de la classe médiane.
$N(x_{i-1})$ = Effectif cumulé strictement inférieur à x_i.
x_i = Classe médiane .
a_i = Amplitude de la classe médiane.

La moyenne arithmétique

Définition 2.1.3 *[Hamdani (1988)] Une moyenne arithmétique d'une variable statistique X se définit comme étant le rapport de la somme des valeurs prises par cette variable,divisée par le nombre d'observations [Grais (1991)].*

La moyenne arithmétique simple Est dit simple lorsque chaque modalité x_i n'apparissent qu'une seul fois

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} x_i \tag{2.5}$$

La moyenne arithmétique pondérée Soit une variable statistique pouvant prendre les $x_1, ..., x_k$ aux quelles correspondent respectivement les effectifs $n_1, ..., n_k$.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i \tag{2.6}$$

On associe à chaque n_i la fréquence $f_i = \frac{n_i}{N}$. Donc la moyenne arithmétique égale

$$\overline{X} = \sum_{i=1}^{k} f_i x_i \tag{2.7}$$

Cas d'une caractère statistique quantitatif discréte

Exemple 2.1.5 *[Grais (1991)] Calculons le nombre moyenne de ventes par jour d'ouverture d'un magasin suivant le nombre d'un appareil A.*

Pour calculer la moyenne arithmétique, il est toujours possible d'utiliser la formule

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i \tag{2.8}$$

$$\overline{X} = \frac{692}{240} = 2.883$$

Jours x_i	Nomber de ventes n_i	$n_i * x_i$
1	53	53
2	65	130
3	50	150
4	43	172
5	7	35
6	2	12
7	20	140
Total	240	692

TAB. 2.5 – Table Le nombre de ventes par jour d'ouverture d'un appareil A.

Cas d'une caractère quantitatif statistique continue :

Exemple 2.1.6 *[Grais (1991)] Considérons la distribution des ouvriers selon le salaire mensuel net.*

Les saclaire x_i	Centre de classe C_i	Nomber de travailleurs n_i
800 à moins de 1000	900	20
1000 à moins de 1200	1100	30
1200 à moins de 1400	1300	20
1400 à moins de 1600	1500	7
1600 à moins de 1800	1700	70
Total		147

TAB. 2.6 – Table distribution des ouvriers selon le salaire mensuel net

Calculons la moyenne en utilisant la formule

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{k} n_i c_i \tag{2.9}$$

[3]

$$\overline{X} = \frac{1}{147} * (206500) = 1404.76$$

Propriété 2.1.1 *[Grais (1991)] La somme algébrique des écarts des observations à la moyenne est nule.*

[3] Centres des classes : $C_i = \frac{b_i + b_{i-1}}{2}$

$$\sum_{i=1}^{k} n_i(x_i - \overline{X}) = 0 \tag{2.10}$$

Preuve. [Grais (1991)] ■

$$\sum_{i=1}^{k} n_i(x_i - \overline{X}) = n_1x_1 + ... + n_kx_k - \overline{X}(n_1 + ... + n_k),$$

c'est-à-dire :

$$\sum_{i=1}^{k} n_i(x_i - \overline{X}) = \sum_{i=1}^{k} n_ix_i - \overline{X}\sum_{i=1}^{k} n_i$$

Or :

$$\sum_{i=1}^{k} n_i = N$$

et

$$\sum_{i=1}^{k} n_ix_i = N\overline{X},$$

par définition de la moyenne :

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{k} n_ix_i$$

d'où :

$$\sum_{i=1}^{k} n_i(x_i - \overline{X}) = N\overline{X} - N\overline{X} = 0$$

Généralisation de la notion de moyenne

La moyenne géométrique G :[Monino et al (2010)]De la distribution (x_i, n_i) est obtenue à partir de la moyenne arithmétique de la distribution (y_i, n_i), en posant $y_i = \log x_i$ [4] avec x_i positif.

$$\log G = \frac{n_1 \log x_1 + ... + n_i \log x_i + ... + n_k \log x_k}{n_1 + ... + n_i + ... + n_k} = \frac{1}{N}\sum_{i=1}^{k} n_i \log x_i = \sum_{i=1}^{k} f_i \log x_i$$

[4]log : Logarithme népérien.

qui s'écrit également :

$$G = \sqrt[N]{x_1^{n_1} \cdots x_i^{n_i} \cdots x_k^{n_k}} = \prod_{i=1}^{k} x_i^{f_i} o\grave{u} f_i = \frac{n_i}{N} \tag{2.11}$$

Propriété 2.1.2 *[Grais (1991)]*

1) Formons les produits : $z_i = x_i * y_i \rightarrow G(z_i) = G(x_i) * G(y_i)$.

2) Formons les rapports : $q_i = \frac{x_i}{y_i} \rightarrow G(q_i) = \frac{G(x_i)}{G(y_i)}$.

Preuve. [Grais (1991)] ■

1) $G(z_i) = \sqrt[n]{z_1.z_2...z_n} = \sqrt[n]{x_1 x_2...x_n}\sqrt[n]{y_1 y_2...y_n} = G(x_i)G(y_i)$.

2) $G(q_i) = \sqrt[n]{q_1 q_2...q_n} = \frac{\sqrt[n]{x_1 x_2...x_n}}{\sqrt[n]{y_1 y_2...y_n}} = \frac{G(x_i)}{G(y_i)}$.

La moyenne harmonique H :[Monino et al (2010)]De la distribution (x_i, n_i) est obtenue à partir de la moyenne arithmétique de la distribution (y_i, n_i), on posant $y_i = \frac{1}{x_i}$ avec x_i non nul.

$$\frac{1}{H} = \frac{1}{N}\sum_{i=1}^{k} \frac{n_i}{x_i} = \sum_{i=1}^{k} f_i \frac{1}{x_i}$$

qui s'écrit également :

$$H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}} \tag{2.12}$$

La moyenne quadratique Q :[Monino et al (2010)]De la distribution (x_i, n_i) est obtenue à partir de la moyenne arithmétique de la distribution (y_i, n_i), en posant $y_i = x_i^2$.

$Q^2 = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 = \sum_{i=1}^{k} f_i x_i^2$

qui s'écrit également :

$$Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i x_i^2} = \sqrt{\sum_{i=1}^{k} f_i x_i^2} \tag{2.13}$$

Propriété 2.1.3 *[Grais (1991)]Il existe une relation d'ordre entre les moyennes :*

1)

$$x_{\min} < H < G < \overline{X} < Q < x_{\max}$$

[5]

2) Si $x_1 = x_2 = ... = x_i = ... = x_k = a$, alors $H = G = \overline{X} = Q = a$.

Preuve. Vérifions cette propriété : ■

Exemple 2.1.7 *Calculons la moyenne géométrique,La moyenne harmonique et La moyenne quadratique des nombres :* $2, 3, 5, 7, 13$.

$$G = \sqrt[5]{2 * 3 * 7 * 13 * 5} = \sqrt[5]{2730} = 4.867$$

$$H = \frac{5}{\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{13}} = 3.990$$

$$Q = \sqrt{\frac{1}{5}(2 * 2 + 3 * 3 + 5 * 5 + 7 * 7 + 13 * 13)} = 7.155$$

$$\overline{X} = \frac{1}{5}(2 + 3 + 5 + 7 + 13) = 6.000$$

$$x_{\min} = 2; x_{\max} = 13$$

Alors $2.000 < 3.990 < 4.867 < 6.000 < 7.155 < 13.000$

[5] min : Munimun d'une fonction $f(.)$.
max : Maximun d'une fonction $f(.)$.

2.1.2 Les caractéristiques de dispersion

Les caractéristique de la dispersion sont nombreuses [Fabrice (2006)], nous étudierons ici les plus fréquemment utilisées : La variance, l'écart type, le coefficient de varition, ...

Nous verrons également deux outils graphiques utiles pour l'analyse de la dispersion d'une distribution : Le graphique "boîte à moustaches", ainsi que la courbe de concentration.

L'étendue

Définition 2.1.4 *[Meghlaoui (2011)] L'étendue d'une distribution statistique, notée E, est la différence entre la plus grande et la plus petite des valeurs observées, ie.*

$$E = x_{\max} - x_{\min} \tag{2.14}$$

Remarque 2.1.3 *[Meghlaoui (2011)] La forme de la distribution entre les extrèrnes n'est pas prise en compte. Donc, l'étendue est une caratéristique de dispersion imparfaite.*

Les quantiles

Les quantiles sont généralisation de la notion de la médiane, qui représente un cas particulier.

Définition 2.1.5 *[Meghlaoui (2011)] Le quantile d'ordre α $(0 \leq \alpha \leq 1)$, noté Q_α est la solution de l'équation $F(x) = \alpha$. Ainsi, en désignant par F^{-1} la fonction inverse[6] de la fonction F on a alors :*

$$Q_\alpha = F^{-1}(\alpha) \tag{2.15}$$

Il existe quatre type de quantiles :[Hamdani (1988)]

On utilise souvent les quantiles d'ordre $\frac{1}{4}$ ou 25% et d'ordre $\frac{3}{4}$ ou 75%, ces quantiles sont appellés quartiles et notés Q_1 et Q_3, la médiane est quartile d'ordre $\frac{1}{2}$ ou 50% notée Q_2.

1.Les quartiles $(Q_1, ..., Q_4)$ divisent la population statistique en 4 effectifs égaux.

[6]Fonction inverse : On appelle fonction inverse la fonction définie pour tout réel non nul par $f(x) = \frac{1}{x}$.

2.Les déciles $(D_1, ..., D_9)$ on divise la population total en 10 effectifs égaux.

3.Les centiles $(C_1, ..., C_{99})$ on divise la population total en 100 effectifs égaux.

4.Les milliles $(M_1, ..., M_{999})$ on divise la population statistique en 1000 effectifs égaux.

Les intervalles interquantiles

Définition 2.1.6 *[Hamdani (1988)] On appelle intervalle interquantile, la différeence entre le dernier et du premier quantile calculé.*

– **Intervalle interquartile** : $IQ = (Q_3 - Q_1) = x_{\frac{3}{4}} - x_{\frac{1}{4}}$ où Q_1et Q_3désigne le premier et le troisiéme quartile. Cet indice fournit un renseignement sur l'étalement des valeurs de part et d'autre de la médiane. Contenant 50% d'observations.

–**Intervalle interdécile :** $ID = (D_9 - D_1) = x_{\frac{9}{10}} - x_{\frac{1}{10}}$ où D_1et D_9 désigne le premier et le neuvième décile. Cet indice fournit un renseignement sur l'étalement des valeurs de part et d'autre de la médiane. Contenant 80% d'observations.

–**Intervalle intercentile :** $IC = (C_{99} - C_1) = x_{\frac{99}{100}} - x_{\frac{1}{100}}$. Contenant 98% d'observations.

–**Intervalle intermillile :** $IM = (M_{999} - M_1) = x_{\frac{999}{1000}} - x_{\frac{1}{1000}}$. Contenant 99.8% d'observations.

Diagramme en boîte (ou boîte à moustaches)

Il s'agit d'un diagramme permettant de positionner les quartiles Q_1, Q_2, Q_3, au moyen de rectangles de largeur arbitraire, prolongés par des "moustaches" de part et d'autre, de longueur au plus égale à une fois et demie $Q_3 - Q_1$.

Remarque 2.1.4 *[Meghlaoui (2011)] Ces diagrammes sont surtout utiles pour comparer rapidement l'allure générale de plusieurs distributions.*

La variance

Définition 2.1.7 *[Fabrice (2006)] Soit une série de valeurs d'une variable* $X : (x_1, x_2, ..., x_k)$. *Soit les effectifs associés :* $\{n_1, n_2, .., n_k\}$ *. La variance de cette série s'écrit :*

Si l'effectif considéré est celui d'une population

$$\sigma_X^2 = \frac{1}{N} \sum_{i=1}^{k} n_i (x_i - \overline{X})^2 \tag{2.16}$$

Si l'effectif considéré est celui d'une échantillon

$$\sigma_X^2 = \frac{1}{N-1} \sum_{i=1}^{k} n_i (x_i - \overline{X})^2 \tag{2.17}$$

Remarque 2.1.5 *[Hamdani (1988)]*

1.Variance est toujours positive ou nulle

$$V(x_i) = \sigma^2(x_i) \geq 0 \tag{2.18}$$

2.La Variance d'une constante est nulle.[Goldfarb, Catherine (2011)]

Propriété 2.1.4 *[Chekroun (2018)] (Théorème de König-Huygens)*

Soit (x_i, n_i) une série statistique de moyenne $\overline{X}$ et de variance $V(x)$. Alors,

$$V(X) = \sum_{i=1}^{K} f_i x_i^2 - \overline{X}^2 \tag{2.19}$$

Preuve. [Chekroun (2018)] Par définition (2.16), nous avons ■

$$\begin{aligned}\sigma_X^2 = V(X) &= \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \overline{X})^2 \\ &= \frac{\sum_{i=1}^{k} n_i(x_i - \overline{X})^2}{N} \\ &= \frac{\sum_{i=1}^{k} n_i(x_i^2 + \overline{X}^2 - 2x_i\overline{X})}{N} \\ &= \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 + \frac{1}{N}\sum_{i=1}^{k} n_i\overline{X}^2 - \frac{1}{N}\sum_{i=1}^{k} 2n_i x_i\overline{X}\end{aligned}$$

$$= \sum_{i=1}^{k} \frac{n_i}{N}x_i^2 + \overline{X}^2 - 2\overline{X}\frac{1}{N}\sum_{i=1}^{k} x_i$$

$$= \sum_{i=1}^{k} f_i x_i^2 - \overline{X}^2$$

Propriété 2.1.5 *[Leboucher, Marie (2013)] Transformation linéaire :*[7]

$$V(aX + b) = a^2 V(X) \tag{2.20}$$

Preuve. [Leboucher, Marie (2013)] Soit $Y = aX + b$ où a, b sont des nombres réels quelques.

On a $\overline{Y} = a\overline{X} + b$ et $y_i = ax_i + b$ pour tout $i = \overrightarrow{1,k}$

[7] f application de E dans K; f linaire si :
$\forall (x, y) \in E^2, \forall \beta \in k$;
$f(\beta x + y) = \beta f(x) + f(y)$

$$\begin{aligned} V(Y) &= \frac{1}{N}\sum_{i=1}^{k} n_i(y_i - \overline{Y})^2 \\ &= \frac{1}{N}\sum_{i=1}^{k} n_i(ax_i + b - a\overline{X} - b)^2 \\ &= \frac{1}{N}\sum_{i=1}^{k} n_i a^2(x_i - \overline{X})^2 \\ &= a^2\frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \overline{X})^2 \\ &= a^2 V(X). \end{aligned}$$

■

L'écart-type

Définition 2.1.8 *[Goldfarb, Catherine (2011)]L'écart-type σ_x d'une variable statistique X est la mesure de dispersion la plus couramment utilisée. Algébriquement, il se définit comme la racine carrée de la variance.*

$$\sigma_X = \sqrt{V(X)} \tag{2.21}$$

Propriété 2.1.6 *[Goldfarb, Catherine (2011)]*

$$\sigma(aX + b) = |a|\,\sigma(X) \tag{2.22}$$

Remarque 2.1.6 *[Goldfarb, Catherine (2011)]Dans le cas d'une variable statistique continue, on ramène la valeur de chaque individu au milieu de sa classe d'affectation. Là encore, le choix des bornes des classes extrêmes non limitées doit être fait avec précaution.*

Le cofficient de variation

Définition 2.1.9 *[Grais (1991)]Le coefficient de variation est défini comme le rapport de l'écart-type à la moyenne :*

$$CV = \frac{\sigma_X}{\overline{X}} \tag{2.23}$$

Exemple 2.1.8 *Lors d'une contrôle de connaissances, on fait subir à étudiants un test noté sur* 60 *points. La série des notes obtenues est la suivante :*

30 45 45 20 40 25 34 50

20 25 30 34 40 45 45 50

Calculer statistiques suivants :

–L'étendue : $E = x_{\max} - x_{\min} = 30$

–Les quartiles

$x_{\min} = 20$, $Q_1 = x_{\frac{1}{4}} = 28.75$, $mediane = Q_2 = x_{\frac{1}{2}} = 37$, $Q_3 = x_{\frac{3}{4}} = 45$, $x_{\max} = 50$

–Intervalle interquartile : $IQ = (Q_3 - Q_1) = x_{\frac{3}{4}} - x_{\frac{1}{4}} = 45 - 28.75 = 16.25$

–La variance : $\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \overline{X})^2 = 113$

–L'écart-type : $\sigma_X = \sqrt{V(X)} = 10.63$

–Le cofficient de variation : $CV = \frac{\sigma_X}{\overline{X}} = 0.29$

–Diagramme en boîte à moustaches

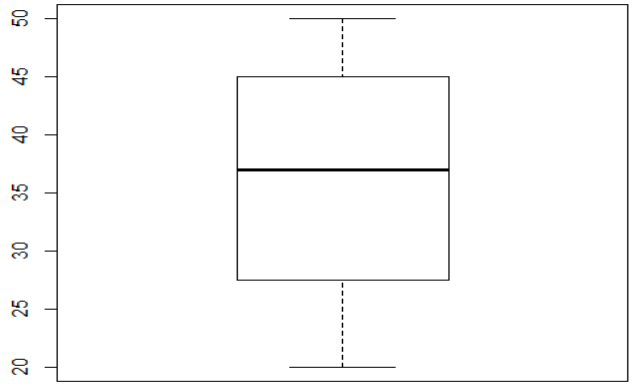

FIG. 2.3 – Boîtes à moustaches

2.1.3 Mesures de forme :

Comme nous l'avons vu précédemment, la tendance centrale et à dispersion nous aide à connaissance la diffusion des données sur la valeur centrale. Nous examinerons les mesures qui nous aident à connaître la centraliser et la forme de la distribution statistique sans recourir à la représentation des données sont appelées mesures de forme ce qui dépend de son calcul les moments.

Les différents indicateurs d'asymétrie et d'aplatissement permettent en premier lieu la comparaison entre les distributions statistiques :

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

2. L'aplatissement peut être approché par l'étude des observations aux alentours du mode : plus le nombre d'individus ayant une valeur proche du mode de la distribution est élevé, plus la courbe sera concentrée et plus l'aplatissement sera faible.[Monino et al (2010)]

Les moments non centrés et les moments centrés d'ordre p. **[Monino et al (2010)]**

Les moments non centrés d'ordre p Soit la distribution statistique (x_i, n_i) où $i \in \{1, ..., k\}$.On appelle moment non centré d'ordre p de la variable statistique X, la quantité définie par :

$$m_p = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^p = \sum_{i=1}^{k} f_i x_i^p \tag{2.24}$$

Les moment centrés d'ordre p Soit la distribution statistique (x_i, n_i) où $i \in \{1, ..., k\}$.On appelle moment centré (sur la moyenne arithmétique) d'ordre p de la variable statistique X, la quantité définie par :

$$\mu_p = \frac{1}{N}\sum_{i=1}^{k} n_i (x_i - \overline{X})^p = \sum_{i=1}^{k} f_i (x_i - \overline{X})^p \tag{2.25}$$

L'asymétrie

Le coefficient d'asymétrie de Pearson

$$A_P = \frac{\mu_3^2}{\mu_2^3} \tag{2.26}$$

Où m est le moment centré sur la moyenne arithmétique. Ce coefficient s'écrit d'une façon plus simple en utilisant les moments non centrés.

Si A_p est nul, alors la distribution est symétrique $\Rightarrow \overline{X} = Me = M_o$.

Si A_p est positif, alors il y a asymétrie$\Rightarrow M_o < Me < \overline{X}$.

Le coefficient d'asymétrie de Fisher C'est la racine carrée du coefficient de Pearson :

$$A_F = \sqrt{A_P} = \sqrt{\frac{\mu_3^2}{\mu_2^3}} = \frac{\mu_3}{\sigma_x^3} \tag{2.27}$$

où $\sigma_X^2 = V(X) = \mu_2$

Si $A_F = 0$, la distribution est symétrique$\Rightarrow \overline{X} = Me = M_o$.

Si $A_F \succ 0$, la disribution est étalées vers la droite$\Rightarrow M_o < Me < \overline{X}$.

Si $A_F < 0$, la disribution est étalées vers la gouche$\Rightarrow M_o \succ Me \succ \overline{X}$.

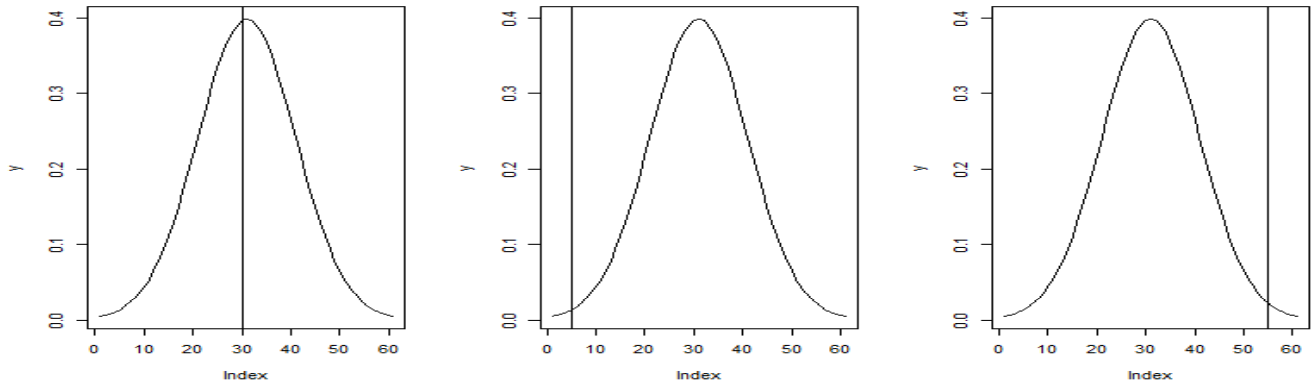

FIG. 2.4 – Asymétrie d'une distribution

L'aplatissement[Monino et al (2010)]

Le coefficient d'aplatissement de Pearson

$$AP_P = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma_x^4} \tag{2.28}$$

Le coefficient d'aplatissement de Fisher

$$AP_F = AP_P = \frac{\mu_4}{\mu_2^2} - 3 \tag{2.29}$$

Si $AP_P = 3$ (ou $AP_F = 0$), alors la courbe mésokurtique.

Si $AP_P < 3$ (ou $AP_F < 0$), alors la courbe platykurtique.

Si $AP_P \succ 3$ (ou $AP_F \succ 0$), alors la courbe leptokurtique.

2.2 Le test du Chi-deux(X^2)

2.2.1 La loi du X^2

Définition 2.2.1 *[Grammont (2003)] Soient $x_1, ..., x_k$ des variables aléatoires indépendantes de même loi normal $N(0,1)$,*[8] *On appelle loi du X^2 à k degrés de liberté, la loi de la variable aléatoire :*

$$X_k^2 = \sum_{i=1}^{k} x_i^2 \tag{2.30}$$

2.2.2 Effectif théorique

Définition 2.2.2 *[Grammont (2003)] On appelle* ***Effectif théorique*** *le produit $Np_i \geq 5$.(Ce n'est pas forcément un entier).*

2.2.3 Le test du X^2 d'ajustement :

La méthode consiste à comparer l'histogramme des fréquences et la distribution de la loi de probabilité servant de modèle théorique[Meraghni (2017)] d'employer le test lorsque certains effectifs théoriques sont inférieurs à 5. Pour cela, après avoir découpé l'intervalle d'observation en k classes, on construit un indice d^2 mesurant l'écart constaté entre les effectifs réels et les effectifs théoriques[Alalouf et al (2002)]

$$d^2 = \sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i} \tag{2.31}$$

Exemple 2.2.1 *[Meraghni (2017)] : On veut tester si un dé n'est pas truqué. Pour cela on lance le dé* 60 *fois et on obtient les résultats suivants*

[8] X suit la loi normalde paramètres m et σ^2 note $N(m, \sigma^2)$ si $X \to N(0,1)$ la loi normale centrée réduite.

Face x_i	Nombre de fois n_i	effectif théorique Np_i
1	15	$10 \geq 5$
2	7	$10 \geq 5$
3	4	$10 \geq 5$
4	11	$10 \geq 5$
5	6	$10 \geq 5$
6	17	$10 \geq 5$

TAB. 2.7 – Table Calcul la distance de khi-deux

$$d^2 = \sum_{i=1}^{k} \frac{(n_i - Np_i)^2}{Np_i} = \frac{(15-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(4-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(17-10)^2}{10} = 13.6$$

Chapitre 3

Statistique descriptive bivarié

Le chapitre précédent traitait de la **statistique descriptive univariée**, c'est-à-dire de la description d'une série statistique selon un seul caractère. On veut maintenant étudier, visualiser et mesurer (le cas échéant) les liens existant entre deux variables : C'est l'objet de la **statistique descriptive bivariée**.

Dans cette chapitre, on s'intéresse à l'étude simultanée de deux variables X et Y , étudiées sur le même population, toujours noté Ω. L'objectif essentiel des méthodes présentées est de mettre en évidence une éventuelle variation simultanée des deux variables, que nous appellerons alors liaison. Dans certains cas, cette liaison peut être considérée a priori comme causale, une variable X expliquant l'autre Y ; dans d'autres, ce n'est pas le cas, et les deux variables jouent des rôles symétriques. Dans la pratique, il conviendra de bien différencier les deux situations et une liaison n'entraîne pas nécessairement une causalité. Sont ainsi introduites les notions de covariance, coefficient de corrélation linéaire,régression linéaire, Khi-deux et autres indicateurs qui lui sont liés. De même, nous présentons les graphiques illustrant les liaisons entre variables : nuage de points ,diagrammes-boîtes parallèles [Baccini (2010)].

3.1 Série statistique double

Définition 3.1.1 *[Dagnelie (2006)] Les observation relatives à deux variables se présentent le plus simplement sous la forme d'une série statistique double, c'est-à-dire de la suite des N couples de valeurs observées (x_i, y_j), dans l'ordre croissant d'une des deux variables.*

Notations

On notera x_i, $i = 1, ..., k$ les k modalités ou valeurs de la variable X.

On notera y_j, $j = 1, ..., m$ les m modalités ou valeurs de la variable Y.

Les deux variables X et Y sont mesurées simultanément sur chacun des N individus de la population. On notera n_{ij} l'effectif correspondant au couple (x_i, y_j).

3.2 Effectif dans le cas bivarié

3.2.1 Effectifs joints

n_{ij} :Effectif joint de la modalité x_i et de la modalité y_j

$$n_{ij} := card\{\omega \in \Omega \ / X(\omega) = x_i, Y(\omega) = y_i\} \tag{3.1}$$

3.2.2 Effectifs marginaux

L'effectif marginal de la $i-$éme modalité de la variable X, $n_{i.}$,s'obtient selon la formule suivante :

$$n_{i.} = \sum_{j=1}^{m} n_{ij} \tag{3.2}$$

L'effectif marginal de la $j-$éme modalité de la variable Y, $n_{.j}$, s'obtient selon la formule suivante :

$$n_{.j} = \sum_{i=1}^{k} n_{ij} \tag{3.3}$$

Remarque 3.2.1 *[Fabrice (2006)]*

$$N = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{m} n_{.j} = n_{..} \tag{3.4}$$

3.2.3 Tableaux des effectifs

$X \setminus Y$	y_1	y_2	...	y_j	...	y_m	total
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im}	$n_{i.}$
...	...	...	...	...	...	...	...
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{km}	$n_{k.}$
total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.m}$	$N = n_{..}$

TAB. 3.1 – Table Effectif dans le cas bivarié

3.3 Fréquence dans le cas bivarié

3.3.1 fréquences jointes :

On peut, par un calcul semblable à celui réalisé dans le cas univarié, mesurer la fréquence d'un couple, en rapportant sa fréquence sur la taile de la population.[Chekroun (2018)]

$$f_{ij} = \frac{n_{ij}}{N} \tag{3.5}$$

3.3.2 Fréquences marginales

La fréquence marginale d'une modalité de X ou Y se calcule, respectivement, avec les formules suivantes, à partir de l'effectif marginal :

$$f_{i.} = \frac{n_{i.}}{N} \tag{3.6}$$

$$f_{.j} = \frac{n_{.j}}{N} \tag{3.7}$$

Remarque 3.3.1 *[Chekroun (2018)]*

$$1 = \sum_{i=1}^{k}\sum_{j=1}^{m} f_{ij} = \sum_{i=1}^{k} f_{i.} = \sum_{j=1}^{m} f_{.j} = f_{..} \tag{3.8}$$

1)

$$f_{i.} = \frac{n_{i.}}{N} i = 1, 2, ..., k \tag{3.9}$$

2)

$$f_{.j} = \frac{n_{.j}}{N} j = 1, 2, ..., m \tag{3.10}$$

$$f_{i.} = \sum_{j=1}^{m} f_{ij} = \sum_{j=1}^{m} \frac{n_{ij}}{N} = \frac{n_{i.}}{N}$$

$$f_{.j} = \sum_{i=1}^{k} f_{ij} = \sum_{i=1}^{k} \frac{n_{ij}}{N} = \frac{n_{.j}}{N}$$

3.3.3 Tableaux des fréquences :

[Fabrice (2006)]

Exemple 3.3.1 *[Fabrice (2006)]Soit le tableau de contingence suivant d'un groupe de* $N = 50$ *personnes réparties par groupe d'âge* X *et par sexe* Y*, tous âgés de* 45 *ans au plus.*

En repenant la notation du Tableau de effectif on a ici :

$n_{11} = 15$; $n_{12} = 13$; $n_{21} = 7$; $n_{22} = 15$

$n_{i.} = \sum_{j=1}^{m} n_{ij}$

$X \setminus Y$	y_1	y_2	...	y_j	...	y_m	Total
x_1	f_{11}	f_{12}	...	f_{1j}	...	f_{1m}	$f_{1.}$
x_2	f_{21}	f_{22}	...	f_{2j}	...	f_{2m}	$f_{2.}$
...	...	...	...	...	...	...	...
x_i	f_{i1}	f_{i2}	...	f_{ij}	...	f_{im}	$f_{i.}$
...	...	...	...	...	...	...	...
x_k	f_{k1}	f_{k2}	...	f_{kj}	...	f_{km}	$f_{k.}$
Total	$f_{.1}$	$f_{.2}$	...	$f_{.j}$	...	$f_{.m}$	$f_{..} = 1$

TAB. 3.2 – Table Fréquence dans le cas bivarié

$X \setminus Y$	Homme	femme
$[0 - 18[$	15	13
$[18 - 45[$	7	15

TAB. 3.3 – Table Représenter groupe de personnes réparties par groupe d'âge X et par sexe Y

$n_{1.} = n_{11} + n_{12} = 15 + 13 = 28$; $n_{2.} = n_{21} + n_{22} = 7 + 15 = 22$

$n_{.j} = \sum_{i=1}^{k} n_{ij}$

$n_{.1} = n_{11} + n_{21} = 15 + 7 = 22$; $n_{.2} = n_{12} + n_{22} = 13 + 15 = 28$

$\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{m} n_{j.} = n_{..} = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = n_{11} + n_{12} + n_{21} + n_{22} = 15 + 13 + 7 + 15 = 50$

$n_{..} = \sum_{i=1}^{k} n_{i.} = n_{1.} + n_{2.} = 28 + 22 = 50$

$n_{..} = \sum_{j=1}^{m} n_{.j} = n_{.1} + n_{.2} = 22 + 28 = 50$

Fréquences marginales de X :

$f_{1.} = \frac{n_{1.}}{n_{..}} = \frac{28}{50} = 0.56$

$f_{2.} = \frac{n_{2.}}{n_{..}} = \frac{22}{50} = 0.44$

Fréquences marginales de Y :

$f_{.1} = \frac{n_{.1}}{n_{..}} = \frac{22}{50} = 0.44$

$f_{.2} = \frac{n_{.2}}{n_{..}} = \frac{28}{50} = 0.56$

3.4 Représentations graphiques des distributions deux caractères

Le mode de représentation graphique d'une distribution deux caractères n'est strictement possible que dans un espace trois dimensions. Chacun des caractères est porté sur une dimension et la troisième est affectée aux effectifs ou aux fréquences [Grais (1991)].

3.4.1 Cas des caractères quantitatives

Une série statistique double dont les caractères X et Y sont quantitative est représentée par les points M_i de coordonnées (x_i, y_i) dans un repère orthogonal du plan. Cette représentation s'appelle nuage de points de la série statistique double.

3.4.2 Cas des caractères qualitatives

Si les deux variable X et Y sont qualitatives, alors les données observées sont une suite de couples de variable $(x_1, y_1), ..., (x_i, y_j)..., (x_N, y_N)$, il n'est pas possible, dans ce cas, de représenter les deux caractères de façon absolument symétrique.

3.5 La moyenne et la variance marginale

3.5.1 La moyenne marginale

La variable X :

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{i=k} n_{i.} x_i = \sum_{i=1}^{i=k} f_{i.} x_i \tag{3.11}$$

La variable Y :

$$\overline{Y} = \frac{1}{N}\sum_{j=1}^{i=m} n_{.j}y_j = \sum_{j=1}^{i=m} f_{.j}y_j \tag{3.12}$$

Remarque 3.5.1 *[Goldfarb, Catherine (2011)]Moyenne d'une somme de deux variables statistiques* $\overline{X+Y} = \overline{X} + \overline{Y}$

3.5.2 La variance marginale

La variable X :

$$V(X) = \sigma_X^2 = \frac{1}{N}\sum_{i=1}^{i=k} n_{i.}(x_i - \overline{X})^2 = \sum_{i=1}^{i=k} f_{i.}(x_i - \overline{X})^2 \tag{3.13}$$

La variable Y :

$$V(Y) = \sigma_Y^2 = \frac{1}{N}\sum_{j=1}^{j=m} n_{.j}(y_j - \overline{Y})^2 = \sum_{j=1}^{j=m} f_{.j}(y_j - \overline{Y})^2 \tag{3.14}$$

Propriété 3.5.1 *Dans le cas où les séries ou les variables aléatoires X et Y sont indépendantes,*[1] *on a :*

$V(X+Y) = V(X) + V(Y)$

3.6 Écart-type

On utilise les carrés des écarts et non les écarts eux-mêmes afin d'éviter une correction fallacieuse entre des écarts positifs et négatifs. L'usage de l'écart moyen arithmétique est très rarement utilisé car peu opérationnel de par les valeurs absolues et son absence de propriétés additives en présence de variables indépendantes.

L'écart-type ou l'écart quadratique moyen d'une série ou variable aléatoire X ou Y est la racine carrée de la variance marginale X ou Y.[Mehl (1996)]

[1] L'indépendance est une notion probabiliste qualifiant de manière intuitive des événements aléatoires n'ayant aucune influence l'un sur l'autre.

La variable X :

$$\sigma_X = \sqrt{V(X)} \tag{3.15}$$

La variable Y :

$$\sigma_Y = \sqrt{V(Y)} \tag{3.16}$$

Exemple 3.6.1 *[Fabrice (2006)] Soit le tableau de contingence suivant*

X/Y	1	4	$n_{i.}$
2	5	7	12
8	2	12	14
$n_{.j}$	7	19	26

Calculons la moyenne marginale de X :

$\overline{X} = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.} x_i = \frac{1}{26}((12 * 2) + (14 * 8)) = \frac{68}{13} = 5.2308$

Calculons la moyenne marginale de Y :

$\overline{Y} = \frac{1}{n_{..}} \sum_{j=1}^{m} n_{.j} y_j = \frac{1}{26}((7 * 1) + (19 * 4)) = \frac{83}{26} = 3.1923$

Calculons le variance marginale de X :

x_i	$n_{i.}$	x_i^2	$n_{i.}x_i^2$
2	12	4	48
8	14	64	896

$\sigma_X^2 = \frac{1}{N} \sum_{i=1}^{i=k} n_{i.}(x_i - \overline{X})^2 = \frac{1}{N} \sum_{i=1}^{i=k} n_{i.} x_i^2 - (\overline{X})^2 = \frac{1}{26}(48 + 896) - (5.2308)^2 = 8.9464$

$\sigma_X = \sqrt{8.9464} = 2.9911$

Calculons le variance marginale de Y :

y_i	$n_{.j}$	y_j^2	$n_{.j}y_j^2$
1	7	1	7
4	19	16	304

$\sigma_Y^2 = \frac{1}{N}\sum_{j=1}^{j=m} n_{.j}(y_j - \overline{Y})^2 = \frac{1}{N}\sum_{j=1}^{j=m} n_{.j}y_j^2 - (\overline{Y})^2 = \frac{1}{26}(7+304) - (3.1923)^2 = 1.7708$

$\sigma_Y = \sqrt{1.7708} = 1.330\,7$

3.7 Covariance

Définition 3.7.1 *[Meghlaoui (2011)]Soit (X, Y) un couple de variables statistiques pouvant prendre les valeurs (x_i, y_j),$i = 1, 2, ..., k$ et $j = 1, 2, ..., m$ avec les effectifs respectifs (n_{ij}), $i = 1, 2..., k$ et $j = 1, 2, ..., m$. On appelle covariance des variables statistiques X et Y notée $Cov(X, Y)$, la quantité définie telle que :*

$$cov(X,Y) = \frac{1}{N}\sum_{i=i}^{i=k}\sum_{j=1}^{j=m} n_{ij}(x_i - \overline{X})(y_j - \overline{Y}) \tag{3.17}$$

Remarque 3.7.1 *[Meghlaoui (2011)]Pour le calcul pratique, on utilisera souvent la formule développée de la covariance définie telle que*

$$cov(X,Y) = \frac{1}{N}\sum_{i=1}^{i=k}\sum_{j=1}^{j=m} n_{ij}x_iy_j - \overline{XY} = \sum_{i=1}^{i=k}\sum_{j=1}^{j=m} f_{ij}x_iy_j - \overline{XY} \tag{3.18}$$

Dans certaines situations il arrive que que les observations d'une population suivant deux caractères(X, Y) soient appariées, i.e. les observations sont disponibles sous forme d'une suite $(x_i, y_i), i = 1, 2, ..., N$, alors dans cette situation la covariance est définie telle que :

$$cov(X,Y) = \frac{1}{N}\sum_{i=i}^{i=N}(x_i - \overline{X})(y_i - \overline{Y}) = \frac{1}{N}\sum_{i=1}^{i=N} x_iy_i - \overline{XY} \tag{3.19}$$

Propriété 3.7.1 *[Goldfarb, Catherine (2011)]*

$cov(X,Y) = cov(Y,X)$

$cov(X,X) = V(X)$

$V(X+Y) = V(X) + V(Y) + 2cov(X,Y)$

Preuve. [Goldfarb, Catherine (2011)] ■

$cov(X,Y) = \frac{1}{N}\sum_{i=i}^{i=k}\sum_{j=1}^{j=m} n_{ij}(x_i - \overline{X})(y_j - \overline{Y}) = \frac{1}{N}\sum_{i=i}^{i=k}\sum_{j=1}^{j=m} n_{ij}(y_j - \overline{Y})(x_i - \overline{X}) = cov(Y,X)$

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

$V(X+Y) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i + y_i - \overline{(X+Y)})^2 = \frac{1}{N}\sum_{i=1}^{i=N}(x_i + y_i - \overline{X} - \overline{Y})^2 = \frac{1}{N}\sum_{i=1}^{i=N}(x_i - \overline{X} + y_i - \overline{Y})^2$

$= \frac{1}{N}\left[\sum_{i=1}^{i=N}(x_i - \overline{X})^2 + \sum_{i=1}^{i=N}(y_i - \overline{Y})^2 + 2\sum_{i=1}^{i=N}(x_i - \overline{X})(y_i - \overline{Y})\right]$

$= \frac{1}{N}\sum_{i=1}^{i=N}(x_i - \overline{X})^2 + \frac{1}{N}\sum_{i=1}^{i=N}(y_i - \overline{Y})^2 + 2\frac{1}{N}\sum_{i=1}^{i=N}(x_i - \overline{X})(y_i - \overline{Y}) = V(X) + V(Y) + 2cov(X,Y)$

3.8 Deux variables qualitatives

On considère une population sur laquelle on étudie deux variables qualitatives observées simultanément sur N individus [Vessereau (1965)].

On peut alors calculer les critères classiques du khi-deux X^2 de Pearson , ou encore le coefficient de corrélation.

3.8.1 Coefficient de corrélation

Lorsque les séries sont qualitatives, il arrive que les modalités d'un des deux caractères soient ordinales(voir le chapitre1) [Fabrice (2006)], autrement dit que l'on puisse opérer un classement sur ces modalités. Dans ce cas, au lieu de calculer la corrélation entre les valeurs comme on le fait pour une variable, on calcule la corrélation entre les rangs des modalités. On calcule alors un coefficient appelé **coefficient de corrélation de rang de SPEARMAN.**

La formule :

$$r_{sp} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{3.20}$$

où d_i est la différence entre les rangs des valeurs correspondantes de X et de Y et N le nombre d'observations.

3.8.2 Le test du khi-deux de PEARSONS

Définition 3.8.1 *[Fabrice (2006)] Lorsque les caractéres sont qualitatives l'étude de la corrélation se fait par un test statistique développé par* ***Karl PEARSONS*** *et appelé test d'indépendance du "Khi-deux".*

3.8.3 test d'indépendance du "Khi-deux"

On considère ici un couple(X, Y) de variables aléatoires. On suppose que X $(resp.Y)$ prend ses valeurs dans l'ensemble $\{1, ..., k\}$ $(resp.\{1, ..., m\})$. Si $p_{ij} = P(X = i, Y = j)$, on représentera la loi du couple(X, Y).

Le problème qui nous intéresse dans ce paragraphe est de tester l'indépendance des variables X et Y.

Les variables X et Y sont indépendantes si et seulement si la loi P est le produit (tensoriel) de ses lois marginales i.e.

[Meraghni (2017)]

$$\forall i = 1, ..., k; \forall j = 1, ..., m, P_{ij} = P_{i.}P_{.j} \tag{3.21}$$

3.8.4 Effectif théorique deux variable

Définition 3.8.2 *[Grammont (2003)] On appelle* ***Effectif théorique*** *la quantité* $u_{ij} = \frac{n_{i.} * n_{.j}}{N}$.

Effectif théorique= (total de la ligne $\times$ total de la colonne)/N.

On définit la quantité

$X \diagdown Y$	Homme	femme	$n_{i.}$
19ans et moins	30	20	50
De 20 à 24 ans	30	10	40
De 25 à 29 ans	40	50	90
30 ans et plus	10	10	20
$n_{.j}$	110	90	200

TAB. 3.4 – Table le nombre de la Grise cardiaques, subies pas de hommes et des femmes selon leur classé d'âge

$$d^2 = \sum_{j=1}^{m} \sum_{i=1}^{k} \frac{(n_{ij} - u_{ij})^2}{u_{ij}} \tag{3.22}$$

d^2 :Appelé distance de Khi-deux.

Exemple 3.8.1 *[Meraghni (2017)] Le tableau suivant donnée le nombre de Grise cardiaques, subies pas de hommes et des femmes selon leur classé d'âge pour un échantillon de* $200(N = 200)$ *personnes.*

où

X :l'age$\Rightarrow k = 4$

Y :la sexe$\Rightarrow m = 2$

par définition :

$$d^2 = \sum_{j=1}^{m} \sum_{i=1}^{k} \frac{(n_{ij} - u_{ij})^2}{u_{ij}}$$

En appliquant cette définition aux données du tableau.

$-$Calcul effectifs théoriques $u_{ij} = \frac{n_{i.} * n_{.j}}{N}$

$u_{11} = \frac{n_{1.} * n_{.1}}{N} = \frac{50 * 110}{200} = 27.5$

$u_{12} = \frac{n_{1.} * n_{.2}}{N} = \frac{40 * 90}{200} = 18$

$u_{21} = \frac{40 * 110}{200} = 22$

$u_{22} = \frac{40 * 90}{200} = 18$

$u_{31} = \frac{90 * 110}{200} = 49.5$

$u_{32} = \frac{90 * 90}{200} = 40.5$

$u_{41} = \frac{20*110}{200} = 11$

$u_{42} = \frac{20*90}{200} = 9$

effectif observé n_{ij}	30	20	30	10	40	50	10	10
effectif théorique u_{ij}	27.5	18	22	18	49.5	40.5	11	9

par définition :

$d^2 = \sum_{j=1}^{m} \sum_{i=1}^{k} \frac{(n_{ij}-u_{ij})^2}{u_{ij}}.G4$

En appliquant cette définition aux données du tableau , on obtient :

$d^2 = \frac{(30-27.5)^2}{27.5} + \frac{(20-18)^2}{18} + \frac{(30-22)^2}{22} + \frac{(10-18)^2}{18} + \frac{(40-49.5)^2}{49.5} + \frac{(50-40.5)^2}{40.5} + \frac{(10-11)^2}{11} + \frac{(10-9)^2}{9} = 11.168$

3.9 Deux variables quantitatives

On s'intéresses à une statistique ayant deux dimensions que nous désignons par les variables X et Y.On veut savoir si les deux variables sont liés par une liaison fonctionnelle du type $Y = f(X)$ (c'est-à-dire que l'on peut prévoir les valeurs de Y à partir des valeurs de X), ou $X = f(Y)$ (c'est-à-dire que l'on peut prévoir les valeurs de X à partir des valeurs de Y).

3.9.1 Coefficient de corrélation

Définition 3.9.1 *[Meghlaoui (2011)] On appelle coefficient de corrélation de deux variables statistiques X et Y et on le note $Corr(X, Y)$ ou , la quantité définie telle que :*

$$r = corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \tag{3.23}$$

Remarque 3.9.1 *[Hamdani (1988)] Le coefficient de corrélation r est, à une constante prés, le cosinus de l'angle entre les vecteurs $\overline{X}$ et $\overline{Y}$.*

On pour tout $a, b, x_0, y_0, \in \omega$

$$r(aX + x_0, bY + y_0) = \frac{cov(aX + x_0, bY + y_0)}{\sigma_{aX+x_0} * \sigma_{bY+y_0}} = \frac{ab * cov(X, Y)}{|ab|\, \sigma_X \sigma_Y} \tag{3.24}$$
$$= \left\{ \begin{array}{l} +r(X, Y) \text{ si } a \text{ et } b \text{ de méme signe} \\ -r(X, Y) \text{ si } a \text{ et } b \text{ de signe opposé} \end{array} \right\}$$

– Si $|r| = 1$, alors il existe une relation linéaire entre X et Y.

– Si $r = 0$, il y a indépendance linéaire entre X et Y (mais il peut exister une autre forme de dépendance).

– $0 < |r| < 1$ traduit une dépendance linéaire d'autant plus forte que $|r|$ est grand.

3.9.2 Droite de régression

Dans le cas où on peut mettre en évidence l'existence d'une relation linéaire significative entre deux caractères quantitatifs continus X et Y (la silhouette du nuage de points est étirée dans une direction), on peut chercher à formaliser la relation moyenne qui unit ces deux variables à l'aide d'une équation de droite l'idée est de transformer un nuage de point en une droite. Celle-ci doit être la plus proche possible de chacun des points. On cherchera donc à minimiser les écarts entre les points et la droite.

Principe des moindres carrés ordinaire (MCO)

En général, les données prennent la forme de N couples $(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)$ que l'on peut représenter par autant de points sur un plan cartésien. L'équation d'une droite est de la forme

$$Y = aX + b \tag{3.25}$$

Soit une droite donnée $Y = a + bX$, et soit $d_1, d_2, ..., d_N$ les distances verticales entre les points et la droite. Ces distances sont représentées par les traits verticaux.

La somme des carrés de ces distances servira de mesure globale de la distance entre les points et la droite. On définit formellement la distance D entre les points et la droite par :

$$D = d_1^2 + d_2^2 + ... + d_N^2 = \sum d_i^2 \tag{3.26}$$

Si l'on dénote par $\hat{y_i}$ la hauteur de la droite au point x_i, c'est-à-dire

$\hat{y}_i = ax_i + b$

alors d_i est donné par

$$d_i = |y_i - \hat{y}_i| \tag{3.27}$$

et

$$D = \sum |y_i - \hat{y}_i|^2 = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - ax_i - b)^2 = f(a, b) \tag{3.28}$$

Nous souhaitons que cette distance soit petite : plus elle est petite, mieux la droite est ajustée aux données. Puisque notre objectif est de trouver une droite qui s'ajuste le mieux possible aux données, nous devons chercher la droite pour laquelle **la distance D est minimale.**

Propriété 3.9.1 *[Meghlaoui (2011)] Soient X et Y deux variables statistiques définies sur la même population. La fonction numérique définie sur R^2 par l'équation* (3.28) *admet un minimum au point (a, b) tel que :*

$$a = \frac{cov(X, Y)}{V(X)} \tag{3.29}$$

$$b = \overline{Y} - a\overline{X} \tag{3.30}$$

Preuve. [Meghlaoui (2011)]On cherche les valeurs a et b définissant la droite de régression $Y = aX + b$ sont minimal la fonction $f(a, b)$ cette méthode s'appelle moinde carrés ordinaire.

Alors (a, b) est solution du systeme ■

$$\begin{array}{l} \frac{\partial f(a,b)}{\partial a} = 0 \\ \frac{\partial f(a,b)}{\partial b} = 0 \end{array} \tag{3.31}$$

$$f(a,b) = \sum_{i=1}^{N} (y_i - ax_i - b)^2$$

$$\begin{aligned} \frac{\partial f(a,b)}{\partial b} &= -2 \sum_{i=1}^{N} (y_i - ax_i - b) \\ &= -2 \sum_{i=1}^{N} y_i + 2a \sum_{i=1}^{N} x_i + 2b \sum_{i=1}^{N} 1 = 0 \end{aligned}$$

$$= \sum_{i=1}^{N} y_i - a \sum_{i=1}^{N} x_i - bN = \frac{1}{N} \sum_{i=1}^{N} y_i - a \frac{1}{N} \sum_{i=1}^{N} x_i - b = 0$$

alors

$$b = \overline{Y} - a\overline{X}$$

$$\begin{aligned} \frac{\partial f(a,b)}{\partial a} &= -2 \sum_{i=1}^{N} x_i (y_i - ax_i - b) \\ &= -2 \sum_{i=1}^{N} x_i y_i + 2a \sum_{i=1}^{N} x_i^2 + 2b \sum_{i=1}^{N} x_i = 0 \end{aligned}$$

$$\sum_{i=1}^{N} x_i y_i - a\sum_{i=1}^{N} x_i^2 - b\sum_{i=1}^{N} x_i = 0$$

$$\sum_{i=1}^{N} x_i y_i = a\sum_{i=1}^{N} x_i^2 + (\overline{Y} - a\overline{X})\sum_{i=1}^{N} x_i$$

$$\sum_{i=1}^{N} x_i y_i = a\sum_{i=1}^{N} x_i^2 + N\overline{YX} - Na\overline{X}^2$$

$$\sum_{i=1}^{N} x_i y_i = a(\sum_{i=1}^{N} x_i^2 - N\overline{X}^2) + N\overline{YX}$$

alors

$$a = \frac{\sum_{i=1}^{N} x_i y_i - N\overline{YX}}{\sum_{i=1}^{N} x_i^2 - N\overline{X}^2}$$

Sachant que :

Par l'équation (2.15)

$cov(X,Y) = \frac{1}{N}\sum_{i=i}^{i=N}(x_i - \overline{X})(y_i - \overline{Y}) = \frac{1}{N}\sum_{i=1}^{i=N} x_i y_i - \overline{XY} = \frac{1}{N}(\sum_{i=1}^{i=N} x_i y_i - N\overline{XY})$

$\Rightarrow N \times cov(X,Y) = \sum_{i=1}^{i=N} x_i y_i - N\overline{XY}$.....(1)

$cov(X,X) = V(X) = \frac{1}{N}\sum_{i=i}^{i=N}(x_i - \overline{X})^2 = \frac{1}{N}\sum_{i=i}^{i=N} x_i^2 - \overline{X}^2 = \frac{1}{N}(\sum_{i=i}^{i=N} x_i^2 - N\overline{X}^2)$

$\Rightarrow N \times V(X) = \sum_{i=i}^{i=N} x_i^2 - N\overline{X}^2$.....(2)

D'apré (1) et (2)

$$a = \frac{N \times cov(X,Y)}{N \times V(X)} = \frac{cov(X,Y)}{V(X)}$$

Remarque 3.9.2 *[Meghlaoui (2011)]* $a = \frac{cov(X,Y)}{V(X)} = corr(X,Y)\frac{\sigma_y}{\sigma_x}$

En effet

$a = \frac{cov(X,Y)}{\sigma_x\sigma_x}\frac{\sigma_y}{\sigma_y} = \{\frac{cov(X,Y)}{\sigma_x\sigma_y}\}\frac{\sigma_y}{\sigma_x} = corr(X,Y)\frac{\sigma_y}{\sigma_x}$

Exemple 3.9.1 *[Monino et al (2010)]On veur étudier la liaison entre la consommation et revenu des ménages, pour cela vous avez ci-dessous le tableau des données.*

		X	Y
Observation	Années	Revenu	Consommation
1	2005	238	199
2	2006	257	208
3	2007	270	221
4	2008	290	237
5	2009	303	254
6	2010	319	268
7	2011	333	280
8	2012	351	293
9	2013	369	307
10	2014	387	323
Total		3117	2590

TAB. 3.5 – Table Représenter relation entre la consommation et revenu des ménages

Tableau des observations revenu/consommation des ménages

Calculer la covariance entre X et Y.

Calculer le coefficient de corrélation linéaire entre X et Y.

Calculer le droite de régression $Y = aX + b$.

Solution 3.9.1 *:*

1) Calcul la covariance

$cov(X,Y) = \frac{1}{N}\sum_{i=i}^{i=N}(x_i - \overline{X})(y_i - \overline{Y}) = \frac{1}{N}\sum_{i=1}^{i=N} x_i y_i - \overline{XY} = \frac{1}{10} * 826039 - 311.7 * 259 = 1873.6$

2) Calcul de coefficient de corrélation linéaire

$r = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

Calcul les écarts-types

$\sigma_X = \sqrt{V(X)} = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}(x_i - \overline{X})^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2 - (\frac{\sum_{i=1}^{N} x_i}{N})^2} = \sqrt{\frac{1}{10}993283 - (311.7)^2} = 46.6 =$

$\sigma_Y = \sqrt{V(Y)} = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N}(y_i - \overline{Y})^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i^2 - (\frac{\sum_{i=1}^{N} y_i}{N})^2} = \sqrt{\frac{1}{10}687042 - (259)^2} = 40.289$

Alors

$r = \frac{1873.6}{46.598 * 40.289} = 0.998$

3) Étude droite d'ajustement $Y = aX + b$

$a = \frac{cov(X,Y)}{V(X)} = \frac{cov(X,Y)}{\sigma_X^2} = \frac{1873.6}{(46.598)^2} = 0.863$

$b = \overline{Y} - a\overline{X} = \frac{\sum_{i=1}^{N} y_i}{N} - a * \frac{\sum_{i=1}^{N} x_i}{N} = 259 - 0.863 * 311.7 = -9.997.$

$Y = 0.863X - 9.997.$

Chapitre 4

Application avec Logiciel R

4.1 Statistique descriptive univariée

Exemple 4.1.1 *Le tableau suivant représente l'évolution du virus Corona en Algérie pour le mois de Ramadan et Shawwal*

Effectifs groupés par classes d'amplitudes égal

Population : Le nombre de malades du virus en Algérie pour le mois de Ramadan et Shawwal.

Unité statistique : Une personne malade du virus Corona.

Caractère : Les caractères quantitatives continue

Echantillon : Le nombre de malades du virus en Algérie c'est en fait un nombre moins.

Jours x	Líamplitude a_i	Centres C_i	De nombre des infectés n_i
$[1,5]$	4	3	642
$[6,10]$	4	8	825
$[11,15]$	4	13	895
$[16,20]$	4	18	884
$[21,25]$	4	23	948
$[26,30]$	4	28	912
$[31,35]$	4	33	884
$[36,40]$	4	38	629
$[41,45]$	4	43	528
$[46,50]$	4	48	544
$[51,55]$	4	53	570
$[56,60]$	4	58	652
Total			8913

TAB. 4.1 – Table représente l'évolution du virus Corona en Algérie

Solution 4.1.1 *:*

1) Les caractéristiques de tendance centrale :

```
x=c(3,8,13,18,23,28,33,38,43,48,53,58)

n=c(642,825,895,884,948,912,884,629,528,544,570,652)

Y=rep(x,n)

N=sum(n)

hist(Y,xlab="x",ylab="Effectifs",main="")

segments(20,948,25,912,col=2)

segments(20,884,25,948,col=2)

arrows(x0=23.25,y0=0,x1=23.25,y1=925,col="blue")

text(23.25,0,labels="mode",col="blue")# Mode
```

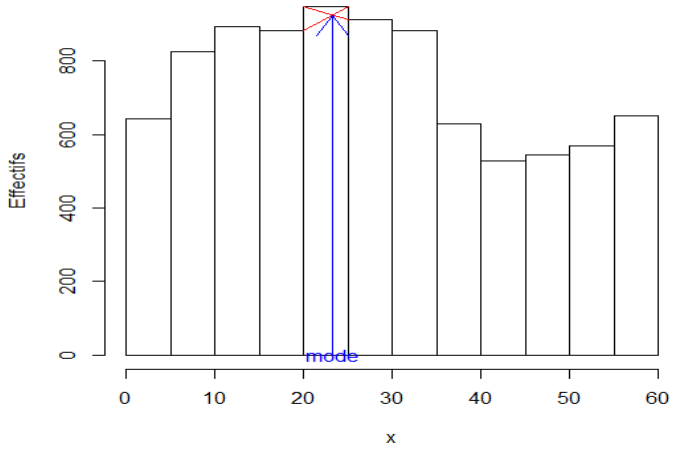

FIG. 4.1 – Histogramme des effectifs et le mode

Commentaire :

Je constate d'après le tableau et le graphique une augmentation notable du nombre des infectés pendant le mois de Ramadan, puis on remarque une diminution progressive des infectés avec le temps.

Hypothèses :

La diminuation de nombre des infectés est dû à plusieurs facteurs.

L'effet de l'augmentation température l'application de confinement.

L'orientation naturelle reconnue des épidémies haute e et baisse.

Le manque flagrant des appareils développés pour détecter le Corona virus.

En langage R

median(Y)# Le mediane

[1] 28

mean(Y)# La moyenne

[1] 28.42242

2) Les caractéristiques de dispersion :

En langage R

max(Y)# Maximum

[1] 58

min(Y)# Minimum

[1] 3

E=max(Y)-min(Y)# L'étendue

[1] 55

quantile(Y)#Les quartiles

0%	25%	50%	75%	100%
3	13	28	43	58

IQR(Y)# Intervalle interquartile

[1] 30

var(Y)# La variance

[1] 270.1509

N*var(Y)/(N-1)# Estimateur sans biais de la variance

[1] 270.1812

sqrt(N*var(Y)/(N-1))# Lécart-type s_2

[1] 16.43719

sd(Y)# L'écart-type

[1] 16.43627

CV=sd(Y)/mean(Y)# Le cofficient de variation

[1] 0.5782853

3) Mesures de forme :

En langage R

AP=(sum((Y-mean(Y))^3)/N)^2/(sum((Y-mean(Y))^2)/N)^3# Le coefficient d'asymétrie de Pearson

[1] 0.05981674

AF=sqrt(AP)# Le coefficient d'asymétrie de Fisher

[1] 0.2445746

Remarque 4.1.1 *:*

Nous remarquons que : AP$\simeq$ 0 (ou AF est nul) alors la disribution symétrique$\rightarrow M_o \prec Me \prec \overline{X}$.

En langage R

APP=(sum((Y-mean(Y))^4)/N)/(sum((Y-mean(Y))^2)/N)^2# Le coefficient d'aplatissement de Pearson

[1] 1.983556

APF=APP-3# Le coefficient d'aplatissement de Fisher

[1] -1.016444

Remarque 4.1.2 *Nous remarquons que : APP$\prec$3 (ou APF$\prec$0) alors la courbe platykurtique.*

résumé

Après la fin de ka crise l'épidémie de corona virus le covid-19, le monde sera différent de ce qu'il était auparavant, notre immunité sera plus forte ainsi que notre résistance et nous puissions trouver des nouvelles solutions innovantes et à changer notre façon de travailler et de vivre.

4.2 Statistique descriptive bivarié

4.2.1 Deux variables quantitatives

Exemple 4.2.1 *On veur étudier la liaison entre la consommation et revenu des ménages, pour cela vous avez ci-dessous le tableau des données.*

		X	Y
Observation	Années	Revenu	Consommation
1	2005	238	199
2	2006	257	208
3	2007	270	221
4	2008	290	237
5	2009	303	254
6	2010	319	268
7	2011	333	280
8	2012	351	293
9	2013	369	307
10	2014	387	323
Total		3117	2590

TAB. 4.2 – Table Représenter relation entre la consommation et revenu des ménages

Solution 4.2.1 *:*

On peut obtenir la moyenne marginale et la variance marginale et l'écart-type et covariance et coefficient de corrélation.

En langage R :

> Revenu=c(238,257,270,290,303,319,333,351,369,387)

> Consommation=c(199,208,221,237,254,268,280,293,307,323)

> N=length(Revenu)# Taille d'un population

> N=length(Consommation)

> mean(Revenu)# Moyenne marginale de X

[1] 311.7

> mean(Consommation)# Moyenne marginale de Y

[1] 259

> v1=sum((x-mean(Revenu))^2)/N# Variance marginale de X

[1] 2171.41

> var(Revenu)# Variance par échantillon

[1] 2412.678

> v2=sum((y-mean(Consommation))^2)/N# Variance marginale de Y

[1] 1623.2

> var(Consommation)# Variance cas echanti

[1] 1803.556

> s1=sqrt(v1)# L'écart-type marginale de X

[1] 46.59839

> sd(Revenu)# L'écart-type cas échantillon

[1] 49.11902

> s2=sqrt(v2)# L'écart-type marginale de Y

[1] 40.28896

> sd(Consommation)

[1] 42.46829

> c12=sum((x-mean(Revenu))*(y-mean(Consommation)))/N# Covariance de X, Y

[1] 1873.6

> r=c12/(s1*s2)# Coefficient de corrélation.

[1] 0.9979756

Remarque 4.2.1 *Nous remarquons que :* $|r| \simeq 1$ *alors il existe une relation linéaire entre* X *et* Y.

Calculer la droite d'ajustement : $Y = aX + b$

En langage R :

> a=c12/v1

[1] 0.8628495

> b=mean(Consommation)-a*mean(Revenu)

[1] -9.950184

alors $Y = 0.863X - 9.950$.

Représentation graphique du nuage de poits de la relation : $Y = aX + b$.

En langage R :

> plot(Revenu,Consommation)

> x=220 :20 :400

> lines(x,0.863*x-9.950)

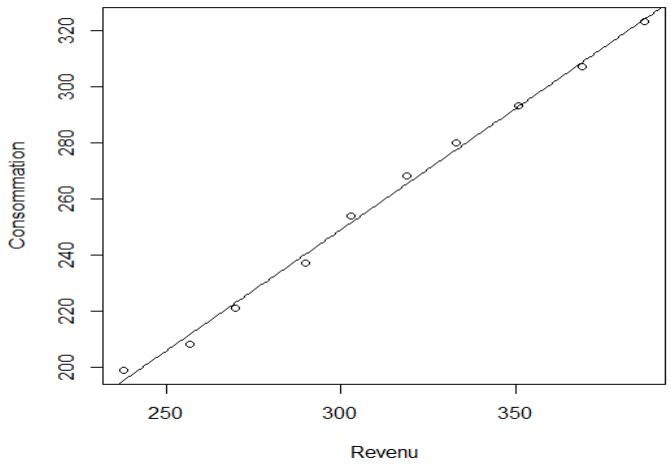

FIG. 4.2 – Le droite de régression

Conclusion

Nous sommes arrivés au terme de la thèse scientifique liée aux statistiques descriptives. J'ai essayé autant que possible de développer ce sujet. De nombreux exemples simples et clairs ont été utilisés pour transmettre des informations, des technologies modernes telles que Logicial R ont été utilisées pour résoudre le problème dans les plus brefs délais et des graphiques ont été utilisés, bien qu'ils n'aient rien à voir avec des concepts mathématiques, mais ils illustrent des concepts mathématiques dans Mindful People, en particulier ceux qui n'étudient pas Statistiques.

Mon objectif principal dans cette thèse était de déterminer l'importance des statistiques descriptives dans notre vie quotidienne.

La descriptive des statistiques nous aide à rendre beaucoup de choses plus faciles. Le résultat est négatif à moins que les informations ne soient inexactes, nous devons confirmer la source.

De nombreux aspects peuvent être ajoutés, car le sujet des statistiques descriptives est vaste, nous proposons :

L'étude des statistiques descriptives univariée dans une cas conditionnelle.

Une étude descriptive deux variables dans le cas d'une variable quantitative et d'une autre variable qualitative.L'étude des statistiques descriptives pour plusieurs variables.

En fin de compte, j'espère à Dieu que ma thèse sera bénéfique à la prochaine génération, et je conclus ma recherche en disant au nom de Dieu, le Compatissant, le Miséricordieux " Allah élèvera en degrés ceux d'entre vous qui auront cru et ceux qui auront reçu le savoir".Al-Mujadalah-11.

Al-Shafei, que Dieu ait pitié de lui, a dit : "La connaissance n'est pas préservée, mais la connaissance est bénéfique."

Bibliographie

[Monino et al (2010)] Monino, Jean-Louis, Jean-Michel Kosianski, and François Le Cornu. Statistique descriptive. Dunod, 2010.

[Hamdani (1988)] Hamdani, Hocine. "Statistique descriptive et expression graphique." (1988).

[Lethielleux (2016)] Lethielleux, Maurice. Statistique descriptive-8e éd. : en 27 fiches. Dunod, 2016.

[Grais (1991)] Grais, Bernard. "Statistique descriptive." (1991).

[Dagnelie (2006)] .Dagnelie, Pierre. Statistique théorique et appliquée : 2. Inférence statistique à 1 et 2 dimensions. Vol. 2. De Boeck Supérieur, 2006.

[Alalouf et al (2002)] Alalouf, Serge, Denis Labelle, and Jean Ménard. Introduction à la statistique appliquée. Loze-Dion, 2002.

[Chekroun (2018)] Chekroun, Abdennasser. "Statistiques descriptives et exercices." (2018).

[Goldfarb, Catherine (2011)] Goldfarb, Bernard, and Catherine Pardoux. Introduction à la méthode statistique : manuel et exercices corrigés. Dunod, 2011.

[Leboucher, Marie (2013)] Leboucher, Lucien, and Marie-José Voisin. Introduction à la statistique descriptive : cours et exercices avec tableur. Cépaduès éd., 2013

[Lethielleux, Chevalier (2017)] Lethielleux, M., & Chevalier, C. (2017). Exercices de statistique et probabilités-3e éd. : Avec rappels de cours. Dunod.

[Bahouayila (2016)] Bahouayila, Bardin. "Cours de statistique descriptive." (2016).

[Baccini (2010)] Baccini, Alain. "Statistique descriptive élémentaire." Institut de Mathématiques de Toulouse (2010).

[Fabrice (2006)] Fabrice, Mazerolle. "Statistique descriptive." (2006).

[Hachmi] Hachmi, Ababsa.Cours statistiques descriptives de première année èconomique.Université Mohamed Khider de Biskra.

[Hammed (2012)] Hammed, Mountassir. "La statistique descriptive." (2012).

[Mémoire (2018)] Mémoire Master. "Statistique descriptive univariée". (2018).

[Meghlaoui (2011)] Meghlaoui, Dakhmouche. "Introduction à la statistique descriptive." (2011).

[Grammont (2003)] Grammont, Laurence. "Cours de statistiques infrentielles." (2003).

[Meraghni (2017)] Meraghni, Djamel. "Cours de tests statistiques première master." (2017).Université Mohamed Khider de Biskra.

[Vessereau (1965)] Vessereau, A. "Les méthodes statistiques appliquées au test des caractères organoleptiques." Revue de statistique appliquée 13.3 (1965) : 7-38.

[Mehl (1996)] Mehl, S. "Chronomath, Une chronologie des mathématiques à l'usage des professeurs de mathématiques et des élèves des lycées &; collèges." (1996).

[Université (2010)] Université libre de bruxelles. "Statistique descriptive univariée." http ://www.itse.be/statistique2010/co/Module_statistique_FSP.h

Annexe : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

Symbole		Signification
Ω	:	Population l'ensemble sur lequel porte notre étude statistique.
ω	:	Individu tout élément de la population Ω.
$V.S$	:	La variable statistique.
X	:	Caractère.
C	:	Ensemble des valeurs du caracère X.
N	:	La taille d'un population ou effectif total.
$\mathbb{N}$	:	Ensemble des nombres entiers naturels.
$\mathbb{Q}$	:	Ensemble des nombres entiers décimaux.
$Card(\Omega)$	:	Le cardinal : Nombre d'éléments de l'ensemble ■
$:=$	:	Est défini comme étant (symbole d'affectation).
n_i	:	Effectif observé dans la classe i
$\sum_{i=1}^{k}$	:	La somme pour i variant de 1 à k.
f	:	Fonction de densitée.
ECC	:	Effectif cumulé croissant.

FCC	:	Fréquence cumulée croissante.
ECD	:	Effectif cumulé décroissante.
FCD	:	Fréquence cumulée décroissante.
$F(x)$	:	Fonction de rèpartition.
f	:	Fréquence
C_i	:	Centre de classe
A_P	:	Coefficient d'asymétrie de Pearson
A_F	:	Coefficient d'asymétrie de Fisher
AP_P	:	Coefficient d'aplatissement de Pearson
AP_F	:	Coefficient d'aplatissement de Fisher
X^2	:	Loi khi-deux
$N(0,1)$	:	Loi normale standard
P	:	Probabilité
d^2	:	Distance
Y	:	Caractère
MOC	:	Moindres carrés ordinaire
r	:	Cofficient de corrélation
$f(.)$	:	fonction