République Algérienne Démocratique et Populaire

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud MAMMERI de Tizi Ouzou

Faculté des Sciences Economiques, Commerciales et des Sciences de Gestion

Département des Sciences Economiques

Polycopié de cours

Statistique I

1ère année Tronc Commun

Semestre 1

Réalisé par Dr. Gouraya BELBACHIR

Maître de conférences B

Email : gouraya.belbachir@ummto.dz

Année universitaire : 2023 – 2024

To access the online courses, please scan the **QR** code using the anonymous login	

Identification du module :

Unité d'enseignement	**Méthodologique**	
Crédits	**5**	
Coefficient	**3**	
Volume horaire hebdomadaire	**Cours**	**3h**
	TD	**1h 30 min**
	∑	**4 h 30 min**
Méthode d'évaluation	**Cours**	**20 / 20**
	TD	**20 / 20**
Poids relatif d'évaluation	**Cours**	**60 %**
	TD	**40 %**
	∑	**100 %**
La Moyenne du Module	**= (Note EMD * 0,6) + (Note TD * 0,4)**	
Exemple	Note EMD = **16 / 20** Note TD = **14 / 20** La Moyenne du Module = (16 * 0,6) + (14 * 0,4) La Moyenne du Module = (9,6 + 5,6) **La Moyenne du Module = 15,2 / 20**	
	Note EMD = **14 / 20** Note TD = **16 / 20** La Moyenne du Module = (14 * 0,6) + (16 * 0,4) La Moyenne du Module = (8,4 + 6,4) **La Moyenne du Module = 14,8 / 20**	

Objectifs de l'enseignement :

Le présent cours, conçu conformément au programme ministériel (Arrêté n° 808 du 22 juillet 2022), vise à ce que l'étudiant acquiert des compétences en la matière, à savoir :

- Maitriser les concepts clés de la statistique descriptive ;
- Résumer et présenter des données sous forme de tableaux et graphes ;
- Calculer et analyser les différents paramètres statistiques (de tendance central, de dispersion, de forme et de concentration) ;
- Analyser et quantifier la relation entre deux variables et mesurer leur corrélation ;
- Calculer les indices de la vie économique et la compréhension de leur signification et leur utilité et leur usage ;
- S'initier à l'usage des logiciels statistiques utilitaires, notamment le logiciel de base Excel, pour les graphiques et le calcul des paramètres.

Les chapitres dispensés en cours magistral font l'objet chacun d'un traitement appliqué, sous forme de séries d'exercices, en séances de Travaux Dirigés (TD). Lors de ces dernières des éclaircissements et des informations supplémentaires, surtout d'ordre pratique, sont fournis aux étudiants.

Une bibliographie révisée et mise à jour est fournie en annexe, dans le but de permettre aux étudiants d'approfondir leurs connaissances.

Les chapitres du présent cours sont également publiés en ligne, sur la plateforme Moodle (Elearning) en accès anonyme et mon e-mail (**gouraya.belbachir@ummto.dz**) est mis à la disposition des étudiants pour toute question ou renseignement éventuel.

Plan du cours :

Chapitre 7 : Les paramètres de concentration

Section 1 : L'analyse algébrique de la concentration

Section 2 : L'analyse graphique de la concentration

Chapitre 8 : Les indices

Section 1 : Les indices élémentaires

Section 2 : Les indices synthétiques

Chapitre 9 : Distributions à deux caractères, corrélation et régression

Section 1 : Présentation et notions fondamentales des distributions à deux caractères

Section 2 : Caractéristiques des distributions à deux caractères

Section 3 : Analyse de la relation entre deux variables quantitatives

- Conclusion générale

- Bibliographie

Introduction générale

Introduction générale

Ce module présente les notions fondamentales de statistique, Les unités statistiques regroupées au sein d'une population peuvent être étudiées à l'aide de plusieurs caractères qualitatifs ou quantitatifs, les informations obtenues permettent alors d'élaborer les séries à plusieurs caractères. Par exemple, les salariés d'une entreprise peuvent être étudiés à partir de caractères tels que l'âge, le sexe, le niveau de formation ou de qualification, la rémunération, l'ancienneté dans l'entreprise.

La plupart des études statistiques effectuées sur une population portent donc sur plusieurs caractères simultanément. Cela permet de réduire les coûts liés à la collecte des informations (recensement de la population), ainsi que de mettre en évidence certains liens potentiels entre les variables observées.

Dans ce cours, seules les séries statistiques avec un caractère et à deux caractères feront l'objet d'une analyse qui n'est qu'une introduction à l'étude globale des relations entre plusieurs variables. L'accent sera mis sur l'interprétation des résultats autant que sur les techniques statistiques elles-mêmes.

Chapitre 1 :

Notions générales

Chapitre 1 : Notions générales

L’objet du présent chapitre est de faire découvrir à l’étudiant deux fondamentaux en statistique descriptive : le vocabulaire de base et les soubassements théorique et pratique de l’analyse statistique. Seront ainsi présentés, en trois sections successives, trois volets de définition, en allant du général vers le détail.

Section 1 : Concept des statistiques
Section 2 : La population, l’échantillon et l'individu

Section 1 : Concept des statistiques

1. La Statistique et Les Statistiques :

- La Statistique : est une méthode de raisonnement permettant d'interpréter le genre de données très particulières, qu'on rencontre notamment dans les sciences économiques.

- On appelle statistique l'ensemble des outils scientifiques, à partir des quels on peut recueillir, ordonner, analyser et tirer des conclusions d'un certain nombre de données.

- Les Statistiques : ensemble des données relatives à un groupe d'individus ou d'unités.

- Une série statistique : est la correspondance entre les individus d'une population et les modalités du caractère étudié.

2. Les étapes de statistique descriptive :

- Tout d'abord, **la collecte des informations**. En dépit des apparences, cette étape est essentielle et s'avère souvent complexe. Son bon déroulement suppose d'avoir répondu préalablement à trois questions :

▪ Quelles informations cherche-t-on à recueillir ? La réponse à cette question définit *l'objet de la collecte.*

▪ Auprès de qui ces informations seront-elles recueillies ? La réponse à cette question définit *le sujet de la collecte.*

▪ Comment ces informations seront-elles recueillies ? La réponse à cette question définit *la méthode de la collecte.*

- Deuxième étape, **la présentation des données.** Une fois les données collectées, il importe « d'organiser » la statistique obtenue. Cette présentation prend la forme de *tableaux et de graphiques.*

- Troisième étape, **le résumé des données**. Paradoxalement, l'information exprimée dans un tableau ou visualisée par un graphique, est parfois trop riche pour être véritablement utile. La troisième étape va donc consister à définir et à calculer quelques paramètres qui expriment les caractéristiques principales de la distribution, et qui en quelque sorte, la « résument ».

Les résultats ainsi obtenus seront beaucoup plus évocateurs, plus « parlants », que le tableau ou le graphique. Mais cette opération présente aussi, des inconvénients. Résumer la distribution, c'est accepter une perte d'information, et peut-être une déformation de l'information. Le statisticien doit en être conscient, et ne pas hésiter à revenir dans ses raisonnements ultérieurs sur la série de chiffres initiaux.

Au terme de ces trois étapes, le statisticien *a décrit la population interrogée, sous l'angle particulier qui l'intéressait*. Les méthodes qu'il a mises en œuvre au cours de cette démarche constituent la statistique descriptive.

La statistique descriptive vise ainsi à collecter, présenter et résumer des données[*].

Section 2 : La population, l'échantillon et l'individu

La mise en œuvre d'une démarche de statistique descriptive dépend en pratique de *nature des variables retenues*. Il nous faut donc, pour aller plus avant, éclaircir quelques points de vocabulaire statistique.

1. La population :

L'analyse statistique débute toujours par la collecte de données sur un ensemble de référence concerné par l'objet de l'étude : **la population** (notée **P**). La population est donc un ensemble finis d'éléments de même nature, qui sont objets de l'observation du statisticien. (Exemples de populations : Les véhicules automobiles immatriculés en Algérie, La population des P.M.E. d'un pays, Les salariés d'une entreprise, Les habitants d'un quartier).

2. L'échantillon :

Dans la mesure où il serait trop lourd d'étudier l'ensemble d'une population, on choisit[*] d'en étudier une partie représentative, appelée échantillon.

---

[*] Il existe d'autres branches de la statistique. Elles font appel à des techniques plus élaborées tant sur le plan des mathématiques que des probabilités.

[*] Cette pratique de l'échantillonnage est très fréquente en sociologie politique, avec les sondages d'opinion, mais également en économie d'entreprise, avec les études de marché. A l'inverse, dans le cas des recensements de population, la prétention du statisticien est de tendre à l'exhaustivité. La collecte s'opère non sur échantillon, mais sur la totalité de la population. Mais le traitement statistique définitif sera long et coûteux. De ce fait, les recensements ne sont pratiqués que tous les 6 ou 7 ans.

L'échantillon est représentatif [1] :

- Si sa taille est suffisamment grande ;
- Si il est extrait au hasard de la population (tirage au sort).

3. L'Individu :

L'effectif de la population (ou la taille de l'échantillon) correspond alors au nombre d'individus qui composent cette population. On notera cet effectif **N.** Ainsi pour une population P donnée, on a $P = (U_1, U_2, U_3, \ldots, U_i, \ldots, U_n)$.

Chaque élément d'une population est également dit *individu ou unité statistique* souvent notée $\mathbf{U_i}$.

Ces premières définitions permettent de définir *le sujet* de la collecte. Il reste donc à préciser *l'objet* de l'observation.

4. Le caractère :

L'objet de l'observation est dit **caractère**. Ainsi, sur une population donnée, le statisticien peut s'intéresser simultanément à plusieurs caractéristiques des individus. Sur une population humaine, par exemple, le statisticien peut relever entre autres l'âge, le sexe, le poids, la couleur des yeux, la forme du crâne, l'opinion politique, l'origine sociale …

Le choix des caractères à observer est essentiel. Il doit permettre de répondre à la problématique posée au départ. Il importe donc de ne pas ignorer un caractère indispensable à l'analyse, mais tout autant de ne pas s'encombrer de caractéristique sans importance.

Le statisticien distingue, pour l'analyse, **trois types de caractères** : qualitatif et quantitatif discret ou quantitatif continu. La façon de traiter ces trois types de caractères diffère sensiblement[2].

a) Le caractère qualitatif : Un caractère qualitatif *diffère en nature* d'une unité statistique à une autre, et il ne peut donc ni être mesuré ni se voir (directement) attribuer une valeur numérique. Ainsi, le sexe, la couleur des yeux, la forme du crâne,

[1] L. FOUCAN, **Probabilités et Statistiques**, PAES, 2012 – 2013, P : 4.
[2] Christian DESMARIS, **Informatique & Statistique**, Tome 2, Conférence de Méthode, Institut d'Études Politiques de Lyon, Université Lumière, Lyon II, France, 2003-2004, P : 5.

l'opinion politique ou encore l'origine sociale d'un individu sont des caractères qualitatifs.

b) Le caractère quantitatif : Un caractère est dit quantitatif lorsqu'il est « mesurable », c'est à dire lorsqu'on peut associer, à chaque modalité du caractère, un nombre qui en exprime l'intensité. Les caractères quantitatifs *diffèrent en intensité* d'une unité statistique à une autre. Ainsi, l'âge, le sexe, le poids sont des caractères quantitatifs.

Il importe encore de distinguer entre caractère quantitatif discret et caractère quantitatif continu.

- Le caractère quantitatif discret : Les modalités de la variable sont exprimées par des nombres isolés, entiers en général. Par exemple, si la variable exprime le nombre de personnes dans un ménage, le nombre d'enfants dans une famille, le nombre de places de stationnement ou encore le nombre de véhicules par ménage, nous avons affaire à une variable quantitative discrète.

- Le caractère quantitatif continu. Les modalités de la variable peuvent prendre toutes les valeurs comprises dans un intervalle donné, c'est à dire un nombre infini de valeurs. De façon générale, toutes les grandeurs liées à l'espace, au temps et à la masse sont par nature des variables quantitatives continues.

Mais il n'est pas toujours facile de déterminer si une variable statistique doit être considérée et traitée comme une grandeur discrète ou comme une valeur continue, et dans un grand nombre de cas, le choix peut présenter un caractère relativement arbitraire ou conventionnel.

Par exemple, les notes mises par un correcteur à un examen, peuvent théoriquement prendre toutes les valeurs comprises entre 0 et 20 et la variable être traitée comme continue. Mais, en pratique, le correcteur peut, ne mettre que des notes entières, ou, avec une précision plus grande mais souvent illusoire, noter au demi-point, ce qui incite à considérer la variable comme discrète. En revanche, la moyenne des notes obtenues aux différentes épreuves d'un examen doit toujours être traitée comme une variable continue.

De même, toute grandeur qui s'exprime en unité monétaire est par nature discrète, puisqu'elle ne peut prendre que des valeurs successives distinctes. Mais en pratique,

pour peu que la grandeur étudiée concerne des montants importants par rapport à l'unité monétaire utilisée, l'on pourra traiter la variable comme si elle était continue. C'est généralement le cas pour les études sur les revenus et les patrimoines des ménages.

5. La variable :

D'un point de vue mathématique. Une **variable** est une application pour laquelle on a un ensemble de départ : la population étudiée et un ensemble d'arrivée qui va définir le type de la variable. A chaque individu de l'ensemble de départ, on associe une seule valeur de l'ensemble d'arrivée.

Habituellement, une variable est désignée par une lettre majuscule, sauf si elle prend une valeur particulière, auquel cas on utilise une lettre minuscule : par exemple x_i est la valeur de X prise par le $i^{ème}$ élément, et $\overline{X}$ est la valeur moyenne de X dans l'ensemble étudiée.

Nous retrouvons donc les **deux types de variables** étudiées précédemment[1] :

a) Les variables qualitatives : l'ensemble d'arrivée est un ensemble fini d'éléments sans structure particulière ;

b) Les variables quantitatives : l'ensemble d'arrivée est l'ensemble des nombres réels.

6. La modalité :

On appellera **modalité** d'un caractère chacun des états que peuvent présenter les unités statistiques. Par exemple, pour le sexe, deux états sont possibles : mâle ou femelle. Pour l'origine sociale ou pour l'opinion politique, les choses sont plus délicates et le statisticien aura en général recours à une typologie construite, **une nomenclature** qui regroupe l'ensemble des modalités possibles.

Toute bonne nomenclature doit se conformer à deux principes : l'incompatibilité et l'exhaustivité. Chaque individu étudié appartient à un seul sous-ensemble, c'est-à-dire ne peut prendre qu'une seule modalité : c'est la propriété de l'**incompatibilité**. En outre, la réunion des sous-ensembles recouvre la population étudiée. Toutes les

[1] Christian DESMARIS, Op.cit., P : 7.

situations doivent être prévues, c'est-à-dire qu'un individu possède toujours l'une des modalités : c'est la propriété de l'**exhaustivité**.

Cette exigence conduit fréquemment en pratique à prévoir des modalités « fourre-tout » qui permettent d'enregistrer des cas particuliers, généralement peu nombreux, pour lesquels on a volontairement refusé de créer des modalités supplémentaires ou pour lesquelles on ne dispose pas d'informations suffisamment précises. Ainsi, la plupart des nomenclatures incorporent une rubrique « Autres » ou « ND : Non Défini ».

7. La fréquence :

La fréquence correspond au nombre de fois où la modalité apparaît proportionnellement à la population totale étudiée. Les fréquences sont obtenues en faisant le rapport des effectifs sur l'effectif total et sont donc comprises entre 0 et 1. Les fréquences sont généralement exprimées en pourcentages et sont alors comprises entre 0 et 100.

L'avantage d'une distribution en fréquence est de permettre une meilleure lisibilité et comparabilité de l'information de départ. En effet, quel que soit l'effectif, toutes les lectures seront effectuées au regard d'une base 100.

On note :

Effectif total = N

Effectif d'une variable : pour chaque valeur xi de la variable X on note ni son effectif, c'est à dire le nombre d'individus de la population qui présentent la modalité i.

Fréquence de la variable xi = fi

On a donc :

$\sum ni = n_1 + n_2 + n_3 + \ldots + n_n = N$

$fi = \frac{n_i}{N}$

$\sum fi = f_1 + f_2 + f_3 + \ldots + f_n = 1$

8. Effectifs cumulés et fréquences cumulées d'une variable :

Il existe en fait deux définitions des fréquences cumulées[*] :

- Définition française : pourcentage d'individus dont le caractère est strictement inférieur à xi (somme des fréquences jusqu'à i-1) ;
- Définition anglo-saxonne : pourcentage d'individus dont le caractère est inférieur ou égal à xi (somme des fréquences jusqu'à i).

Exemples :

a) Etude des différentes spécialités choisies par les étudiants de la 2ème année S. T.

- Population : Les étudiants de la 1ere année LMD / FSECSG / UMMTO.
- Individu : Un étudiant.
- Caractère étudié : Les spécialités choisies.
- Modalités : Sciences économiques, Sciences commerciales, Sciences de gestion, Sciences financières et comptabilité.

b) Etude du poids des nouveaux née.

- Population : Les nouveaux née.
- Individu : Un nouveau née.
- Caractère étudié : Le poids.
- Modalités 3.400 kg, 2.900 kg, . . . etc.

c) Etude du nombre des travailleurs dans un certain nombre de petites entreprises.

- Population : Les petites entreprises.
- Individu : Une petite entreprise
- Caractère étudié : Le nombre des travailleurs.
- Modalités : 10, 25, 5, . . . etc.

[*] Nous retenons ici la définition anglaise de la fréquence.

Chapitre 2 :
Représentation des données statistiques (Tabulation)

Chapitre 2 : Représentation des données statistiques (Tabulation)

L'objectif de la statistique est de collecter, d'analyser et d'interpréter des données (des ensembles d'observations) relative à un même phénomène et susceptible d'être caractérisée par un nombre.

A cette fin, le travail du statisticien comprend trois étapes :

- La collecte des données ;
- La présentation des données ;
- Le résumé des données.

La présentation des données. Une fois les données collectées, il importe « d'organiser » la statistique obtenue. Cette présentation prend la forme de tableaux et de graphiques.

Section 1 : Représentation des données statistiques qualitatives

Section 2 : Représentation des données statistiques quantitatives

Section 1 : Représentation des données statistiques qualitatives

S'agissant d'un caractère qualitatif, il faut ordonner les modalités à l'intérieur du tableau soit : - par ordre alphabétique s'il s'agit d'un caractère qualitatif nominal ;

- par ordre d'importance ou hiérarchique s'il s'agit d'un caractère qualitatif ordinal, ce qui facilite encore davantage leur ordonnancement.

Exemple :

Les chiffres d'affaires trimestriels dans un magasin de matériel informatique, en fonction de la marque des produits vendus, se répartissent comme suit :

(U.I : 1000€)

Marque	HP	Apple	Toshiba	Samsung	**Total**
Chiffre d'affaires	55	30	15	20	**120**

Section 2 : Représentation des données statistiques quantitatives

1. Cas d'une variable discrète (VSD) :

Considérons une population à N individus, décrite suivant une variable statistique discrète X ayant les valeurs (x_1, x_2, …, x_k). On s'intéresse donc à connaître, pour chaque valeur x_i , le nombre d'individus prenant cette valeur, ce nombre est noté par $\boldsymbol{n_i}$, $i = 1,...,k$. Nous obtenons donc le ***tableau statistique*** suivant :

Les valeurs x_i	$\boldsymbol{n_i}$
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_k	n_k
Total	$\boldsymbol{N}$

a) Effectif : Le nombre d'individus n_i de la population, pour lesquels la variable X prend la valeur x_i, est appelé **effectif** ou **fréquence absolue** de la valeur x_i.

b) La fréquence relative : f_i de la valeur x_i d'effectif n_i est donnée par la formule $f_i = \frac{n_i}{N}$, ou N est l'**effectif total** de la population.

c) Le pourcentage : p_i de la valeur x_i d'effectif n_i est donné par la formule

$p_i = f_i \times 100 = \frac{n_i}{N} \times 100.$

Remarque :

- $\sum_{i=1}^{k} n_i = N \; et \; 0 \le n_i \le N$, ou k est le nombre des valeurs différentes.

- $\sum_{i=1}^{k} f_i = 1$ et $\sum_{i=1}^{k} p_i = 100.$

- La correspondance entre les valeurs de x_i et leurs effectifs s'appelle distribution d'effectifs.

d) Effectifs cumulés : Il peut être intéressant par la lecture du tableau de répondre à des questions de la forme:

-Quel est le nombre d'individus pour les quels la variable X prend au plus x_j ?

-Quel est le nombre d'individus pour les quels la variable X prend au moins x_j ?

La réponse à la 1ére question se fait en additionnant les effectifs à partir de la première valeur n_1 jusqu'à n_j ($1 \le j \le k$). Les nombres ainsi obtenus sont appelés effectifs cumulés croissants ou fréquences absolus cumulées croissantes, notés par $\boldsymbol{n_{ic}\uparrow}$.

La réponse à la 2ème question se fait en additionnant les effectifs à partir de n_j ($1 \le j \le k$) jusqu'à la dernière valeur n_k. Les nombres ainsi obtenus sont appelés effectifs cumulés décroissants ou fréquences absolus cumulées décroissantes, notés par $n_{ic}\downarrow$.

De la même manière on peut définir :

e) Les fréquences relatives cumulées (croissantes et décroissantes).

f) Les pourcentages cumulés (croissants et décroissants).

Exemple :

A study on the number of milk litres bought each week by 100 consumers gives the following results :

Number of bought milk litres	0	1	2	3	4	5
Number of consumers	5	20	35	25	10	5

1. Calculer les effectifs cumulés croissants et décroissants.

Solution :

X_i	n_i	$n_i\uparrow$	$n_i\uparrow$	$n_i\downarrow$	$n_i\downarrow$
0	5	5	5	100	100
1	20	5 + 20 = 25	25	100 - 5 = 95	95
2	35	25 + 35 = 60	60	95 - 20 = 75	75
3	25	60 + 25 = 85	85	75 - 35 = 40	40
4	10	85 + 10 = 95	95	40 - 25 = 15	15
5	5	95 + 5 = 100	100	15 - 10 = 5	5
$\sum$	**100**	-	-	-	-

2. Cas d'une variable continue (VSC) :

Dans le cas d'une variable continue, théoriquement les valeurs recueillies sont infinies et très proches l'une de l'autre. Alors, pour simplifier l'étude on construit des **classes** (intervalles) en divisant l'**étendue** de la série statistique en plusieurs intervalles.

a) L'étendue : d'une série statistique est la différence entre la plus grande et la plus petite valeur dans la série.

b) Les classes : sont des intervalles de la forme $[a_i\ ,\ a_{i+1}[$ ou a_0 et a_k sont respectivement la plus petite et la plus grande valeur de la série.

c) Dans la classe $[a_i\ ,\ a_{i+1}[$, les valeurs a_i et a_{i+1} sont les **bornes** ou les **limites** de cette classe.

d) Le nombre $x_i = \frac{a_i + a_{i+1}}{2}$ s'appelle le **centre** de la classe $[a_i\ ,\ a_{i+1}[$.

e) Le nombre $\alpha_i = a_{i+1} - a_i$ s'appelle l'**étendue** ou **amplitude** de la classe $[a_i\ ,\ a_{i+1}[$.

f) L'effectif n_i la classe $[a_i\ ,\ a_{i+1}[$.correspond au nombre de valeurs appartenant à cette classe

Remarque : Le nombre de classes k ne doit pas être trop petit, perte d'information, ni trop grand, le regroupement en classe est alors inutile. Le nombre de classe qu'on peut construire est donné par la formule $k = \left[\sqrt{N}\right]$.

Exemple :

Soit la distribution statistique suivante qui donne la répartition des entreprises commerciales selon la part des marchés réalisée dans un territoire économique :

La part des marchés (%)	2 - 4	4 - 8	8 - 20	20 - 40	40 - 45	45 - 100
Nombre des entreprises	24	36	22	18	14	6

1. Calculer les effectifs cumulés croissants et décroissants.

Solution :

Classes	$\mathbf{n_i}$	**ai**	$\mathbf{n_{ic}}$	$\mathbf{n_i}\uparrow$
2 - 4	24	2	24	24
4 - 8	36	4	18	60
8 - 20	22	12	3,66	82
20 -40	18	20	1,8	100
40 - 45	14	5	5,6	114
45 - 100	6	55	0,21	120
∑	**120**	**-**	**-**	**-**

▪ $\mathbf{N = \sum n_i}$

$N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6$

$N = 24 + 36 + 22 + 18 + 14 + 6$

$N = 120$

▪ $\mathbf{a_i = L_s - L_i}$

$a_1 = 4 – 2 = 2$

$a_2 = 8 – 4 = 4$

$a_3 = 20 – 8 = 12$

$a_4 = 40 – 20 = 20$

$a_5 = 45 – 40 = 5$

$a_6 = 100 – 45 = 55$

$a_i \neq$ Constante $\Rightarrow n_{ic} = \frac{n_i}{a_i} * a_0$ / a_0 : l'amplitude de base = 2

▪ $\mathbf{n_{ic} = \frac{n_i}{a_i} * a_0}$

$$n_{1c} = \frac{n_1}{a_1} * a_0 \Rightarrow n_{1c} = \frac{24}{2} * 2 \Rightarrow n_{1c} = 24$$

$$n_{2c} = \frac{n_2}{a_2} * a_0 \Rightarrow n_{2c} = \frac{36}{4} * 2 \Rightarrow n_{2c} = 18$$

$$n_{3c} = \frac{n_3}{a_3} * a_0 \Rightarrow n_{3c} = \frac{22}{12} * 2 \Rightarrow n_{3c} = 3,66$$

$$n_{4c} = \frac{n_4}{a_4} * a_0 \Rightarrow n_{4c} = \frac{18}{20} * 2 \Rightarrow n_{4c} = 1,8$$

$$n_{5c} = \frac{n_5}{a_5} * a_0 \Rightarrow n_{5c} = \frac{14}{5} * 2 \Rightarrow n_{5c} = 5,6$$

$$n_{6c} = \frac{n_6}{a_6} * a_0 \Rightarrow n_{6c} = \frac{6}{55} * 2 \Rightarrow n_{6c} = 0,21$$

Ministry of Higher Education & Scientific Research **Mouloud MAMMERI University of Tizi Ouzou – Algeria** **Faculty of Economics, Business & Management** **Department of Economics** **1st Year LMD - Section E3**		**Academic Year : 2024 – 2025** **Module : Statistics I** **Worksheet 1st** **Lecturer : G. BELBACHIR** **Email : gouraya.belbachir@ummto.dz**

To access the online answer key, please scan the **QR** code using the anonymous login.	

Exercise 1 :

Classer les caractères suivants selon le type et la nature :

Le nombre d'étudiants par groupe - Le prix d'un produit - Le nombre d'enfants par ménage - Les filières dispensées par une université - La durée de la mission - La mention au BAC - L'origine géographique - Le code postal.

Exercise 2 :

Précisez la population, le caractère et la nature du caractère (qualitatif ou quantitatif), lorsqu'on considère:

1. Le montant du salaire annuel des employés d'une entreprise.

2. La taille des élèves d'une classe de seconde.

3. La couleur des ours en peluche dans un magasin de jouets.

4. L'âge des députés de l'assemblée nationale.

Exercise 3 :

Un professeur interroge les 35 élèves d'une classe de baccalauréat sur le nombre de leurs frères et sœurs. Voici les résultats obtenus : 1 3 1 1 2 2 1 1 2 2 1 3 5 3 1 1 0 3 5 3 2 2 2 1 1 2 3 2 1 1 2 4 1 1 2

1. Dresser le tableau statistique.

2. Donner le tableau des fréquences.

3. Indiquer dans un tableau les fréquences cumulées croissantes et les fréquences cumulées décroissantes.

4. Quel est le pourcentage des élèves qui ont au plus deux frères et sœurs ?

5. Quel est le pourcentage des élèves qui ont au moins trois frères et sœurs ?

Exercise 4 :

Le recensement des ménages d'une ville selon le nombre d'enfants a donné les résultats ci-dessous.

Individus par ménage	1	2	3	4	5	6
Nombre de ménages	380	274	250	188	120	85

1. Calculer les effectifs cumulés, les fréquences relatives et les fréquences cumulées.

2. Combien de ménages ont plus de 2 enfants ? Combien en ont moins de 2 ? Combien en ont plus de 1 et moins de 6 ?

Exercise 5 :

Les données suivantes représentent l'âge d'un groupe d'individus : 15 19 31 30 23 76 13 35 27 32 77 35 24 18 18 15 45 76 81 27 76 23 18 18 75 15 69 14 75 63 29 19 81 15 29 81 45 17 15 31 18 31

1. Grouper les nombres précédents en classes de façon à mettre en évidence les effectifs, les centres de classes, les amplitudes, les fréquences relatives cumulées et les bornes de classes.

~ ... ~ ... ~

Chapitre 3 :
Représentation des données statistiques (Graphe)

Chapitre 3 : Représentation des données statistiques (Graphe)

Aux tableaux statistiques, on associe des représentations graphiques qui permettent une compréhension plus globale des données. Les représentations graphiques sont différentes selon le type de la variable.

Les représentations graphiques permettent d'avoir rapidement une vue d'ensemble d'un tableau de données.

Section 1 : Représentation des données statistiques qualitatives

Section 2 : Représentation des données statistiques quantitatives

Section 1 : Représentation des données statistiques qualitatives

La distribution des fréquences d'une variable qualitative peut être représentée soit par un diagramme en secteurs, soit par un diagramme en tuyaux d'orgue.

Dans le cas du diagramme en secteur, les modalités sont représentées par des aires. Si l'on ne dispose pas de tableur nous permettant d'obtenir automatiquement un tel graphique, il faudra calculer les angles des différents secteurs. Pour cela, on effectuera un produit en croix en utilisant les fréquences correspondant aux modalités (fi * 360 / 100). Ainsi, on obtient la valeur de l'angle de la modalité.

Dans le cas du diagramme en tuyaux d'orgue les barres sont des rectangles de même base et de hauteurs proportionnelles aux effectifs.

Notez que lorsque le nombre de modalités de la variable est important (plus que 5) et les valeurs insuffisamment contrastées, il est préférable de recourir au diagramme en barres, plus lisible.

1. Le diagramme circulaire :

Appelé aussi diagramme à secteurs ou camembert, ce graphique repose sur le système de coordonnées polaires. Le principe de sa représentation consiste (conformément au principe de proportionnalité des surfaces aux effectifs) à considérer que l'angle total formé par le cercle, soit 360°, correspond à l'effectif total (N) de la distribution à la somme des fréquences relatives, soit 1.

L'angle (α_i) se calcule comme suit :

$$\boldsymbol{\alpha_i = (n_i / N) \cdot 360 = f_i \cdot 360}$$

Exemple :

Les chiffres d'affaires trimestriels dans un magasin de matériel informatique, en fonction de la marque des produits vendus, se répartissent comme suit :

(U.I : 1000€)

Marque	HP	Apple	Toshiba	Samsung	**Total**
Chiffre d'affaires	55	30	15	20	**120**
f_i (%)	45,83	25,00	12,50	16,67	**100**

1. Représenter graphiquement cette distribution statistique par un diagramme circulaire.

Pour représenter graphiquement cette distribution, il faut d'abord compléter le tableau en calculant les fréquences et les angles (α_i), comme suit :

Marque	HP	Apple	Toshiba	Samsung	**Total**
Chiffre d'affaires	55	30	15	20	**120**
f_i (%)	45,83	25,00	12,50	16,67	**100**
αi	164,98	90	45	60,01	**360**

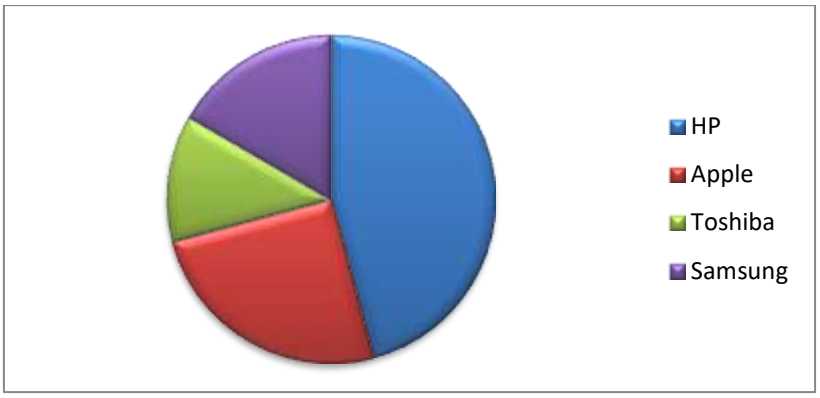


Diagramme circulaire

2. Le diagramme à barres :

Ce type de graphique repose sur un système de coordonnées cartésiennes avec une échelle ordinale. Sur un plan orthonormé, on représente chaque modalité du caractère par une barre dont la hauteur est proportionnelle à l'effectif ou à la fréquence, représentée par l'axe vertical yy', et la largeur de dimension arbitraire puisqu'il s'agit d'un caractère qualitatif qui ne reflète pas une mesure, et de préférence constante pour des raisons d'esthétique.

Reprenant l'exemple précédent, on aura graphiquement :

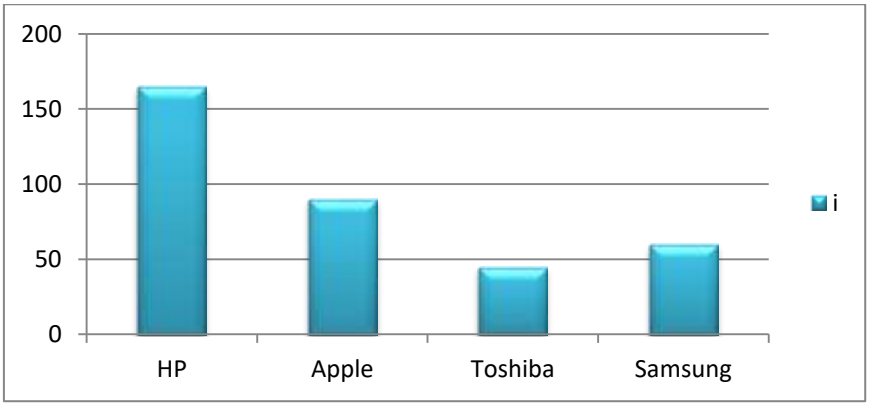


Diagramme à barres

Section 2 : Représentation des données statistiques quantitatives

Parmi les variables quantitatives on distingue les variables quantitatives discrètes et les variables quantitatives continues.

Les variables discrètes ne prennent qu'un nombre fini de valeurs, par exemple le nombre d'enfants par famille ;

Les variables continues prennent toutes les valeurs possibles d'un intervalle de nombres réels. Par exemple le poids ou la taille d'une personne (240,23 Kg et 2.35 mètres !), le temps d'attente à un guichet ...

On ne peut pas toujours faire facilement la distinction entre les deux types de variables.

Généralement on a :

Si la variable étudiée peut prendre un petit nombre de valeurs distinctes on la considère comme une variable quantitative discrète (VSD) ;

Si la variable étudiée peut prendre un grand nombre de valeurs distinctes on la regroupe en classes et on l'étudie comme une variable quantitative continue (VSC).

Dans tous les cas, le choix de l'étude doit être expliqué et interprété.

1. Distribution d'effectif et de fréquence d'une variable quantitative discrète (VSD) :

Pour obtenir la distribution d'effectifs et de fréquences d'une variable quantitative discrète, on procède en trois étapes :

- on classe les valeurs de la variable dans l'ordre croissant ;
- on compte les effectifs qui s'y rapportent ;
- enfin on calcule les fréquences pour chacune des modalités de la variable.

a) Diagramme en bâtons :

Cette représentation se fait en portant sur l'axe des abscisses, les valeurs xi prisent par la v. s. puis, on trace à partir de chaque point xi un bâton dont la longueur est proportionnelle à ni ou fi.

Exemple :

Classement de 100 familles en fonction du nombre d'enfants

On a relevé le nombre d'enfants de 100 familles choisies au hasard. Le tableau suivant donne les principales caractéristiques de cette étude.

Tableau : Statistique sur le nombre d'enfants de 100 familles.

Xi	ni	fi	fi↑
0	20	0,2	0,20
1	25	0,25	0,45
2	30	0,3	0,75
3	10	0,1	0,85
4	5	0,05	0,90
5	5	0,05	0,95
6	3	0,03	0,98
7	2	0,02	1
Σ	**100**	**1**	**-**

xi : nombre d'enfants compris entre 0 et 7.

ni : nombre de familles ayant *xi* enfants.

fi : fréquence relative des familles ayant *xi* enfants.

fi↑*:* fréquence cumulée des familles ayant au plus *xi* enfants.

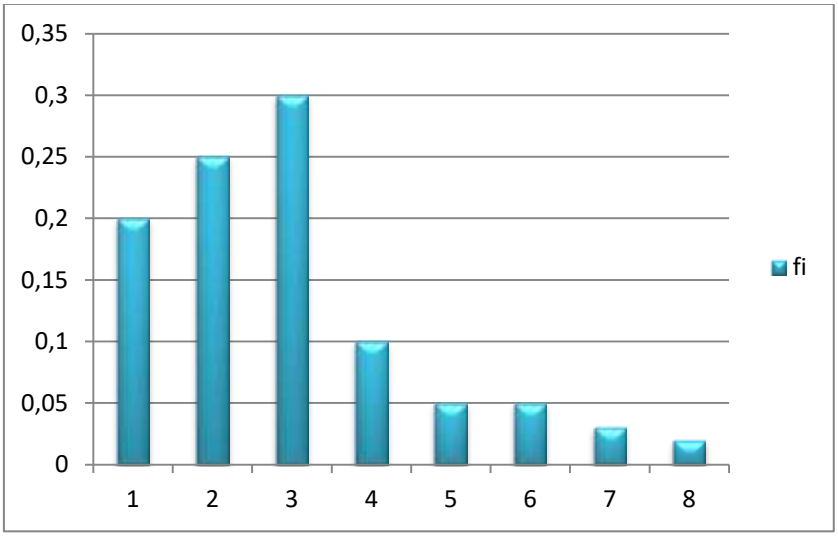


Figure : Diagramme en bâtons de la distribution de l'exemple précédent

b) Le polygone des fréquences (ou des effectifs) :

Cette représentation est obtenue en joignant les sommets bâtons.

c) Le polygone des effectifs cumulés (ou courbe cumulative des effectifs) :

On définit la fonction qui associe à chaque valeur $x \in \mathbb{R}$, la somme des effectifs de tous les xi < x et qu'on appelle fonction de distribution des effectifs. La représentation graphique de cette fonction est appelée polygone des effectifs cumulés.

2. Distribution d'effectif et de fréquence d'une variable quantitative continue (VSC) :

a) Cas des classes à étendues égale :

- Histogramme : L'histogramme est un outil statistique facile à utiliser, donnant rapidement[1], et par principe l'histogramme représente une distribution statistique d'une variable continue.

Sur l'axe des abscisses, on représente les bornes des différentes classes et on associe à chaque classe un rectangle, dont la base est une partie de l'axe des abscisses comprise entre les bornes de cette classe et dont la longueur est proportionnelle à ni ou fi. Sa construction nécessite de respecter l'hypothèse d'équirépartition des effectifs dans chaque classe.

Si toutes les classes ont la même amplitude (ai), on porte directement en ordonnée les effectifs (ou les fréquences).

- Le polygone des fréquences (ou des effectifs) : Cette représentation est obtenue en joignant les points (xi, ni) par des segments de droite. On complète par 2 classes extrêmes de même amplitude.

Remarque :

- L'aire de tous les rectangles est égale à 1 si on représente les fréquences relatives et N si on représente les effectifs.
- L'aire comprise entre < le polygone des effectifs et l'axe des abscisses est égale à l'aire de l'histogramme.

[1] Renée VEYSSEYRE, **Aide-mémoire Statistique et probabilités pour l'ingénieur**, 2e édition, Dunod, Paris, France, 2006, P : 11.

- Courbes de fréquences cumulées :

▪ **Courbe cumulative croissante :** on joint les points ayant pour abscisses la limite supérieure des classes et pour ordonnées les fréquences cumulées croissantes correspondant à la classe considérée (pour le premier point, on porte la valeur 0). Elle donne le nombre d'observations inférieures à une valeur quelconque de la série.

▪ **Courbe cumulative décroissante :** la construction de cette courbe est analogue à la précédente. Les points ont pour abscisses, les limites inférieures des classes et pour ordonnées, les fréquences cumulées décroissantes (pour le dernier point, la valeur est 0). Elle donne le nombre d'observations supérieures à une valeur quelconque de la série.

Exemple :

Étude de la dispersion d'un lot de 400 résistances.

On a contrôlé 400 résistances dont la valeur nominale est égale à 100 KΩ et on a regroupé les résultats en classes d'amplitude 2 KΩ qui représente environ le dixième de la dispersion totale de l'échantillon contrôlé.

Tableau : Étude statistique des mesures de la résistance d'un lot de 400 pièces.

Les classes	**ni**	**ni↑**	**fi**	**fi↑**	**fi↓**
92-94	10	10	0,025	0,025	1
94-96	15	25	0,0375	0,0625	0,975
96-98	40	65	0,1	0,1625	0,9375
98-100	60	125	0,15	0,3125	0,8375
100-102	90	215	0,225	0,5375	0,6875
102-104	70	285	0,175	0,7125	0,4625
104-106	50	335	0,125	0,8375	0,2875
106-108	35	370	0,0875	0,9250	0,1625
108-110	20	390	0,05	0,9750	0,075
110-112	10	400	0,025	1	0,025
∑	**400**	**-**	**1**	**-**	**-**

Les classes étant toutes de même amplitude, l'histogramme est facile à tracer ; il suffit de construire des rectangles dont l'aire est proportionnelle à la fréquence des résistances de la classe correspondante.

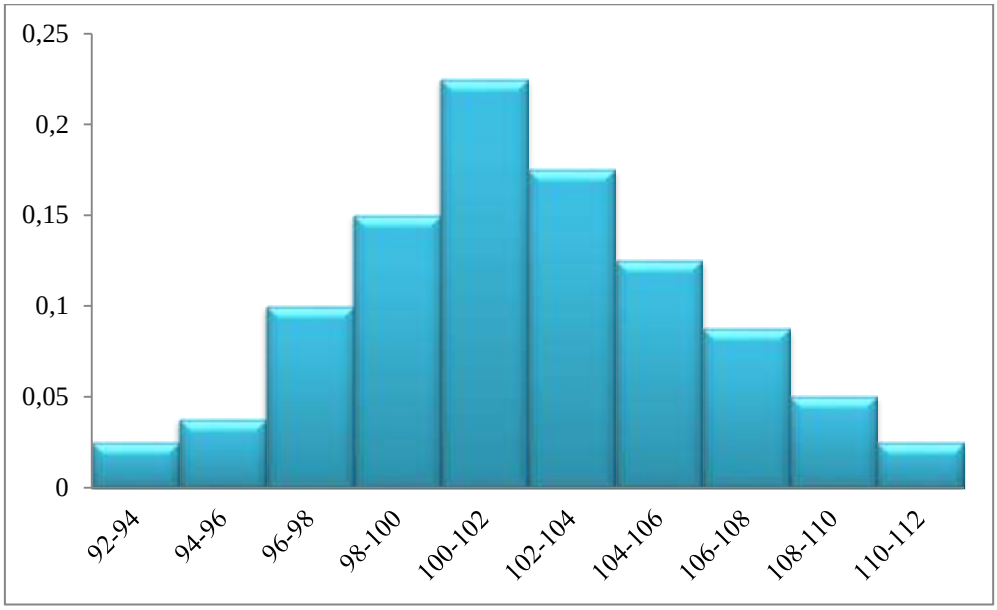


Figure : Histogramme de la distribution de l'exemple

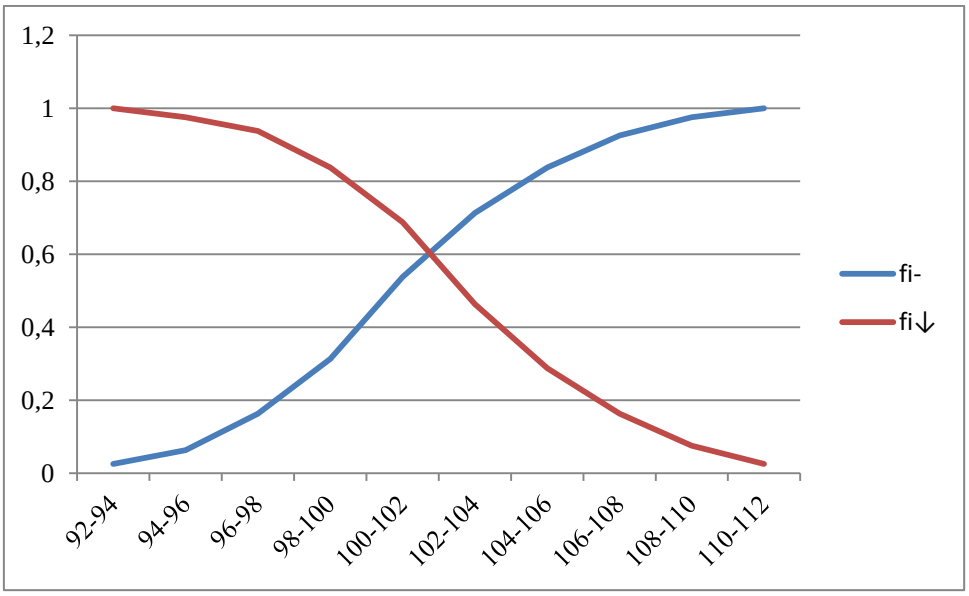


Figure : Courbe cumulative croissante (trait bleu) et courbe cumulative décroissante (trait rouge) de la distribution de l'exemple

b) Cas des classes à différents étendues :

- Histogramme : L'histogramme par principe représente une distribution statistique d'une variable continue.

Dans le cas où les classes ne sont pas d'amplitude égale, il faut corriger les effectifs (nic) (ou les fréquences) de façon à ce que les surfaces des rectangles soient proportionnelles aux effectifs.

Pour cela, on va retenir une amplitude de base a_o, (correspond au Plus Grand Diviseur Commun PGDC des amplitudes). On va déterminer le multiple de cette amplitude de base pour chacune des classes : $k_i = a_i/a_o$.

On détermine enfin les effectifs corrigés n_ic, $n_ic = n_i/k_i$. On peut également utiliser les fréquences corrigées, $f_i' = f_i/k_i$. L'ensemble de ces calculs est généralement présenté dans un tableau.

- Le polygone des effectifs cumulés croissants : On obtient le polygone des effectifs cumulés croissants en joignant, par des segments droits, les points ayant pour abscisse les bornes supérieurs des classes et pour ordonnés les effectifs cumulés (ou les fréquences relatives cumulées) croissants correspondant à la classe considérée. Le premier point est (a_0, 0).

- Le polygone des effectifs cumulés décroissants : On obtient le polygone des effectifs cumulés décroissants en joignant, par des segments droits, les points ayant pour abscisse les bornes inférieurs des classes et pour ordonnés les effectifs cumulés (ou les fréquences relatives cumulées) décroissants correspondant à la classe considérée. Le dernier point est (a_k, 0).

Ministry of Higher Education & Scientific Research **Mouloud MAMMERI University of Tizi Ouzou – Algeria** **Faculty of Economics, Business & Management** **Department of Economics** **1st Year LMD - Section E3**		**Academic Year : 2024 – 2025** **Module : Statistics I** **Worksheet 2nd** **Lecturer : G. BELBACHIR** **Email : gouraya.belbachir@ummto.dz**

To access the online answer key, please scan the **QR** code using the anonymous login.	

Exercise 1 :

Les chiffres d'affaires trimestriels dans un magasin de matériel informatique, en fonction de la marque des produits vendus, se répartissent comme suit :

(U.I : 1000€)

Marque	HP	Apple	Toshiba	Samsung	**Total**
Chiffre d'affaires	55	30	15	20	**120**
f_i (%)	45,83	25,00	12,50	16,67	**100**

1. Représenter graphiquement cette distribution statistique par un diagramme circulaire.

Exercise 2 :

La répartition des décès liés à la Covid-19 au Québec selon le sexe est comme suite :

Sexe Vague	**Vague 1**	**Vague 2**	**Vague 3**	**Vague 4**
Femmes (%)	55	50.6	43.6	41.3
Hommes (%)	45	49.4	56.4	58.7

La source : Site web de l'Institut National de Santé Publique du Québec, [En ligne].https://mobile.inspq.qc.ca/covid-19/donnees/age-sexe, [consulté le 20 novembre 2021].

1. Représenter graphiquement cette distribution statistique par un diagramme en barres.

Exercise 3 :

On a relevé le nombre de garçons dans 200 familles de quatre enfants. On a obtenu les résultats suivants :

Nombre de garçons	0	1	2	3	4	**Total**
Nombre de familles	8	42	67	70	13	**200**

1. Représenter cette distribution statistique par un diagramme en bâtons.

2. Représenter cette distribution statistique par une courbe cumulative.

Exercise 4 :

La série ci-dessous donne la répartition de 140 employés selon la prime de fin d'année exprimée en dinar.

Prime	[8000-10000[	[10000-12000[	[12000-14000[	[14000-16000[	[16000-18000[	**Total**
Nombre d'employés	23	30	70	07	10	**140**

1. Représenter graphiquement cette distribution par un histogramme, un polygone et une courbe des fréquences cumulées.

Supposant que les 70 employés sont répartis ainsi :

30 ont une prime de [12000 - 13000[et 40 ont une prime de [13000 - 14000[DA.

2. Représenter cette nouvelle série par un histogramme et un polygone.

~ ... ~ ... ~

Chapitre 4 :
Les paramètres de tendance centrale

Chapitre 4 : Les paramètres de tendance centrale

L'objectif de ce chapitre est de permettre à l'étudiant de dépasser le stade de l'explication graphique des distributions qu'il pourra observer et de présenter un certain nombre de valeurs calculées que l'on nomme paramètres et qui résument cette distribution observée. Ces paramètres qui sont calculés n'intéressent donc que les variables aléatoires quantitatives.

Lorsque l'on étudie une distribution statistique, il est rapidement nécessaire de simplifier le grand nombre des observations par un plus petit nombre de caractéristiques qui résument cette distribution. Chacune de ces caractéristiques doit remplir quatre conditions. Elle doit être objective, elle doit tenir compte de toutes les observations (il peut être utile d'éliminer certaines observations jugées aberrantes par l'analyse graphique), elle doit avoir une signification concrète, aisée à comprendre, elle doit être simple à calculer.

Certaines caractéristiques ont en plus la propriété d'être peu sensible aux fluctuations d'échantillonnage (c'est à dire que deux expérimentateurs effectuant le même travail sur la même population obtiendront les mêmes résultats), et de pouvoir se prêter aux calculs algébriques ultérieurs (c'est à dire de pouvoir servir à un autre calcul algébrique).

Les principaux paramètres de tendance centrale sont le mode, la médiane et la moyenne arithmétique.

Section 1 : Le mode

Section 2 : La médiane

Section 3 : La moyenne arithmétique

Section 1 : Le mode

Le mode ou dominante est la valeur du caractère ayant l'effectif le plus important. Il peut y avoir plusieurs modes. Dans la distribution d'une variable, le mode peut ne pas exister ou ne pas être unique[1].

C'est un paramètre que l'on peut lire directement dans le tableau des effectifs de la série statistique dans le cas de variables aléatoires discrètes.

X = (1, 3, 5, 2, 4, 7) pas de mode

X = (1, 2, 5, 2, 4, 2, 5) a pour mode = 2

X = (2, 7, 5, 2, 5, 8, 9) a pour mode 2 et 5. On parle de distribution bimodale.

Sur un plan graphique, le mode est la valeur de x sur l'axe des abscisses dont l'ordonnée est la plus grande.

1. Le mode pour une variable statistique discrète (VSD) :

De manière algébrique : Il s'agit de repérer dans la colonne (ni) ou (fi) du tableau statistique l'effectif le plus élevé (ou le chiffre le plus élevé), la modalité (xi) (dans la colonne xi) correspondant à celui-ci est le mode.

Exemple :

Soit la distribution suivante :

Xi	10	20	30	40	50	**Total**
ni	14	23	78	43	32	**190**

1. Calculer le mode de cette distribution.

Il suffit de regarder la colonne (ni), le plus grand effectif c'est ni = 78, la modalité correspondante dans la colonne (xi) c'est 30. Donc le Mode = 33.

2. Le mode pour une variable statistique discrète (VSC) :

Pour une variable continue, la classe modale est celle qui correspond au plus grand effectif si toutes les classes ont la même amplitude. Dans le cas où les amplitudes de classes diffèrent, il faut corriger les effectifs (n_{ic}). La classe modale est alors celle qui

[1] L. FOUCAN, Op.cit., P : 9.

représente l'effectif le plus élevé par unité d'amplitude. Donc c'est la classe d'effectif corrigé maximal. Il est en effet nécessaire d'avoir des classes de même largeur.

a) Amplitude des classes constante (ai = C^{te}) :

Dans ce cas la classe la plus dense est celle aussi correspondant au plus grand effectif, car dans ce cas di = ni/ai revient à diviser tous les effectifs par la même amplitude, donc la classe qui a le plus grand effectif sera aussi la plus dense.

De manière algébrique, la détermination du mode suit les étapes suivantes :

- Repérer dans la colonne (n_i) ou (f_i) la plus grande valeur (le plus grand effectif),

- Repérer la classe modale correspondant au plus grand effectif,

- Appliquer la formule du mode réservée au cas continue :

$$\mathbf{Mo = X_o + a \left(\frac{(n_{mo} - n_{mo-1})}{(n_{mo} - n_{mo-1}) + (n_{mo} - n_{mo+1})} \right)}$$

Avec :

- X_o = borne inférieure de la classe modale.
- a = amplitude la classe modale ou des classes puisque a_i est constante.
- n_{mo} = effectif ou fréquence de la classe modale.
- $n_{mo\ -1}$ = effectif ou fréquence de la classe avant ou précédant la classe modale.
- $n_{mo+}1$ = effectif ou fréquence de la classe après ou suivant la classe modale.

Exemple : (ai = Cte)

Calculer le mode de la distribution suivante :

Les classes	10-20	20-30	30-40	40-50
ni	10	10	15	5

Il faut d'abord, au préalable, calculer les amplitudes et voir si elles sont constantes ou pas. On ajoute alors une colonne (ai) au tableau, on obtient :

Classes	ni	ai
10-20	10	10
20-30	10	10
30-40	15	10
40-50	5	10
Total	40	-

On constate donc que l'amplitude (ai) est constante (ai =10). Par conséquent, la classe modale est celle qui correspond au plus grand effectif. Si on regarde dans la colonne des (ni), la plus grande valeur c'est 15, (ni = 15). La classe correspondant (c'est-à-dire sur la même ligne) c'est [30 - 40[, $\Rightarrow$ c'est la classe modale, donc Xo = 30.

On applique la formule du mode :

$$Mo = Xo + a\left(\frac{(n_{mo} - n_{mo-1})}{(n_{mo} - n_{mo-1}) + (n_{mo} - n_{mo+1})}\right)$$

$$Mo = 30 + 10\left(\frac{(15 - 10)}{(15 - 10) + (15 - 5)}\right) = 33{,}33 \longrightarrow \underline{Mo = 33{,}33}$$

b) Amplitude des classes non constante (ai ≠ C^{te}) :

Dans ce cas la classe la plus dense n'est pas forcément celle qui correspond au plus grand effectif. En effet, di = ni/ai, (ai≠ C^{te}), signifie que l'on divise les effectifs ou fréquences par des valeurs de ai différentes, ce qui implique que l'on peut tomber sur des situations où la classe correspondant au plus grand effectif n'est pas la classe la plus dense.

Aussi, avant de calculer le Mode, il faut comme précédemment, calculer les densités (ou calculer les effectifs corrigés n_{ic}).

La détermination du mode suit les étapes suivantes :

- Calculer les densités (d_i) ou effectifs corrigés (n_{ic}),

- Repérer dans la colonne (d_i) ou (n_{ic}) la plus grande valeur (plus grande densité),

- Repérer la classe modale correspondant à cette plus grande valeur,

- Appliquer la formule du mode réservée au cas continu avec amplitude non constante :

$$\mathbf{Mo = Xo + a}\left[\frac{\mathbf{(n_{icmo} - n_{icmo-1})}}{\mathbf{(n_{icmo} - n_{icmo-1}) + (n_{icmo} - n_{icmo+1})}}\right]$$

Avec :

- Xo = borne inférieure de la classe modale.
- a = amplitude la classe modale.
- n_{icmo} = effectif (ou fréquence) corrigé de la classe modale.
- $n_{icmo\ -1}$ = effectif (ou fréquence) corrigé de la classe avant ou précédant la classe modale.
- $n_{icmo+}1$ = effectif (ou fréquence) corrigé de la classe après ou suivant la classe modale.
- d_{imo} = densité de la classe modale.
- d_{imo-1} = densité de la classe avant ou précédant la classe modale.
- d_{imo+1} = densité de la classe après ou suivant la classe modale.

Exemple : (ai ≠ Cte)

Calculer le mode de la distribution suivante :

Les classes	10-20	20-30	30-50	50-80
ni	10	20	30	25

Pour calculer le mode, il faut d'abord vérifier les amplitudes (ai). On construit comme précédemment une colonne (ai) de laquelle dépendra le nombre de colonnes à ajouter au tableau, selon que ai soit C^{te} ou pas.

Classes	ni	ai	n_{ic}
10-20	10	10	10
20-30	20	10	20
30-50	30	20	15
50-80	25	30	8,33
Total	85	-	-

Si on regarde la colonne (a_i), on constate que l'amplitude de classe n'est pas constante. La classe modale est donc la classe la plus dense, c'est-à-dire celle qui a la plus grande densité ou le plus grand effectif corrigé, et qui n'est pas forcément celle qui a le plus grand effectif. Par conséquent, avant de calculer le mode il faut corriger les effectifs, c'est-à-dire ; soit calculer les densités (d_i), soit calculé les effectifs corrigés (n_{ic}).
L'amplitude de base (la plus petite amplitude), d'après le tableau est $a_o = 10$. On en déduit les (n_{ic}). Ou bien, on calcule directement les densités (d_i).

Dans la colonne (di), la plus grande valeur c'est 2. Sur la même ligne, la plus grande valeur correspondante dans la colonne (n_{ic}) doit automatiquement être la plus grande valeur de la colonne (n_{ic}). Si ce n'est pas le cas, cela voudrait dire qu'il y a erreur de calcul quelque part que l'étudiant doit vérifier et corriger.

Dans notre tableau, la densité la plus élevée $d_2 = 2$, correspond également sur la même ligne, dans la colonne (n_{ic}), à la plus grande valeur n_{ic}, $n_{2c} = 20$. La classe correspondante, à savoir ; [20 - 30[, est la classe modale ou la plus dense.

Donc Mo sera :

$$Mo = Xo + a\left[\frac{(n_{icmo} - n_{icmo-1})}{(n_{icmo} - n_{icmo-1}) + (n_{icmo} - n_{icmo+1})}\right]$$

$$Mo = 20 + 10\left[\frac{(20 - 10)}{(20 - 10) + (20 - 15)}\right] = 26,67 \Rightarrow Mo = 26,67$$

Section 2 : La médiane

La médiane elle correspond à la valeur qui sépare la population en deux sous-ensembles d'effectifs égaux. Au regard du diagramme des effectifs cumulés et des fréquences cumulées, la médiane est la valeur de la variable correspondant à la fréquence cumulée 50%.

1. La médiane pour une variable statistique discrète (VSD) :

La valeur médiane est la valeur Me de X telle que, immédiatement à gauche de Me, la fréquence cumulée F(Me) soit inférieure à 50% et, immédiatement à droite de Me, la fréquence cumulée F(Me) soit supérieure à 50%.

- Si le nombre des observations N est **impair**, le rang de la valeur médiane Me est (N+1)/2.

$M_e = x_k$ (La valeur de rang k) telle que $k = \frac{N+1}{2}$

Exemple :

N = 5

Xi = 6 ; 8 ; 4 ; 10 ; 9.

Me(X) = ?

Commencez par classer par ordre (croissant) les valeurs observées (de 1 à N dans l'ordre croissant).

Xi = 4 ; 6 ; 8 ; 9 ; 10.

Calculez le rang de la médiane.

Me = (N+1)/2.

Ici (5+1)/2 = 3.

Le rang de la médiane est 3.

La valeur médiane obtenue est donc 8.

- Si le nombre des observations N est **pair**, la solution réside dans un intervalle médian.

Exemple :

N = 4

Xi = 6 ; 8 ; 4 ; 10.

Me(X) = ?

Commencez par classer par ordre (croissant) les valeurs observées (de 1 à N dans l'ordre croissant).

Xi = 4 ; 6 ; 8 ; 9.

Calculez le rang de la médiane.

Me = 2,5

Les valeurs de la médiane sont donc celles de l'intervalle [6-8[.

Certains statisticiens retiendront le centre de cet intervalle, à savoir la valeur 7.

On résume donc le cas d'une **Variable Statistique Discrète :**

Soit $x_1, x_2, \ldots, x_N$ une série ordonnée de valeurs de la v. s.

- Si N est impaire i. e. N = 2p + 1, alors la médiane est donnée par : $Me = x_{p+1}$.

- Si N est impaire i. e. N = 2p, alors la médiane est donnée par : $Me = \frac{x_{p+1} + x_p}{2}$.

2. La médiane pour une variable statistique continue (VSC) :

Le parcours est ici plus long. A partir d'une lecture sur les fréquences cumulées, on repère d'abord dans quelle classe est située la médiane.

On fait ensuite l'hypothèse d'équipartition. On suppose que les valeurs à l'intérieur de cette classe sont uniformément réparties.

A partir du tableau des fréquences cumulées la médiane correspond à la valeur de la variable aléatoire X telle que F(X=x) = 0,5. On la calcule souvent par interpolation linéaire dans la classe médiane de façon à obtenir $F(M_e) = 0{,}5$

De façon analytique, on calcule précisément :

$$M_E = x_i + A_i\left(\frac{0{,}5 - F(x_{i-1})}{F(x_i) - F(x_{i-1})}\right)$$

Avec :

i : Représentant la classe médiane.

x_i : La borne inférieure de la classe médiane.

A_i : L'amplitude de la classe médiane.

$F(x_{i-1})$: La fréquence cumulée de la classe précédent la classe médiane.

$F(x_i)$: La fréquence cumulée de la classe médiane.

La qualité statistique essentielle de la médiane est qu'elle est peu sensible aux aléas d'échantillonnage, en particulier aux erreurs d'échantillonnage.

En effet comme il s'agit d'un paramètre de rang, une valeur aberrante ne la modifiera pas. On dit que c'est un paramètre robuste.

Son inconvénient principal est d'être peu exploitable pour des calculs ultérieurs. Si l'expérimentateur a oublié un certain nombre de valeurs, il est nécessaire de tout recalculer.

Graphiquement, en utilisant la courbe des fréquences cumulées ascendantes, la médiane est le point d'abscisse qui correspond à 50% en ordonnée.

Cependant, par la méthode graphique il n'est pas toujours aisé de déterminer avec précision la valeur de la médiane.

3. Les quantiles :

a) Les quartiles :

Les quartiles, notés Q_1, Q_2, Q_3, sont les valeurs ordonnées de la variable qui partagent les valeurs de la population N en quatre sous-ensembles d'effectifs égaux de 25%.

Ils correspondent donc aux fréquences cumulées 25%, 50%, 75% du diagramme de fréquences cumulées.

- Le premier quartile Q_1 est la valeur de la variable située au quart de l'effectif.
- Le deuxième quartile est la médiane.
- Le troisième quartile Q_3 est la valeur de la variable située aux trois quarts de l'effectif.

On calcule les quartiles de la même façon que la médiane, par interpolation linéaire à l'intérieur de la classe qui les contient.

Exemple :

Nous allons analyser le nombre d'appels que reçoit un central téléphonique. Nous disposons d'un tableau de données contenant le nombre d'appel par heures.

Classe (nombre d'appels par heure compris entre … et …)	**Nombre d'heures**
[10 ; 20[	10
[20 ; 30[	40
[30 ; 50[	140
[50 ; 90[	220
[90 ; 100[	10
Total	**420**

Il s'agit d'une variable quantitative continue (VSC).

- Détermination de la médiane :

Il faut calculer les fréquences ou les effectifs cumulés. Nous avons choisi de calculer la médiane à partir des fréquences cumulées, la médiane étant la valeur de la variable aléatoire pour $F_i = 0.5$.

Le calcul de la médiane se fait par interpolation linéaire à l'intérieur de la classe médiane par application de la formule suivante :

$$M_E = x_i + A_i\left(\frac{0{,}5 - F(x_{i-1})}{F(x_i) - F(x_{i-1})}\right)$$

Dans laquelle :

x_i : La borne inférieure de la classe médiane est égale à 50

A_i : L'amplitude de la classe médiane est égale à 40 (dans le tableau nous avons présenté les classes unités c'est-à-dire 40/10 = 4)

$F(x_{i-1})$: La fréquence cumulée de la classe précédente la classe médiane est égale à 0,45

$F(x_i)$: La fréquence cumulée de la classe médiane est égale à 0,98.

Ce qui donne comme résultat 53,77 appels par heure.

La médiane peut aussi être obtenue par application de la règle de proportionnalité suivante :

$M_E = 50 + \left(\frac{(50-45)(90-50)}{98-45} \right) = 53{,}77$ appels par heure.

b) Les déciles :

De la même façon que les quartiles divisent la distribution en quatre parties d'effectifs égaux, les déciles divisent la distribution en 10 parties d'effectifs égaux.

On les calcule par interpolation linéaire dans la classe qui les contient, comme pour la médiane.

c) Les centiles :

De la même façon que les quartiles divisent la distribution en quatre parties d'effectifs égaux, les déciles divisent la distribution en 10 parties d'effectifs égaux et les centiles en cent parties d'effectifs égaux.

On les calcule par interpolation linéaire dans la classe qui les contient, comme pour la médiane.

Section 3 : La moyenne arithmétique

1. Définition de la moyenne arithmétique :

La moyenne arithmétique d'une variable X, généralement notée $\overline{X}$, s'obtient en faisant la somme des valeurs de la variable pondérées par leurs effectifs et en la divisant par le nombre d'individus.

C'est un caractère de position essentiel car elle permet des calculs ultérieurs. En particulier, elle sert au calcul de la variance. Par contre c'est un paramètre très sensible aux variations d'échantillonnage.

▪ Soit X une variable quantitative discrète (VSD) définie sur une population de N individus ;

- $x_1, \ldots, x_i$, les valeurs distinctes prises par X ;
- $n_1, \ldots, n_i$, les effectifs de ces valeurs ;
- $f_1, \ldots, f_i$, leurs fréquences.

On a les relations suivantes : $f_i = \frac{n_i}{N}$, $\sum n_i = N$, $\sum f_i = 1$

Selon que l'on dispose des effectifs ou des fréquences, la formule analytique est la suivante :

$$\overline{X} = \frac{\sum_i n_i x_i}{N} = \sum_i f_i x_i$$

▪ Pour une variable continue (VSC), ce calcul se fait à l'aide des centres de classes :

$$\overline{X} = \frac{\sum_i n_i c_i}{N} = \sum_i f_i c_i \quad avec \; c_i = \left(\frac{e_i + e_{i-1}}{2}\right)_i$$

C'est-à-dire que c'est le barycentre de la distribution.

Pour trouver la valeur de la moyenne, on rajoute une colonne dans le tableau statistique dont laquelle on calcule le produit $n_i x_i$, pour tout i. Si la VS est continue (VSC), les valeurs x_1, x_2, …, x_k représentent les centres de classes.

2. Les propriétés de la moyenne :

- La moyenne ne change pas si on remplace un nombre déterminé de valeurs par leur moyenne multipliée par la somme de leurs effectifs.
- La moyenne conserve les changements de l'axe et l'origine

$$X(x_i, n_i) \to Y(y_i = ax_i + b, n_i)$$
$$\bar{x} \mapsto \bar{y} = a\bar{x} + b$$

- On peut réaliser un changement d'origine et/ou d'échelle pour simplifier les calculs

Changement d'origine : (méthode de la moyenne provisoire) [1]

Soit la variable $X' = X - x_0$

On démontre que : $\overline{X}' = \overline{X} - x_0$ $\qquad$ $\overline{X} = \overline{X}' + x0$

On a intérêt à choisir x_0 de manière à obtenir une simplification des calculs et donc des valeurs très petites de X'. Il faut choisir de préférence le mode.

Changement d'échelle :

$X' = \frac{X}{h}$ $\qquad$ $\overline{X}' = \frac{\overline{X}}{h}$ $\qquad$ $\overline{X} = h\ \overline{X}'$

Changement d'origine et Changement d'échelle :

$X' = \frac{X - x_0}{h}$ $\qquad$ $\overline{X}' = \frac{\overline{X} - x_0}{h}$ $\qquad$ $\overline{X} = h\ \overline{X}' + x_0$

- Autre propriété : La somme algébrique des écarts à la moyenne est nulle.

[1] L. FOUCAN, Op.cit., P : 8.

3. Démonstration de la formule :

$$\overline{X} = \frac{\sum_i n_i x_i}{N} = \sum_i f_i x_i$$

car $\frac{n_i}{N} = f_i$ la fréquence observée de la classe de rang i.

Dans le cas d'une variable aléatoire continue, si on ne dispose que d'une représentation par classe de la distribution, il faut attribuer au centre de chaque classe, c_i = xi, la totalité de l'effectif de la classe puisqu'on ne connaît pas la répartition exacte de l'effectif à l'intérieur de la classe.

Analyse du nombre d'appels que reçoit un central téléphonique (suite).

4. Détermination de la moyenne arithmétique :

Pour calculer la moyenne arithmétique, il faut d'abord calculer le centre des classes, puis utiliser la formule :

$$\overline{X} = \frac{\sum_i n_i x_i}{N} = \sum_i f_i x_i$$

Ou xi (ci) le centre de la classe de rang i remplace xi.

x_i	n_i	c_i	$n_i c_i$
[10 ; 20[	10	15	150
[20 ; 30[	40	25	1000
[30 ; 50[	140	40	5600
[50 ; 90[	220	70	15400
[90 ; 100[	10	95	950
Total	**420**	**-**	**23100**

Au total nous obtenons $\overline{X} = m = \frac{23100}{420} = 55$ appels par heure.

Chapitre 5 :

Les paramètres de dispersion

Chapitre 5 : Les paramètres de dispersion

Les données que nous avons recueillies ont été présentées par un graphique et résumées par un certain nombre de paramètres de tendance. Mais nous restons interpellés par le fait que deux graphiques distincts mais ayant des étalements différents (des dispersions autour de la tendance centrale différentes) seront résumées par des paramètres de tendance identique. Cela n'est pas satisfaisant.
Prenons l'exemple de deux échantillons issus de deux entreprises dont on souhaite étudier les salaires. Les caractères de dispersion vont nous permettre de quantifier cette différence.

Ils ont d'autres usages comme par exemple de mesurer la qualité d'un échantillonnage ou de permettre de réaliser des tests statistiques
Les principaux paramètres de dispersion sont l'étendue, l'écart interquartile, l'écart inter décile, la variance et l'écart-type et enfin l'intervalle de confiance.

Section 1 : La dispersion dans un intervalle (Les écarts simples)
Section 2 : La dispersion autour d'une valeur centrale (Les écarts moyens)
Section 3 : La comparaison des dispersions des séries statistiques

Section 1 : La dispersion dans un intervalle (Les écarts simples)

1. L'étendue :

a) Définition[1] :

L'étendue est la quantité $x_{max} - x_{min}$. C'est-à-dire la différence entre les valeurs extrêmes de la variable $e = x_{max} - x_{min}$. Elle permet de mettre en évidence des valeurs aberrantes.

b) Propriétés :

- L'étendue est facile à calculer.
- Elle ne tient compte que des valeurs extrêmes de la série ; elle ne dépend ni du nombre, ni des valeurs intermédiaires ; elle est très peu utilisée dès que le nombre de données dépasse 10.

Elle est utilisée en contrôle industriel où le nombre de pièces prélevées dépasse rarement 4 ou 5 ; elle donne une idée appréciable de la dispersion. Cependant, dès que cela est possible, on préfère prélever 15 à 20 unités et utiliser l'écart-type pour apprécier la dispersion.

2. L'écart interquartile :

C'est la différence entre les deux quartiles extrêmes $Q_3 - Q_1$. Il comprend 50% des mesures les plus médianes.

3. L'écart inter décile :

De façon à pallier à l'inconvénient qui consiste à laisser de côté 50% des mesures on peut utiliser l'écart interdécile. Celui-ci consiste en la différence entre le dernier et le premier décile : $D_9 - D_1$. Il ne laisse de coté que 20% des mesures.

L'écart interquartile et l'écart interdécile servent à calculer des indices de forme de distribution.

[1] Renée VEYSSEYRE, op.cit., P : 22.

Section 2 : La dispersion autour d'une valeur centrale (Les écarts moyens)

1. L'écart-type :

Nous avons deux séries de données :

Série 1	1100 €	1300 €	1600 €	2000 €
Série 2	1300 €	1400 €	1600 €	1700 €

Les paramètres précédents ont l'inconvénient de ne pas être faciles à calculer et surtout, en cas de nouvelles mesures, de nécessiter de refaire l'ensemble des calculs. Une possibilité est de synthétiser les écarts à la moyenne ou à la médiane.

a) Ecart moyen ou écart médian :

On peut utiliser les écarts absolus à la moyenne (ou à la médiane) et calculer la moyenne des écarts absolus à la moyenne (ou à la médiane) :

$$e_m = \frac{1}{N}\sum_{i=1}^{i=p} n_i |x_i - m|$$

.

Malgré les ordinateurs, le calcul avec des valeurs absolues n'est pas très prisé. Pourtant, cet indice à l'avantage d'être croissant quelque soit $(x_i - m)$.

En effet, la somme des écarts à la moyenne est nulle alors que la somme des écarts absolus à la moyenne est croissante.

Démonstration :

$$\sum_i (x_i - m) = \sum_i (x_i) - \sum_i m = \sum_i x_i - Nm = \sum_i x_i - N\frac{\sum_i x_i}{N} = \sum_i x_i - \sum_i x_i = 0$$

En effet, les écarts tantôt positifs, tantôt négatifs s'annulent entre eux. Il n'est donc pas possible d'utiliser un tel indice et il a donc fallu utiliser un indice qui s'accroît systématiquement quand N grandit.

b) Ecart quadratique moyen et moyenne quadratique :

La fonction élévation au carré présente cet avantage. Elle est continuellement croissante et facilement dérivable (donc facile à utiliser en pratique contrairement à la valeur absolue).

C'est la quantité $(x_i - m)^2$ qui est continuellement croissante quelque soit $(x_i - m)$.

c) La variance :

La variance est égale à la moyenne des carrés des écarts à la moyenne, c'est-à-dire :

$$\text{VAR (X)} = \boxed{\frac{1}{N}\sum_i n_i\left(x_i - \overline{X}\right)^2 = \sum_i f_i\left(x_i - \overline{X}\right)^2}$$

$$\text{VAR (X)} = \boxed{\frac{1}{N}\sum_i n_i {x_i}^2 - \overline{X}^2 = \sum_i f_i {x_i}^2 - \overline{X}^2}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{i=N}(x_i - m)^2$$

Dans le cas des variables continues présentées par classe, il faut utiliser le centre des classes c_i au lieu des x_i.

Dans son expression :

$$\sigma^2 = \frac{1}{N}\left(N\sum_{i=1}^{i=N} x_i^2 - \left(\sum_{i=1}^{i=N} x_i\right)^2\right)$$

On voit apparaître la moyenne quadratique et la moyenne arithmétique. Mais la façon pratique de la calculer est d'utiliser l'expression :

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{i=N} x_i^2 - m^2$$

- Démonstration :

Ecrivons la variance comme la moyenne des écarts carrés à la moyenne :

$\sigma^2 = \frac{1}{N}\sum_{i=1}^{i=N}(x_i - m)^2$ et développons $\sum_{i=1}^{i=N}(x_i - m)^2$.

$$\sum_{i=1}^{i=p}(x_i - m)^2 = \sum\left(x_i^2 - 2mx_i + m^2\right) = \sum x_i^2 - 2\sum mx_i + Nm^2$$

Or $m = \frac{\sum x_i}{N}$ d'où $Nm = \sum x_i$ alors

$\sum(x_i^2 - 2mx_i + m^2) = \sum x_i^2 - 2\sum mx_i + Nm^2 = \sum x_i^2 - m\sum x_i$ d'où la variance : $\sigma^2 = \frac{\sum x_i^2}{N} - m\frac{\sum x_i}{N} = \frac{\sum x_i^2}{N} - m^2$ qui est la formule calculatoire.

- Propriétés de la variance :

- La variance est toujours positive ou nulle.
- Changement d'échelle et d'origine.

$$X(x_i, n_i) \to Y(y_i = ax_i + b, n_i)$$
$$V_X \mapsto V_Y = a^2 V_X$$

d) L'écart-type :

L'écart type est égal à la racine carrée de la variance, et est donc mesuré dans la même unité que la variable X.

$\sigma = \sqrt{VAR(X)} = (VAR(X))^{1/2}$

$\sigma = \sqrt{\frac{\sum x_i^2}{n_i} - m^2}$

Résolvons notre exemple :

Série 1	1100 €	1300 €	1600 €	2000 €
Série 2	1300 €	1400 €	1600 €	1700 €

Série 1	**Série 2**	**Série 1**	**Série 2**	**Série 1**	**Série 2**
x_i	x_i	$(x_i - m)$	$(x_i - m)$	$(x_i - m)^2$	$(x_i - m)^2$
1 100,00 €	1 300,00 €	-400,00 €	-200,00 €	160 000,00 €	40 000,00 €
1 300,00 €	1 400,00 €	-200,00 €	-100,00 €	40 000,00 €	10 000,00 €
1 600,00 €	1 600,00 €	100,00 €	100,00 €	10 000,00 €	10 000,00 €
2 000,00 €	1 700,00 €	500,00 €	200,00 €	250 000,00 €	40 000,00 €
Somme		**0,00 €**	**0,00 €**	**460 000,00 €**	**100 000,00 €**
Variance				**115 000,00 €**	**25 000,00 €**
Ecart-type				**339,12 €**	**158,11 €**

Les moyennes m sont les mêmes.

En effet (1100 + 1300 + 1600 + 2000)/4 = 1500

Pour la première série et (1300 + 1400 + 1600 + 1700)/4 = 1500.

La somme des $(x_i - m)$ = 0. Les sommes quadratiques sont d'autant plus importantes que les $(x_i - m)$ sont plus grandes en valeur absolue.

Exemple :

Calculons quelques caractères de dispersion de la série suivante (il s'agit de la température en degrés Celsius relevées dans un entrepôt pendant un an). Les données sont présentées par intervalles.

Classes	n_i
-2 à 0	7
0 à 2	10
2 à 4	12
4 à 6	13
6 à 8	12
8 à 10	11
10 à 15	2

- Complétons d'abord le tableau de données :

Classes	c_i	n_i	$n_i c_i$	$n_i c^2{}_i$	f_i	F_i
-2 à 0	-1	7	-7	7	0,11	0,11
0 à 2	1	10	10	10	0,15	0,26
2 à 4	3	12	36	108	0,18	0,45
4 à 6	5	13	65	325	0,20	0,65
6 à 8	7	12	84	588	0,18	0,83
8 à 10	9	11	99	891	0,17	1,00
10 à 15	12,5	2	25	312,5	-	-
Totaux	**-**	**67**	**312**	**2241,5**	**1**	**-**

- La moyenne : est donnée par $m = \frac{\sum n_i c_i}{\sum n_i} = \frac{312}{67} = 4{,}66°$.

En moyenne les températures relevées dans cet entrepôt sont de 4,66°.

- La variance : est donnée par $\frac{\sum n_i c_i^2}{\sum n_i} - m^2 = \frac{2241{,}5}{67} - 4{,}66^2$ = 11.73.

- L'écart-type : est donné par $\sigma = \sqrt[2]{\sigma^2} = 3.43°$.

- A titre d'exercice : vous pouvez vérifier que : Q1 = 1,95° , que Q3 = 7,375° et que l'intervalle interquartile est de 5,425°.

e) Propriétés de l'écart-type :

L'écart-type réunit un grand nombre des qualités d'un paramètre de dispersion. Il se prête facilement au calcul algébrique, il est intuitif qu'il est d'autant plus grand que la dispersion est plus grande, il possède une unité de mesure, etc.

Dans le cadre d'une distribution « normale » ou proche de la normale, on note que :

- 50% des valeurs sont situées dans l'intervalle [m-2/3σ ; m+2/3σ],
- 70% des valeurs environ sont situées dans l'intervalle [m-σ ; m+σ],
- et enfin que 95% environ des valeurs sont situées dans l'intervalle [m-2σ ; m+2σ].

2. L'intervalle de confiance :

Si X la variable aléatoire suit une loi normale de moyenne m et d'écart type σ, alors l'intervalle de confiance à 95% est donné par :

IC 95% = [m-1,96σ ; m+1,96σ].

Nous utiliserons 2 par approximation de 1,96 qui est la valeur lue dans la table de la Loi Normale centrée réduite pour Fi ≤ 0,975 (nous le démontrerons ultérieurement). C'est un paramètre très utilisé.

Section 3 : La comparaison des dispersions des séries statistiques

1. Le coefficient de variation :

Etant donné que l'écart type d'une distribution s'exprime dans la même unité de mesure que les observations, il peut s'avérer intéressant de disposer d'un paramètre qui soit indépendant des unités de mesure. Cela permet de pouvoir comparer des dispersions de deux distributions qui diffèrent soit par les unités employées, soit par leur nature. Le coefficient de variation est le quotient de l'écart type par la moyenne.

$$Cv_X = 100 \times \frac{\sigma_X}{\bar{x}}$$

Le coefficient de variation est un nombre sans dimension, indépendant de l'unité de mesure. Il est souvent exprimé en pourcentages (%) sans que cela soit une obligation.

2. Propriétés[1] :

a) Le coefficient de variation ne dépend pas des unités choisies.

b) Il permet d'apprécier la représentativité de la moyenne arithmétique x par rapport à l'ensemble des données.

c) Il permet d'apprécier l'homogénéité de la distribution, une valeur du coefficient de variation inférieure à 15 % traduit une bonne homogénéité de la distribution.

d) Il permet de comparer deux distributions, même si les données ne sont pas exprimées avec la même unité ou si les moyennes arithmétiques des deux séries sont très différentes.

3. Interprétation :

Plus le coefficient se rapproche de zéro, meilleur est le résultat. A l'extrême, si l'écart-type était nul, on aurait CV = 0.

Le coefficient permet de comparer les dispersions des séries statistiques qui ne sont pas exprimée dans les mêmes unités de mesure ou des séries ayant des moyennes très différentes.

[1] Renée VEYSSEYRE, op.cit., P : 21.

Exemple :

On souhaite comparer la dispersion des tailles et poids d'un groupe d'étudiants.

Variable taille exprimée en cm : moyenne = 180,4 cm ; écart-type : 5,1 cm.

Variable poids exprimée en kg : moyenne = 72,2 kg ; écart-type : 9,0 kg.

Comment comparer 5,1 cm et 9,0 kg les dispersions respectives de la taille et du poids des étudiants ?

Solution :

$$CV_{(Taille)} = \frac{\sigma}{\mu} = \frac{5{,}1}{180.4} = 2.83\%$$

$$CV_{(Poids)} = \frac{\sigma}{\mu} = \frac{9.0}{72.2} = 12.47\%$$

La dispersion des poids autour de la moyenne 72,2 kg est plus forte que la dispersion des tailles autour de la moyenne 180,4 cm.

Chapitre 6 :
Les paramètres de forme

Chapitre 6 : Les paramètres de forme

Pour résumer et caractériser les séries statistiques, nous avons jusque-là analysé :

- Les paramètres de tendance centrale, qui donnent une idée de la grandeur de la série,

- Les caractéristiques de dispersion qui mesurent l'intensité de l'éloignement ou de la fluctuation des modalités autour d'une valeur centrale.

Nous avons ainsi à notre disposition une multitude de paramètres qui nous renseignent sur l'allure générale de la série et qu'on retrouve en traçant la courbe des fréquences.

Section 1 : Mesure de la symétrie

Section 2 : Mesure de l'aplatissement

Section 1 : Mesure de la symétrie

1. Distribution symétrique :

Une distribution est symétrique si les valeurs de la variable statistique sont également distribuées de part et d'autre d'une valeur centrale.

Pour une distribution symétrique : mode = médiane = moyenne arithmétique.

2. Coefficient d'asymétrie ou de dissymétrie ou skewness :

$$\gamma_1 = \frac{\mu_3}{s^3} \quad \text{où} \quad \mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3$$

Selon la valeur de ce coefficient, on peut donner quelques caractéristiques sur la forme de la distribution[1]:

– si $\gamma_1 > 0$, la distribution est étalée vers la droite,

– si $\gamma_1 < 0$, la distribution est étalée vers la gauche,

– si $\gamma_1 = 0$, on ne peut pas conclure que la distribution est symétrique mais la réciproque est vraie.

3. Synthèse graphique :

Les positions relatives du mode, de la médiane et de la moyenne arithmétique nous renseignent sur l'allure de la distribution.

a) Si la distribution est symétrique et uni modale, alors le mode, la médiane et la moyenne sont confondus.

b) En revanche, si la moyenne est inférieure à la médiane elle-même inférieure au mode alors la distribution est dissymétrique avec étalement à gauche.

c) Si la moyenne est supérieure à la médiane elle-même supérieure au mode alors la distribution est dissymétrique avec étalement à droite.

[1] Renée VEYSSEYRE, op.cit., P : 22-23.

Section 2 : Mesure de l'aplatissement

1. Coefficient d'aplatissement ou kurtosis :

$$\gamma_2 = \frac{\mu_4}{s^4} \quad \text{où} \quad \mu_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4$$

Le coefficient d'aplatissement est égal à 3. Selon la valeur de ce coefficient, on peut donner quelques caractéristiques sur la forme de la distribution[1]:

– si $\gamma_2 > 3$, la distribution est moins aplatie qu'une distribution gaussienne,

– si $\gamma_2 < 3$, la distribution est plus aplatie qu'une distribution gaussienne.

[1] Renée VEYSSEYRE, op.cit., P : 23.

Ministry of Higher Education & Scientific Research
Mouloud MAMMERI University of Tizi Ouzou – Algeria
Faculty of Economics, Business & Management
Department of Economics
1st Year LMD - Section E3

Academic Year : 2024 – 2025
Module : Statistics I
Worksheet 3rd
Lecturer : G. BELBACHIR
Email : gouraya.belbachir@ummto.dz

To access the online answer key, please scan the **QR** code using the anonymous login.

Exercise 1 :

Calculer le mode, la médiane et la moyenne arithmétique des séries suivantes :

1. Série 1 : 10 ; 7 ; 8 ; 11 ; 6 ; 7 ; 9.

2. Série 2 : 84 ; 91 ; 72 ; 68 ; 87 ; 78.

Exercise 2 :

Soient deux séries statistiques. L'une composée de 240 observations et l'autre de 315 observations.

1. Trouver la position de la valeur médiane dans chacune d'elles.

Exercise 3 :

Dans une classe de 50 élèves le poids moyen des élèves est de 55,7 Kg. Le poids moyen chez les garçons est de 63, 4 Kg, tandis que chez les filles il est de 51,8 Kg.

1. Déterminer les proportions des garçons et des filles ainsi que leurs effectifs.

Exercise 4 :

La répartition des bureaux dans une administration publique en fonction du nombre d'employés est donnée par le tableau suivant.

Nombre d'employés par bureau	0	1	2	3	4	5
Nombre de bureaux	3	7	25	18	12	8

1. Calculer le mode, la médiane et la moyenne arithmétique de cette distribution statistique.

2. Déterminer graphiquement le mode et la médiane.

Exercise 5 :

Le CHU de Tizi Ouzou étudie la létalité du variant Delta de la COVID-19, pour les personnes âgées de 10 à 50 ans. Le tableau suivant indique le nombre de fatalités de la région Tizi Ouzou centre par tranches d'âge du mois août 2021 :

Classe d'âge (an)	10 - 15	15 - 25	25 - 30	30 - 45	45 - 50
Nombre de décès	4	8	5	12	24

1. Déterminer la population et la variable.

2. Préciser la nature de la variable et les déférentes modalités.

3. Calculer les indicateurs de tendance centrale en explicitant vos calculs.

4. Déterminer le mode et la médiane graphiquement.

5. Recalculer la moyenne arithmétique en utilisant la méthode de changement de variable.

6. Déterminer Q_1 ; Q_2 ; Q_3 ; D_1 ; D_5 ; C_4 et C_{50} et déterminer l'intervalle interquartile.

7. La série étudiée est-elle symétrique ou asymétrique ? justifier votre réponse. Pouvait-on prévoir ce résultat ?

~ ... ~ ... ~

Ministry of Higher Education & Scientific Research **Mouloud MAMMERI University of Tizi Ouzou – Algeria** **Faculty of Economics, Business & Management** **Department of Economics** **1st Year LMD - Section E3**		**Academic Year : 2024 – 2025** **Module : Statistics I** **Worksheet 4th** **Lecturer : G. BELBACHIR** **Email : gouraya.belbachir@ummto.dz**

To access the online answer key, please scan the **QR** code using the anonymous login.	

Exercise 1 :

Soit la distribution statistique suivante.

X_i	2	3	8	12	17
n_i	2	2	3	3	1

1. Calculer la moyenne arithmétique ($\bar{X}$), la moyenne géométrique (G) et la moyenne harmonique (H).

2. Vérifier que $\bar{X} > G > H$.

Exercise 2 :

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

Années	1994	1995	1996	1997	1998	1999
Pourcentage de variation	-3,8	-18,3	+14,2	-3,1	-1,5	+8,9

1. Calculer le taux annuel moyen de variation au cours de la période 1993-1999.

Sachant que le taux moyen d'accroissement des exportations était de 9,2% par an pendant la période 1969-1982, de 12,4% pendant la période 1982-1993.

2. Calculer le taux de variation annuel moyen pour la période 1969-1999.

Exercise 3 :

Un courtier réalise les opérations financières suivantes :

▪ Opération 1 : achat pour 20.000 $ d'actions au cours de 20 $ l'action ;

▪ Opération 2 : achat pour 35000 $ d'actions au cours de 10 $ l'action ;

▪ Opération 3 : achat pour 15000 $ d'actions au cours de 30 $ l'action.

1. Quel est le cours moyen de l'action subi par le courtier sur l'ensemble des trois opérations ?

L'année suivante, ce même courtier a réalisé les opérations suivantes :

▪ Opération 1 : achat de 800 actions au cours de 30 $ l'action.

▪ Opération 2 : achat de 1200 actions au cours de 25 $ l'action.

2. Quel est le cours moyen de l'action subi par le courtier sur les deux opérations?

~ ... ~ ... ~

Chapitre 7 :
Les paramètres de concentration

Chapitre 7 : Les paramètres de concentration

Lors d'une étude de variable continue, telle que le revenu, il est intéressant de connaître le degré de concentration de cette variable dans la population. Par exemple, il nous importe ici de pouvoir répondre à la question : quelle est la part de la population de l'entreprise bénéficiaire de 10 %, 80 % du revenu total ? ou à l'inverse, quelle est la part de la population disposant des 20 % de revenus les plus élevés ?
Pour ce faire, le statisticien dispose de deux instruments, un indicateur chiffré, dit indice de GINI et une représentation graphique dite courbe de LORENTZ.

Section 1 : L'analyse algébrique de la concentration
Section 2 : L'analyse graphique de la concentration

Section 1 : L'analyse algébrique de la concentration

1. L'indice de GINI* :

L'indice de Gini (IG), ou indice de concentration, est le double de l'aire située entre la courbe de concentration décrite précédemment et la diagonale du carré défini par F_i =1 et G_i =1.

Ainsi, l'indice de Gini est compris entre 0 et 1.

Un indice proposé par Gini est le suivant :

$$G = \text{aire ODBC} - \text{aire ODBA}$$

$$G = U + L$$

L'indice de Gini est égal au double de l'aire comprise entre la courbe de concentration et la première bissectrice.

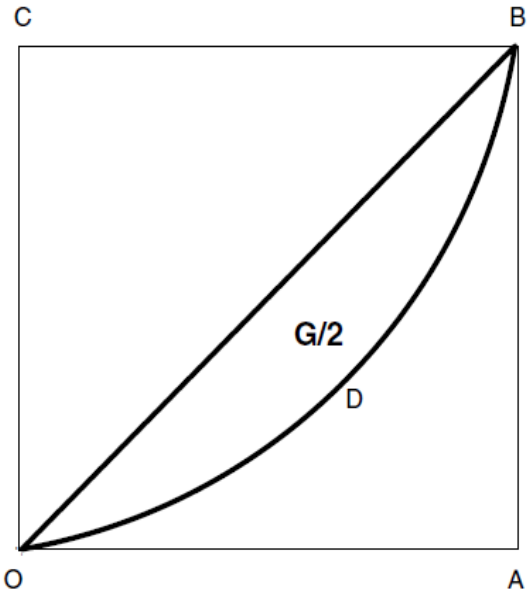


Figure : Courbe de concentration et indice de Gini.

Cet indice est donné par l'intégrale double où *f* est la densité de la loi de la variable *X* et *m* son espérance mathématique :

$$G = \frac{1}{2m} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y| \, f(x) f(y) \, \mathrm{d}x \, \mathrm{d}y$$

* Économiste Italien né en 1884.

Pour un échantillon de taille *n*, l'indice de Gini est calculé de la façon suivante :

$$G = \frac{1}{2n^2 \bar{y}} \sum_i \sum_j |y_i - y_j|$$

avec $\bar{y}$: revenu moyen de la population totale ; $y_i et y_j$: revenu des individus i et j

2. Méthode opératoire :

La méthode de calcul la plus simple vise à estimer l'aire U située au-dessous de la courbe de concentration et de déduire ensuite L, sachant que :

G = U + L = la moitié de l'aire du carré de côté 1= 0,5.

Pour cela, on assimile la courbe de concentration à des segments de droite dont on calcule l'aire : méthode des trapèzes (aire d'un trapèze = hauteur * demi somme des bases).

$$U_i = f_i(G_{e(i-1)} + G_{e(i)})/2$$

puis :

$$U = \sum U_I \; ; L = 0{,}5 - U \text{ et } IG = 2 \text{ x } L$$

3. Interprétation :

a) L'Indice de Gini est égal à 0 quand l'inégalité de répartition est nulle : Alors, la courbe de concentration est confondue avec la droite d'équirépartition. *La distribution est parfaitement égalitaire*. 60% de la population cumulée perçoit 60 % du revenu cumulé, etc.

b) L'Indice de Gini est égal à 1 quand l'inégalité de répartition est maximale : Alors, la courbe de concentration est confondue avec le bord droit du carré. Un individu concentre l'ensemble du revenu cumulé de toute la population. Tous les autres ne perçoivent rien !

Section 2 : L'analyse graphique de la concentration

1. La courbe de LORENTZ :

Soit une distribution de consommation *X* de masse totale *M*. À chaque valeur *xi* de la variable *X*, on associe le point qui a :

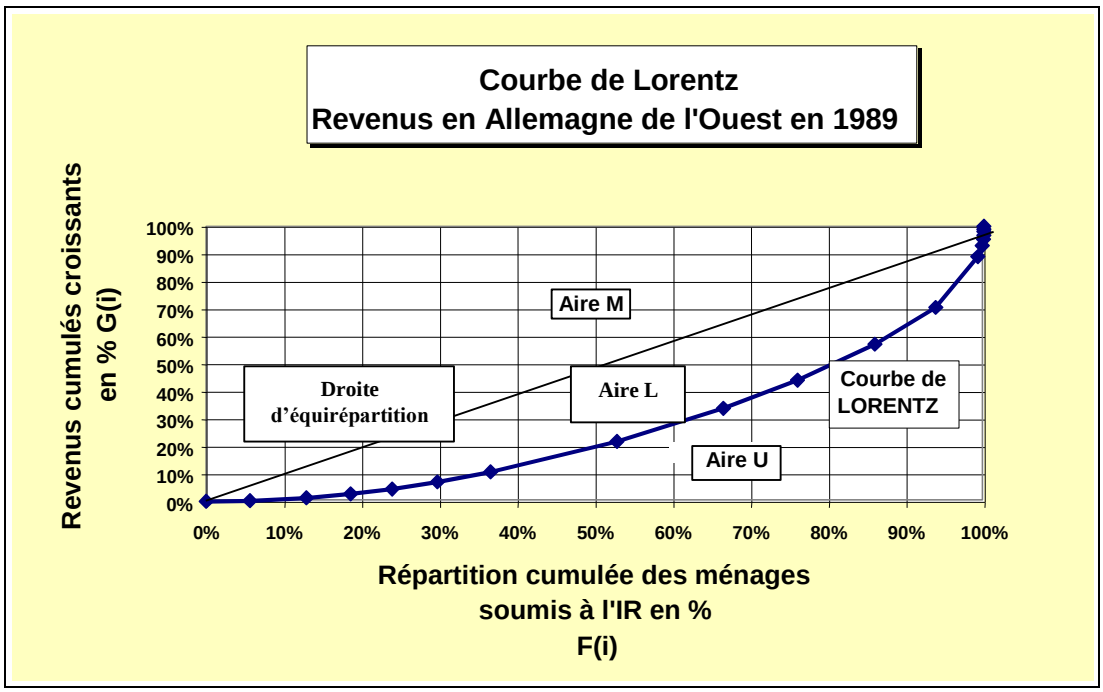
- Pour abscisse *F*(*xi*) = Proportion des individus consommant moins de *xi*
- Pour ordonnée *G* (*xi*) = $\frac{\text{Masse des consommations} < xi}{\text{Masse totale}}$

Pour une distribution non uniforme, cette courbe est toujours en dessous de la première bissectrice ; en effet, *F*(*xi*) est la proportion des individus consommant moins de *xi* ; ils ne peuvent pas globalement consommer autant que les 100 *F*(*xi*) % suivants donc *G*(*xi*) < *F*(*xi*).

La courbe de concentration traduit le pourcentage des individus consommant moins de *xi* à la contribution de ces individus à la moyenne *x* de la masse totale.

Cette courbe met en rapport, sur un espace cartésien, deux distributions cumulées croissantes :

En abscisse, la distribution des effectifs cumulés de la population (F_i), en ordonnées, la distribution des fréquences cumulées de la variable étudiée (G_i)



Exemple :

Une distribution des salaires, appliquée en général à l'étude des revenus, cette courbe met en rapport la distribution des effectifs cumulés (ou fréquences $F(e_i)$) avec les masses de revenus cumulés ($G(e_j)$).

Nous vous proposons, sous tableur, une **méthode de calcul en trois temps** :

a) Calculez d'abord la **proportion de la population** gagnant moins de e_i (fréquences cumulées) :

$$F_{ei} = \frac{\sum_{j=1}^{i} n_j}{N} = \sum_{j=1}^{i} f_j \ pour\ i = 1,2,...,k$$

b) Calculez ensuite la **masse salariale totale** gagnée par les salariés qui sont situés en deçà de la borne e_i soit, si on appelle S_j la masse salariale de la classe j, c'est-à-dire le total des salaires de cette classe :

$$G_{ei} = \frac{\sum_{j=1}^{i} S_j}{MS}$$

On peut définir au préalable la série g_i, c'est à dire la part de la masse salariale totale (MS) que reçoit la classe i dans son ensemble soit : $g_i = \frac{S_i}{MS}$ Il est important de cumuler les différentes classes de masse salariale, soit : $G_{ei} = \sum_{j=1}^{i} g_j$

Il est possible d'estimer S_i si cette donnée est absente.

c) Construire enfin **le graphique** avec une courbe de fréquence cumulée G_i en ordonnées et F_i en abscisses. Parce que ce sont des fréquences cumulées, elles varient de 0 à 1 (ou de 0 à 100 % si l'on raisonne en pourcentage)

Ministry of Higher Education & Scientific Research **Mouloud MAMMERI University of Tizi Ouzou – Algeria** **Faculty of Economics, Business & Management** **Department of Economics** **1st Year LMD - Section E3**		**Academic Year : 2024 – 2025** **Module : Statistics I** **Worksheet 5th** **Lecturer : G. BELBACHIR** **Email : gouraya.belbachir@ummto.dz**

To access the online answer key, please scan the **QR** code using the anonymous login.	

Exercise 1 :

Le score d'un joueur de boules lors de six (6) parties était respectivement de 182, 168, 184, 190, 170 et 174 points. En considérant ces données comme celles d'un échantillon, calculer les statistiques descriptives suivantes :

1. L'étendue.

2. La variance.

3. L'écart type.

4. Le coefficient de variation.

Exercise 2 :

Le tableau suivant donne les températures moyennes par mois à Serbie et à Kosovo en degrés Celsius.

Mois	Janv.	Févr.	Mars	Avr.	Mai	Juin	Juill.	Aout	Sept.	Oct.	Nov.	Déc.
Serbie	-5	-4	4	15	27	31	31	30	26	20	10	-5
Kosovo	3	4	7	10	14	17	19	18	16	17	7	6

1. Calculer l'étendue, la moyenne, la variance et l'écart-type des températures mensuelles pour chacune de ces deux villes.

2. Comparer et analyser les résultats obtenus.

Exercise 3 :

L'étude de la répartition des revenus ***annuels*** des personnes d'un échantillon a donné les résultats suivants :

Revenus en 10^3 DA	- 40	40 à 50	50 à 60	60 à 70	70 à 80	80 à 90	90 et +
Effectifs	45	12	23	46	25	30	15

1. Calculer les revenus annuels moyen et médian sachant qu'aucune personne ne touche ***moins*** de 1666,67 DA par ***mois*** et que le ***maximum*** de revenu ***mensuel*** réalisé dans cette population est de 9166,67 DA.

2. Analyser la dispersion en utilisant l'Ecart-type et le coefficient de variation.

3. Donner l'intervalle contenant 40% des revenus existant au centre de la série.

4. Recalculer la nouvelle moyenne et le nouveau CV sachant que tous les revenus annuels augmentent de 20%.

Exercise 4 :

On considère la distribution suivante : chiffre d'affaires (CA) des entreprises Algériennes :

CA en 10^6 DA	20-40	40-50	50-70	70-80	80-110	110-130	130-150	**Total**
Effectifs des entreprises	40	20	70	30	100	90	50	**400**

1. Calculer l'étendue de cette distribution.

2. Calculer l'écart absolu moyen.

3. Calculer la variance, l'écart type et le coefficient de variation.

4. Calculer la différence médiale-médiane.

~ ... ~ ... ~

Chapitre 8 :

Les indices

Chapitre 8 : Les indices

Le mot indice désigne un nombre sans dimension permettant de faire des comparaisons dans le temps et dans l'espace.

Dans la majorité des cas, la comparaison sera temporelle et portera essentiellement sur des prix et des quantités.

Il existe deux types d'indices :

Ceux qui correspondent à des grandeurs dites « simples » sont exprimés par un nombre appelés « indices élémentaires ». (On appelle grandeur simple toute grandeur repérée par un nombre unique).

Ceux qui correspondent à des grandeurs dites « complexes » composées de différentes grandeurs simples appelées « indices synthétiques ». (On appelle grandeur complexe toute grandeur définie par plusieurs nombres).

Section 1 : Les indices élémentaires

Section 2 : Les indices synthétiques

Section 1 : Les indices élémentaires

1. Définitions :

Un indice est un outil de comparaison, comparaison dans le temps ou dans l'espace.

Soient x_0 et x_t les valeurs prises par une grandeur X respectivement aux temps 0 et t.

On appelle indice élémentaire de X au temps t relativement au temps 0, le rapport noté $I_{t/0}(x)$, défini par : $I_{t/0}(x) = \frac{x_t}{x_0}$.

Cet indice est un indice temporel et s'énonce de la manière suivante : « indice élémentaire de X en t base 1 en 0 ».

Par commodité, on utilise généralement un indice « base 100 en 0 »,c'est-à-dire le nombre $I_{t/0}(x) = 100 \times \frac{x_t}{x_0}$. La date t est appelée « date courante » et la date 0 est appelée « date de référence » ou « de base ».

Interprétation de la « date de base » :

a) Un indice élémentaire base 1 en 0 signifie que $I_{0/0}(x) = 1 \times \frac{x_0}{x_0} = 1.$

b) Un indice élémentaire base 100 en 0 signifie que $I_{0/0}(x) = 100 \times \frac{x_0}{x_0} = 100$.

Soient x_A et x_B les valeurs prises par une grandeur X respectivement dans 2 régions A et B. On appelle indice élémentaire de X en A base 100 en B le nombre : $I_{A/B}(x) = 100 \times \frac{x_A}{x_B}$.

Cet indice élémentaire est qualifié de « spatial ».

2. Types d'indices :

a) Les indices ***élémentaires*** $\Rightarrow$ une seule grandeur (grandeur *simple*)

Ex : Indice du SMIC

b) Les indices ***synthétiques*** $\Rightarrow$ variation d'une grandeur *complexe*

Ex : Indice des salaires, IPC, Indices boursiers

c) Les indices ***composites*** $\Rightarrow$ évolution de l'ens. d'un domaine éco.

Ex : Indice général d'activité d'une branche industrielle

3. Interprétation :

L'indice élémentaire s'interprète comme le facteur de croissance globale de X , F_t, sur la période $[0,t]$. Pour un indice à base 100 par exemple, on a :

a) Si $I_{t/0}(x)>100$, on dit que X a augmenté entre 0 et t de $\left(I_{t/0}(x)-100\right)\%$.

b) Si $I_{t/0}(x)<100$, on dit que X a baissé entre 0 et t de $\left(100-I_{t/0}(x)\right)\%$.

c) Si $I_{t/0}(x)=100$, on dit que X est stationnaire entre 0 et t.

4. Propriétés :

Ces propriétés sont celles des facteurs de croissance.

a) Transitivité :

On a : $I_{t/0}(x)=I_{t/t'}(x)\times I_{t'/0}(x) \quad \forall(t,t')$.

En effet : $I_{t/0}(x)=\dfrac{x_t}{x_0}=\dfrac{x_t}{x_{t'}}\times\dfrac{x_{t'}}{x_0}=I_{t/t'}(x)\times I_{t'/0}(x)$

Cette propriété permet de changer de base, en effet :

$$I_{t/0}(x)=I_{t/t'}(x)\times I_{t'/0}(x)\Leftrightarrow I_{t/t'}(x)=\frac{I_{t/0}(x)}{I_{t'/0}(x)}.$$

Généralisation : Propriété d'enchaînement :

$$I_{t/0}(x)=I_{1/0}(x)\times I_{2/1}(x)\times....\times I_{t/t-1}(x),$$

En effet, on a : $I_{t/0}(x)=\dfrac{x_t}{x_0}=\dfrac{x_1}{x_0}\times\dfrac{x_2}{x_1}\times....\times\dfrac{x_t}{x_{t-1}}=I_{1/0}(x)\times I_{2/1}(x)\times....\times I_{t/t-1}(x)$.

b) La réversibilité :

$$I_{t/0}(x)=\frac{1}{I_{0/t}(x)} \quad \text{ou alors} \quad I_{A/B}(x)=\frac{1}{I_{B/A}(x)}.$$

En effet : $I_{t/0}(x)=\frac{x_t}{x_0}=\frac{1}{\frac{x_0}{x_t}}=\frac{1}{I_{0/t}(x)}$, de la même façon, on a :

$I_{A/B}(x)=\frac{x_A}{x_B}=\frac{1}{\frac{x_B}{x_A}}=\frac{1}{I_{B/A}(x)}$.

c) Factorité :

Soit $X=Z\times Y$

a) Si on connaît $I_{t/0}(Z)$ et $I_{t/0}(Y)$,alors on a : $I_{t/0}(X)=I_{t/0}(Z)\times I_{t/0}(Y)$.

En effet, si $X=Z\times Y$, alors : $x_t=z_t\times y_t$ et $x_0=z_0\times y_0$.

Comme $I_{t/0}(X)=\frac{x_t}{x_0}$, donc $I_{t/0}(x)=\frac{z_t\times y_t}{z_0\times y_0}=\left(\frac{z_t}{z_0}\right)\times\left(\frac{y_t}{y_0}\right)=I_{t/0}(Z)\times I_{t/0}(Y)$.

De même si : $x_t=\frac{z_t}{y_t}$, alors : $x_0=\frac{z_0}{y_0}$, alors : $I_{t/0}(x)=\frac{\frac{z_t}{y_t}}{\frac{z_0}{y_0}}=\frac{\left(\frac{z_t}{z_0}\right)}{\left(\frac{y_t}{y_0}\right)}=\frac{I_{t/0}(Z)}{I_{t/0}(Y)}$.

Cette propriété permet énoncer la relation fondamentale suivante :

Indice élémentaire de valeur = Indice élémentaire des prix × Indice élémentaire des quantités (ou des volumes).

d) Proportionnalité :

b) Si $x_t=k\times x_0$, l'indice élémentaire de X à la date t de base 1 en 0 est égale à k. En effet, $I_{t/0}(x)=\frac{x_t}{x_0}=\frac{k\times x_0}{x_0}=k$.

Exemple :

Le tableau suivant décrit l'évolution du chiffre d'affaires d'une entreprise réalisé sur un produit.

Date	**Prix unitaire** (P_t)	**Quantité vendue** (Q_t)	**Montant des ventes** (V_t)
0	200	5000	100000
t	220	6000	132000

Déterminer par 2 méthodes l'indice élémentaire du chiffre d'affaires de cette entreprise à la date t base 100 en 0. Notons par V le chiffre d'affaires, on a :

1ère méthode : $\left.\begin{array}{l} V_0 = P_o \times Q_o = 200 \times 5000 = 100000 \\ V_t = P_t \times Q_t = 220 \times 6000 = 132000 \end{array}\right\} I_{t/0}(x) = 100 \times \frac{132000}{100000} = 132.$

2ème méthode : on a : $V_t = P_t \times Q_t$, alors : $\frac{V_t}{V_0} = \frac{P_t \times Q_t}{P_o \times Q_o} = \left(\frac{P_t}{P_o}\right) \times \left(\frac{Q_t}{Q_o}\right)$, d'où :

$$I_{t/0}(Q) = 100 \times \frac{220}{200} \times \frac{6000}{5000} = 100 \times 1.10 \times 1.20 = 3.2$$

Augmentation de 10% des prix — Augmentation de 20% des quantités vendues

5. Application en économie :

Soient W_t et W_{t-1} les valeurs aux prix courants d'un bien X et V_t la valeur de X à la date t au prix de la date $t-1$, alors :

$\frac{W_t}{W_{t-1}}$ est un indice de valeur,

$\frac{V_t}{W_{t-1}}$ est un indice de volume,

$\frac{W_t}{V_t}$ est un indice de prix.

Section 2 : Les indices synthétiques

1. Définition :

Ils sont utilisés pour comparer, dans le temps ou dans l'espace, les valeurs prises par une grandeur complexe.

Ex. de grandeur complexe : la consommation des ménages

On s'intéresse à 2 caractéristiques des xi :

Leurs prix $\{p_1, p_2,\ldots, p_k\}$ et es quantités achetées $\{q_1, q_2,\ldots, q_k\}$ à 2 dates différentes, 0 et n.

On notera donc[1]:

c) p_{i0} : le prix de la grandeur i à la date 0,

d) p_{in} : son prix à la date n,

e) q_{i0} : la quantité achetée en 0,

f) q_{in} : la quantité achetée en n

On cherche à rendre compte de l'évolution de la grandeur complexe X entre 2 dates, de l'évolution de « son prix » et de celle de « sa quantité ».

On peut raisonner sur la dépense globale. On calculera l'indice de valeur en n, base en 0 :

$$IVA_{n/0} = I_{\mathbf{n/0}}(\mathbf{pq}) = 100 \cdot \frac{p_{1n}q_{1n} + p_{2n}q_{2n} + \ldots + p_{kn}q_{kn}}{p_{10}q_{10} + p_{20}q_{20} + \ldots + p_{k0}q_{k0}} = 100 \cdot \frac{\sum_{i=1}^{k} p_{\mathbf{in}}q_{\mathbf{in}}}{\sum_{i=1}^{k} p_{\mathbf{i0}}q_{\mathbf{i0}}}$$

Considérons une entreprise qui vend les produits b_1, b_2, b_3, b_4 . On veut expliquer l'évolution du chiffre d'affaires réalisé entre 2 dates 0 et t à partir de l'évolution des prix pratiques et celle des quantités vendues. Les données sont consignées dans le tableau suivant :

[1] Patrica VORNETTI, **Statistique & Informatique**, L1, Sc Eco, PPT, 2016 P : 10.

b	P_{i0}	Q_{i0}	P_{it}	Q_{it}	$V_{i0}=P_{i0}\times Q_{i0}$	$V_{it}=P_{it}\times Q_{it}$	$V_{i0}^{*}=P_{it}\times Q_{i0}$	$V_{it}^{0}=P_{i0}\times Q_{it}$
b_1	5	20	8	15	100	120	160	75
b_2	6	30	7	20	180	140	210	120
b_3	8	40	9	30	320	270	360	240
b_3	10	10	11	20	100	220	110	200
					$V_0=\sum_{i=1}^{4}V_{i0}=700$	$V_t=\sum_{i=1}^{4}V_{it}=750$	$V_0^{*}=\sum_{i=1}^{4}V_{i0}^{*}=840$	$V_{it}^{*}=\sum_{i=1}^{4}V_{it}^{*}=635$

On a : $I_{t/0}(V)=\dfrac{V_t}{V_o}=\dfrac{\sum_{i=1}^{4}V_{it}}{\sum_{i=1}^{4}V_{io}}=\dfrac{750}{700}=1.071.$

Le chiffre d'affaires a augmenté de 7.1% .

Remarque :

Pour un bien b_i, on peut calculer l'indice d'affaires réalisé sur ce bien de la manière suivante : $I_{t/0}(V^i)=\dfrac{V_{it}}{V_{io}}=\dfrac{P_{it}\times Q_{it}}{P_{i0}\times Q_{i0}}=\left(\dfrac{P_{it}}{P_{i0}}\right)\times\left(\dfrac{Q_{it}}{Q_{i0}}\right)=I_{t/0}(p^i)\times I_{t/0}(q^i)$

L'indice élémentaire de V^i s'obtient comme un produit de deux indices élémentaires. Pour V, on ne peut pas obtenir ce résultat, mais on va tout de même décomposer l'indice de V comme produit de deux indices (prix et quantités), mais qui ne seront pas des indices élémentaires.

2. Méthode de décomposition :

On va utiliser la méthode de décomposition de l'indice de valeur vue précédemment : Indice de valeur = indice de volume × indice de prix .

On a alors deux possibilités : soit $I_{t/0}(V)=\dfrac{V_t}{V_o^{*}}\times\dfrac{V_o^{*}}{V_0}$, ou $I_{t/0}(V)=\dfrac{V_t}{V_{t'}}\times\dfrac{V_{t'}}{V_0}$.

$\frac{V_t}{V_0}=\frac{\sum_{i=1}^{4}V_{it}}{\sum_{i=1}^{4}V_{i0}}=\frac{750}{840}=0.893$, $\frac{V_t}{V_0}$ est un indice de volume.

$\frac{V_0^*}{V_0}=\frac{\sum_{i=1}^{4}V_{i0}^*}{\sum_{i=1}^{4}V_{i0}}=\frac{840}{700}=1.20$, $\frac{V_0^*}{V_0}$ est un indice de prix.

$I_{t/0}(V)=\frac{V_t}{V_o^*}\times\frac{V_o^*}{V_0}=\frac{750}{840}\times\frac{840}{700}=(0.893)\times(1.20)=1.071.$

$\frac{V_t}{V_t^*}=\frac{\sum_{i=1}^{4}V_{it}}{\sum_{i=1}^{4}V_{it}^*}=\frac{750}{635}=1.18$, $\frac{V_t}{V_t^*}$ est un indice de prix.

$\frac{V_t^*}{V_0}=\frac{\sum_{i=1}^{4}V_{it}^*}{\sum_{i=1}^{4}V_{i0}}=\frac{635}{700}=0.907$, $\frac{V_t^*}{V_0}$ est un indice de volume.

$I_{t/0}(V)=\frac{V_t}{V_i^*}\times\frac{V_i^*}{V_0}=\frac{750}{635}\times\frac{635}{700}=(1.18)\times(0.907)=1.071$.

Montrons que les indices utilisés précédemment sont des moyennes d'indices élémentaires.

3. Précisons les indices des prix :

$$\frac{V_0^*}{V_0}=\frac{\sum_{i=1}^{4}P_{it}\times Q_{i0}}{\sum_{i=1}^{4}P_{i0}\times Q_{i0}}=\frac{\sum_{i=1}^{4}\left(\frac{P_{it}}{P_{i0}}\right)\times(P_{i0}\times Q_{i0})}{\sum_{i=1}^{4}P_{i0}\times Q_{i0}}=\frac{\sum_{i=1}^{4}I_{t/0}(p^i)\times(P_{i0}\times Q_{i0})}{\sum_{i=1}^{4}P_{i0}\times Q_{i0}}.$$

Si on pose $\alpha_{i0}=\frac{(P_{i0}\times Q_{i0})}{\sum_{i=1}^{4}P_{i0}\times Q_{i0}}$ $\Leftrightarrow$ $\sum_{i=1}^{4}\alpha_{i0}=\frac{(P_{10}\times Q_{10})+.....+(P_{40}\times Q_{40})}{\sum_{i=1}^{4}P_{i0}\times Q_{i0}}=1.$

Par suite :

$\frac{V_0^*}{V_0}=\sum_{i=1}^{r}\alpha_{i0}\times I_{t/0}(p^i)$, il s'agit donc de la moyenne arithmétique pondérée des indices élémentaires de prix $I_{t/0}(p^i)$.

Le coefficient α_{10} représente la part de la contribution du bien b_i dans le chiffre d'affaires global réalisé à la date 0.

$$\frac{V_t}{V_t^*}=\frac{\sum_{i=1}^{4}P_{it}\times Q_{it}}{\sum_{i=1}^{4}P_{i0}\times Q_{it}}=\frac{\sum_{i=1}^{4}P_{it}\times Q_{it}}{\sum_{i=1}^{4}\left(\frac{P_{i0}}{P_{it}}\right)\times\left(P_{it}Q_{it}\right)}=\frac{1}{\frac{1}{\sum_{i=1}^{4}P_{it}\times Q_{it}}\times\sum_{i=1}^{4}\left(\frac{P_{i0}}{P_{it}}\right)\times\left(P_{it}Q_{it}\right)},$$

Comme : $I_{t/0}\left(p^i\right)=\frac{P_{it}}{P_{i0}} \quad\Leftrightarrow\quad \frac{P_{i0}}{P_{it}}=\frac{1}{I_{t/0}\left(p^i\right)}$,

Par suite : $\frac{V_t}{V_t^*}=\frac{1}{\frac{\sum_{i=1}^{4}\frac{1}{I_{t/0}\left(p^i\right)}\times\left(P_{it}Q_{it}\right)}{\sum_{i=1}^{4}P_{it}\times Q_{it}}}$, En posant $\alpha_{it}=\frac{P_{it}\times Q_{it}}{\sum_{i=1}^{4}P_{it}\times Q_{it}}$,

on obtient : $\frac{V_t}{V_t^*}=\frac{1}{\sum_{i=1}^{4}\frac{1}{I_{t/0}\left(p^i\right)}\times\alpha_{it}}$, il s'agit donc de la moyenne harmonique pondérée

des indices élémentaires de prix $I_{t/0}\left(p^i\right)$. Le coefficient α_{it} représente la part de la contribution du bien b_i dans le chiffre d'affaires global réalisé à la date t.

En utilisant les mêmes procédés, on montre que :

$\frac{V_t}{V_0^*}=\frac{1}{\sum_{i=1}^{4}\frac{1}{I_{t/0}\left(q^i\right)}\times\alpha_{it}}$ est la moyenne harmonique pondérée des indices élémentaires

des quantités : $I_{t/0}\left(q^i\right)$.

$\frac{V_0^*}{V_0}=\sum_{i=1}^{4}\frac{1}{I_{t/0}\left(q^i\right)}\times\alpha_{i0}$ est un moyenne arithmétique pondérée des indices élémentaires

des quantités $I_{t/0}\left(q^i\right)$ par les coefficients α_{i0}.

4. Les indices de Laspeyres, Paasche et Fisher :

a) On appelle indice des prix de **Laspeyres**, l'indice synthétique des prix noté $L_{t/0}(P)$

et défini par : $L_{t/0}(P) = \dfrac{\sum_{i=1}^{r} P_{it} \times Q_{i0}}{\sum_{i=1}^{r} P_{i0} \times Q_{i0}}$.

$$\mathbf{L_{n/0}(p)} = 100 \cdot \frac{p_{1n}\mathbf{q_{10}} + p_{2n}\mathbf{q_{20}} + \ldots + p_{kn}\mathbf{q_{k0}}}{p_{10}\mathbf{q_{10}} + p_{20}\mathbf{q_{20}} + \ldots + p_{k0}\mathbf{q_{k0}}} = 100 \cdot \frac{\sum_{i=1}^{k} p_{in}\mathbf{q_{i0}}}{\sum_{i=1}^{k} p_{i0}\mathbf{q_{i0}}}$$

De même, l'indice de quantité de Laspeyres est l'indice synthétique noté $L_{t/0}(Q)$,

défini par : $L_{t/0}(Q) = \dfrac{\sum_{i=1}^{r} P_{i0} \times Q_{it}}{\sum_{i=1}^{r} P_{i0} \times Q_{i0}}$.

$$\mathbf{L_{n/0}(q)} = 100 \cdot \frac{\mathbf{p_{10}}q_{1n} + \mathbf{p_{20}}q_{2n} + \ldots + \mathbf{p_{k0}}q_{kn}}{\mathbf{p_{10}}q_{10} + \mathbf{p_{20}}q_{20} + \ldots + \mathbf{p_{k0}}q_{k0}} = 100 \cdot \frac{\sum_{i=1}^{k} \mathbf{p_{i0}}q_{in}}{\sum_{i=1}^{k} \mathbf{p_{i0}}q_{i0}}$$

L'interprétation de l'indice de Laspeyres est quelque peu problématique puisque la méthode de pondération suppose que les quantités de référence q_0 ne varient pas quand les prix changent. De plus, l'indice de Laspeyres tend à perdre sa représentativité au cours du temps.

b) On appelle indice de prix de **Paasche** l'indice synthétique des prix noté :

$$\mathbf{P_{n/0}(p)} = 100 \cdot \frac{p_{1n}\mathbf{q_{1n}} + p_{2n}\mathbf{q_{2n}} + \ldots + p_{kn}\mathbf{q_{kn}}}{p_{10}\mathbf{q_{1n}} + p_{20}\mathbf{q_{2n}} + \ldots + p_{k0}\mathbf{q_{kn}}} = 100 \cdot \frac{\sum_{i=1}^{k} p_{in}\mathbf{q_{in}}}{\sum_{i=1}^{k} p_{i0}\mathbf{q_{in}}}$$

On appelle indice de prix de **Paasche** l'indice synthétique des quantités noté :

$$\mathbf{P_{n/0}(q)} = 100 \cdot \frac{\mathbf{p_{1n}}q_{1n} + \mathbf{p_{2n}}q_{2n} + \ldots + \mathbf{p_{kn}}q_{kn}}{\mathbf{p_{1n}}q_{10} + \mathbf{p_{2n}}q_{20} + \ldots + \mathbf{p_{kn}}q_{k0}} = 100 \cdot \frac{\sum_{i=1}^{k} \mathbf{p_{in}}q_{in}}{\sum_{i=1}^{k} \mathbf{p_{in}}q_{i0}}$$

L'utilisation de l'indice de Paasche n'est pas aisée car les comparaisons interpériodes sont rendues complexes puisque les pondérations varient dans le temps.

c) On appelle indice de **Fisher** des prix et des quantités :

Problème de Paasche : comparaisons difficiles puisque Δ pondérations.

Problème de Laspeyres : perte de représentativité au fil du temps.

Donc l'indice de Fisher = moyenne géométrique des taux précédents.

$$F_{n/0}(p) = \sqrt{L_{n/0}(p) \cdot P_{n/0}(p)}$$
$$F_{n/0}(q) = \sqrt{L_{n/0}(q) \cdot P_{n/0}(q)}$$

5. Propriétés des indices synthétiques :

a) Les indices de Laspeyres et de Paasche sont les moyennes d'indices élémentaires :

$L_{t/0}(P) = \sum_{i=1}^{h} \alpha_{i0} \times I_{t/0}(p^i)$, moyenne arithmétique pondérée des prix $I_{t/0}(p^i)$.

$L_{t/0}(Q) = \sum_{i=1}^{h} \alpha_{i0} \times I_{t/0}(q^i)$, moyenne arithmétique pondérée des prix $I_{t/0}(q^i)$.

$\sigma_{t/0}(P) = \dfrac{1}{\sum_{i=1}^{h} \alpha_{it} \times \dfrac{1}{I_{t/0}(p^i)}}$, moyenne harmonique pondérée des prix $I_{t/0}(p^i)$.

$\sigma_{t/0}(Q) = \dfrac{1}{\sum_{i=1}^{h} \alpha_{it} \times \dfrac{1}{I_{t/0}(q^i)}}$, moyenne harmonique pondérée des prix $I_{t/0}(q^i)$.

Ou encore : $\begin{cases} L_{t/0}(*) = \sum_{i=1}^{h} \alpha_{i0} \times I_{t/0}(*) \\ \sigma_{t/0}(*) = \dfrac{1}{\sum_{i=1}^{h} \alpha_{it} \times I_{t/0}(*)} \end{cases}$, où $(*)$ désigne prix ou quantité.

$\alpha_{i0} = \dfrac{P_{i0} \times Q_{i0}}{\sum_{i=1}^{h} (P_{i0} \times Q_{i0})} = \dfrac{V_{i0}}{V_0}$, et $\alpha_{it} = \dfrac{P_{it} \times Q_{it}}{\sum_{i=1}^{h} (P_{it} \times Q_{it})} = \dfrac{V_{it}}{V_t}$.

Exemples :

a) Exemple 1 : On a posé des hypothèses pour fixer les quantités consommées en t (1) et t-1 (0).

Calcul des poids et budgets des indices synthétiques [1]:

Produit	p_0^k	q_0^k	p_1^k	q_1^k	$p_0^k q_0^k$	$p_1^k q_1^k$	$p_0^k q_0^k$	$p_0^k q_1^k$
Cinéma	100	50	125	55	5000	6875	6250	5500
Hebdo	50	40	75	30	2000	2250	3000	1500
Piscine	60	25	66	25	1500	1650	1650	1500
Chope	20	300	21	310	6000	6510	6300	6200
Total	-	-	-	-	**14500**	**17285**	**17200**	**14700**

g) Indice de Laspeyres : $\frac{\sum_{k=1}^{4} p_1^k q_0^k}{\sum_{k=1}^{4} p_0^k q_0^k} = \frac{17200}{14500} = 1{,}186$ (ou 118,6 en base 100).

h) Indice de Paasche : $\frac{\sum_{k=1}^{4} p_1^k q_1^k}{\sum_{k=1}^{4} p_0^k q_1^k} = \frac{17285}{14700} = 1{,}176$ (ou 117,6 en base 100).

i) L'indice de Fisher :

L'indice de Fisher = moyenne géométrique des taux précédents.

Dans l'exemple : Fisher = $\sqrt{118{,}6 * 117{,}6}$ = 118,1 (en base 100).

[1] Partie 2 : Les chroniques, chapitre 2 : construction des indices, P 7.

b) Exemple 2 :

b_i	P_{i0}	Q_{i0}	P_{it}	Q_{it}	V_{i0}	V_{it}	α_{i0}	α_{it}	$I_{t/0}(p^i)$	$I_{t/0}(q^i)$
b_1	5	20	8	15	100	120	$\frac{10}{70}$	$\frac{12}{75}$	$\frac{8}{5}$	$\frac{15}{20}$
b_2	6	30	7	20	180	140	$\frac{18}{70}$	$\frac{14}{75}$	$\frac{7}{6}$	$\frac{20}{30}$
b_3	8	40	9	30	320	270	$\frac{32}{70}$	$\frac{27}{75}$	$\frac{9}{8}$	$\frac{30}{40}$
b_4	10	10	11	20	100	220	$\frac{10}{70}$	$\frac{22}{75}$	$\frac{11}{10}$	$\frac{20}{10}$
Σ	-	-	-	-	$V_0 = 700$	$V_t = 750$	-	-	-	-

$$L_{t/0}(P) = \sum_{i=1}^{h} \alpha_{i0} \times I_{t/0}(p^i) = \frac{10}{70} \times \frac{8}{5} + \frac{18}{70} \times \frac{7}{6} + \frac{32}{70} \times \frac{8}{9} + \frac{10}{70} \times \frac{11}{10} = 1.20.$$

$$L_{t/0}(Q) = \sum_{i=1}^{h} \alpha_{i0} \times I_{t/0}(q^i) = \frac{10}{70} \times \frac{15}{20} + \frac{18}{70} \times \frac{20}{30} + \frac{32}{70} \times \frac{30}{10} + \frac{10}{70} \times \frac{20}{10} = 0.907.$$

$$\sigma_{t/0}(P) = \frac{1}{\sum_{i=1}^{4} \alpha_{it} \times \frac{1}{I_{t/0}(P^i)}} = \frac{1}{\frac{12}{75} \times \frac{5}{8} + \frac{14}{75} \times \frac{6}{7} + \frac{27}{75} \times \frac{8}{9} + \frac{22}{75} \times \frac{10}{11}} = 1.18.$$

$$\sigma_{t/0}(Q) = \frac{1}{\sum_{i=1}^{4} \alpha_{it} \times \frac{1}{I_{t/0}(q^i)}} = \frac{1}{\frac{12}{75} \times \frac{20}{15} + \frac{14}{75} \times \frac{20}{20} + \frac{27}{75} \times \frac{40}{30} + \frac{22}{75} \times \frac{10}{20}} = 0.893.$$

b) Aucun des trois indices ne vérifie la propriété de transitivité.

Exemple : $L_{t/t'}(P) \neq \dfrac{L_{t/0}(P)}{L_{t'/0}(P)}$.

En effet : $\left.\begin{aligned} L_{t/0}(P) &= \dfrac{\sum_{i=1}^{h} P_{it} \times Q_{i0}}{\sum_{i=1}^{h} P_{i0} \times Q_{i0}} \\ L_{t'/0}(P) &= \dfrac{\sum_{i=1}^{h} P_{it'} \times Q_{i0}}{\sum_{i=1}^{h} P_{i0} \times Q_{i0}} \end{aligned}\right\} \Rightarrow \dfrac{L_{t/0}(P)}{L_{t'/0}(P)} = \dfrac{\sum_{i=1}^{h} P_{it} \times Q_{i0}}{\sum_{i=1}^{h} P_{it'} \times Q_{i0}}$,

Nous avons donc : $L_{t/t'}(P) = \dfrac{\sum_{i=1}^{h} P_{it} \times Q_{it'}}{\sum_{i=1}^{h} P_{it'} \times Q_{it'}} \neq \dfrac{\sum_{i=1}^{h} P_{it} \times Q_{i0}}{\sum_{i=1}^{h} P_{it'} \times Q_{i0}} = \dfrac{L_{t/0}(P)}{L_{t'/0}(P)}$.

Le problème de changement se résout en introduisant les indices chaînes. L'indice de chaînes de Laspeyres noté $CL_{t/0}(*)$ est défini par $CL_{t/0}(*) = L_{t/t-1}(*) \times L_{t-1/t-2}(*) \times ..... \times L_{1/0}(*)$, où $(*)$ désigne le prix ou la quantité.

Par exemple : $CL_{t/0}(P) = L_{t/t-1}(P) \times L_{t-1/t-2}(P) \times ..... \times L_{1/0}(P)$, d'où :

$CL_{t-1/0}(P) = L_{t-1/t-2}(P) \times ..... \times L_{1/0}(P)$,

Donc on a : $\dfrac{CL_{t/0}(P)}{CL_{t-1/0}(P)} = \dfrac{L_{t/t-1}(P) \times L_{t-1/t-2}(P) \times ..... \times L_{1/0}(P)}{L_{t-1/t-2}(P) \times ..... \times L_{1/0}(P)}$,

Or puisque : $L_{t/t-1}(P) = CL_{t/t-1}(P)$, alors : $CL_{t/t-1}(P) = \dfrac{CL_{t/0}(P)}{CL_{t-1/0}(P)}$.

$CL_{t/0}(*)$ vérifie donc la propriété de transitivité, il en est de même pour $CP_{t/0}(*)$.

c) Les indices de Laspeyres et de Paasche ne vérifient pas la propriété de réversibilité.

$$L_{0/n}(p) = 100 \cdot \frac{\sum p_{i0} q_{in}}{\sum p_{in} q_{in}} = \frac{100}{\dfrac{\sum p_{in} q_{in}}{\sum p_{i0} q_{in}}} \neq \frac{100^2}{L_{n/0}(p)}$$

En effet : $L_{0/t}(*)=\dfrac{1}{\sigma_{t/0}(P)}$ et $\sigma_{t/0}(*)=\dfrac{1}{L_{t/0}(*)}$.

Illustration pour $L_{0/0}(P)$ et $\sigma_{0/t}(P)$.

$L_{0/t}(P)=\dfrac{\sum_{i=1}^{h}P_{i0}\times Q_{it}}{\sum_{i=1}^{h}P_{it}\times Q_{it}}$ et $L_{t/0}(P)=\dfrac{\sum_{i=1}^{h}P_{it}\times Q_{i0}}{\sum_{i=1}^{h}P_{i0}\times Q_{i0}}$, alors que : $\dfrac{1}{L_{t/0}(P)}=\dfrac{\sum_{i=1}^{h}P_{i0}\times Q_{i0}}{\sum_{i=1}^{h}P_{it}\times Q_{i0}}$.

On a donc : $L_{t/0}(P)\neq L_{0/t}(P)$.

De plus : $L_{0/t}(P)=\dfrac{\sum_{i=1}^{h}P_{i0}\times Q_{it}}{\sum_{i=1}^{h}P_{it}\times Q_{it}}=\dfrac{1}{\dfrac{\sum_{i=1}^{h}P_{it}\times Q_{it}}{\sum_{i=1}^{h}P_{i0}\times Q_{it}}}=\dfrac{1}{\sigma_{t/0}(P)}$

Et : $\sigma_{0/t}(P)=\dfrac{\sum_{i=1}^{h}P_{i0}\times Q_{i0}}{\sum_{i=1}^{h}P_{it}\times Q_{i0}}=\dfrac{1}{\dfrac{\sum_{i=1}^{h}P_{it}\times Q_{i0}}{\sum_{i=1}^{h}P_{i0}\times Q_{i0}}}=\dfrac{1}{L_{t/0}(P)}$.

Par contre, les indices de Fisher vérifient la propriété de réversibilité, c'est-à-dire :

$F_{0/t}(*)=\dfrac{1}{F_{t/0}(*)}$.

$F_{0/t}(P)=\sqrt{L_{0/t}(P)\times\sigma_{0/t}(P)}$

Or on a : $L_{t/0}(P)=\dfrac{\sum_{i=1}^{h}P_{i0}\times Q_{it}}{\sum_{i=1}^{h}P_{it}\times Q_{it}}=\dfrac{1}{\sigma_{t/0}(P)}$ et $\sigma_{0/t}(P)=\dfrac{\sum_{i=1}^{h}P_{i0}\times Q_{i0}}{\sum_{i=1}^{h}P_{it}\times Q_{i0}}=\dfrac{1}{L_{t/0}(P)}$, alors :

$F_{0/t}(P)=\sqrt{\dfrac{1}{\sigma_{t/0}(P)}\times\dfrac{1}{L_{t/0}(P)}}=\dfrac{1}{\sqrt{L_{t/0}(P)\times\sigma_{t/0}(P)}}=\dfrac{1}{F_{t/0}(P)}$.

6. Comparaison des indices de Laspeyres et de Paasche :

C'est par le choix de la date de référence t que se distinguent les 2 indices synthétiques les plus courants, l'indice de **Laspeyres** et l'indice de **Paasche** :

a) Laspeyres ⇒ Choix de la **date de départ** (i.e. **t = 0**),

b) Paasche ⇒ Choix de la **date d'arrivée** (i.e. **t = n**).

Les deux indices ne sont pas comparables sans hypothèses particulières.

Cas n°1 : $I_{t/0}\left(q^i\right)=\dfrac{Q_{it}}{Q_{i0}}=C_q \quad (\forall i)$, (taux de croissance identique en volume).

$L_{t/0}(Q)=\sigma_{t/0}(Q)=C_q$, on en déduit que : $L_{t/0}(P)=\sigma_{t/0}(P)$.

En effet, $I_{t/0}(V)=L_{t/0}(P)\times\sigma_{t/0}(Q)=L_{t/0}(Q)\times\sigma_{t/0}(P)$.

Cas n°2 : $I_{t/0}\left(p^i\right)=\dfrac{P_{it}}{P_{i0}}=C_p \quad (\forall i)$, (taux de croissance des prix identiques).

$L_{t/0}(P)=\sigma_{t/0}(P) \quad \Rightarrow \quad L_{t/0}(Q)=\sigma_{t/0}(Q)$.

Cas n°3 : $\dfrac{P_{it}\times Q_{it}}{P_{i0}\times Q_{i0}}=C_v \qquad (\forall i)$, alors : $P_{t/0}(*)\le L_{t/0}(*)$.

Preuve :

On va montrer que $\alpha_{i0}=\alpha_{it} \qquad (\forall i)$:

On a : $\dfrac{P_{it}\times Q_{it}}{P_{i0}\times Q_{i0}}=C_v \quad (\forall i) \quad \Leftrightarrow \quad P_{it}\times Q_{it}=C_v\times P_{i0}\times Q_{i0}$, et $\sum_{i=1}^{h}P_{it}\times Q_{it}=C_v\times\sum_{i=1}^{h}P_{i0}\times Q_{i0}$.

Or, on a : $\alpha_{it}=\dfrac{P_{it}\times Q_{it}}{\sum_{i=1}^{h}P_{it}\times Q_{it}}$, d'où : $\alpha_{it}=\dfrac{C_v\times\left(P_{i0}\times Q_{i0}\right)}{\sum_{i=1}^{h}P_{it}\times Q_{it}}=\dfrac{C_v\times\left(P_{i0}\times Q_{i0}\right)}{C_v\times\sum_{i=1}^{h}\left(P_{i0}\times Q_{i0}\right)}=\alpha_{i0}$.

Comme $\sigma_{t/0}(*)$ est une moyenne harmonique, elle est inférieure ou égale a la moyenne arithmétique $L_{t/0}(*)$.

Cette dernière relation est également vérifiée dans un cas plus général, en effet, comme, en moyenne, prix et quantité varient en sens inverse :

a) L'indice de Laspeyres des prix a tendance à surestimer une hausse.

b) L'indice de Paasche a tendance à la sous-estimer.

En comptabilité, les indices de prix utilisés sont des indices de Paasche. Les indices conjoncturels (indice des prix à la consommation : indice de la production industrielle, …) sont des indices de Laspeyres.

Ministry of Higher Education & Scientific Research **Mouloud MAMMERI University of Tizi Ouzou – Algeria** **Faculty of Economics, Business & Management** **Department of Economics** **1st Year LMD - Section E3**		**Academic Year : 2024 – 2025** **Module : Statistics I** **Worksheet 7th** **Lecturer : G. BELBACHIR** **Email : gouraya.belbachir@ummto.dz**

To access the online answer key, please scan the **QR** code using the anonymous login.	

Exercise 1 :

Une entreprise X désire connaître l'évolution de son chiffre d'affaires à l'exportation depuis sa création en 2017. Le tableau ci-dessous donne le chiffre d'affaires (CA) à l'exportation (en milliers de DA) de cette entreprise au 31 décembre de chaque année :

Année	2017	2018	2019	2020
CA	1040	965	1543	2024

1. Calculer la série des indices simples base 100 en 2017.

2. Calculer la série des indices simples base 100 où la base est à chaque fois l'année précédente.

3. Après avoir défini le taux de variation, calculer les taux de variation du CA de l'entreprise pour chaque année.

4. Après avoir défini le taux de variation *global*, calculer le taux de variation *global* du CA de l'entreprise de 2017 à 2020.

5. Calculer la valeur du CA en décembre 2021 :

- Si l'on prévoit une *augmentation* de 37% par rapport à l'année précédente.
- Si l'on prévoit une *diminution* de 8% par rapport à l'année précédente.

Exercise N° 2 :

Soit le tableau ci-dessous :

Produits	Prix		Quantités	
	Période 0	Période 1	Période 0	Période 1
A	10	12	40	50
B	50	55	80	100
C	25	40	50	30

1. Calculer et interpréter l'indice de valeur base 100.

2. Calculer et interpréter les indices de Laspeyres base 100.

3. Calculer et interpréter les indices de Paasche base 100.

4. Calculer et interpréter les indices de Fischer base 100.

Exercise N° 3 :

Vous travaillez sur des séries de données tirées de la statistique administrative où vous avez été envoyé en mission d'expertise.

1. Il vous faut produire deux *indices composites* (Laspeyres et Paasche, base 1997 = 100) du prix des fleurs en vente sur le marché de la capitale en 2000.

Vous disposez pour ce faire des trois doubles séries suivantes :

Tableau des ventes (milliers d'unités) et des prix moyens observés durant l'année (€) 1997 - 2000

	Roses		Orchidées		Oeillets	
Années	**Prix**	**Ventes**	**Prix**	**Ventes**	**Prix**	**Ventes**
1997	1,23	3256	2,45	597	0,56	5698
1998	1,25	4567	2,79	612	0,63	5893
1999	1,19	3972	3,06	624	0,48	5364
2000	1,32	3587	2,98	658	0,67	6971

2. L'indice de valeur des ventes de fleurs (base 1997 = 100) valait 131,2741 en 2000.

A partir de cette valeur et du calcul à la question 1, calculer les indices (Laspeyres et Paasche) synthétiques des quantités (base 1997 = 100) pour l'année 2000.

3. L'indice de valeur des ventes de fleurs (base 1997 = 100) valait 131,2741 en 2000 :

a) Ceci veut dire qu'entre 1997 et 2000, la valeur des ventes de fleurs a connu un taux de croissance global de : ….

b) Calculez le taux symétrique du taux de croissance global de la valeur des ventes.

~ ... ~ ... ~

Chapitre 9 :
Distributions à deux caractères, corrélation et régression

Chapitre 9 : Distributions à deux caractères, corrélation et régression

La 1ère partie traitait une seule dimension (colonne xi et ni). Dans ce chapitre, on va doubler l'information (xi, yj et ni). Ce tableau porte le nom de tableau de contingence (sous la forme d'un *tableau à double entrée*). On va chercher à calculer des moyennes (pour x et y).

Soient X et Y les deux caractères étudiés, p le nombre de modalités prises par X, q le nombre de modalités prises par Y et n le nombre total d'observations. On étudie, par exemple, le poids et la taille d'un nombre n d'individus, le temps de travail sans pause et le nombre de pièces assemblées ou le nombre d'accidents survenus pendant cette période.

On définit alors la distribution conjointe, les distributions marginales et les distributions conditionnelles. L'étude de la distribution de deux variables se poursuit par celle de leur *liaison*.

L'étude de la liaison entre les variables observées, appelée communément l'étude des corrélations, dépend de leur nature. On envisagera les trois cas suivants :

a) Deux variables quantitatives :
b) Deux variables qualitatives
c) Une variable quantitative et une variable qualitative.

Section 1 : Présentation et notions fondamentales des distributions à deux caractères
Section 2 : Caractéristiques des distributions à deux caractères
Section 3 : Analyse de la relation entre deux variables quantitatives

Section 1 : Présentation et notions fondamentales des distributions à deux caractères

1. Les tableaux de contingence :

On suppose que les deux variables étudiées sont des variables discrètes et que les caractères sont des caractères quantitatifs. Les tableaux statistiques portent le nom de tableaux croisés ou tableaux de contingence.

Dans chaque case du tableau, on écrit l'effectif n_{ij} de l'échantillon, c'est-à-dire le nombre de données tel que X = xi et Y = yj .

Tableau : Tableau de contingence : distribution conjointe de deux variables X et Y

Y / X	**Y_1**	**Y_2**	**……**	**y_j**	**……...**	**y_p**	**$n_{i.}$**
X_1	n_{11}	n_{12}	……	n_{1j}	……	n_{1p}	**$n_{1.}$**
X_2	n_{21}	n_{22}	……	n_{2j}	……	n_{2p}	**$n_{2.}$**
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
X_i	n_{i1}	n_{i2}	……	n_{ij}	……	n_{ip}	**$n_{i.}$**
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
.	.	.		.		.	.
X_k	n_{k1}	n_{k2}	……	n_{kj}	……	n_{kp}	**$n_{k.}$**
$n_{.j}$	**$n_{.1}$**	**$n_{.2}$**	**……**	**$n_{.j}$**	**……**	**$n_{.p}$**	**$n_{..}$**

L'effectif nij désigne le nombre de fois où la modalité xi de la variable X et la modalité yj de la variable Y ont été observées simultanément.

L'effectif ni• est le nombre total d'observations de la modalité xi de X, quelle que soit la modalité de Y[1] :

$\mathbf{n_{i.}} = \sum_{j=1}^{p} \boldsymbol{nij}$

De même, l'effectif n•j est le nombre total d'observations de la modalité yj de Y, quelle que soit la modalité de X[2] :

$\mathbf{n_{.j}} = \sum_{i=1}^{k} \boldsymbol{nij}$

On a évidemment :

$$\sum_{i=1}^{k} \boldsymbol{n.j} = \sum_{j=1}^{p} \boldsymbol{ni.} = \mathbf{n..} = \mathbf{N}$$

La distribution conjointe peut aussi être définie par les fréquences :

$\mathbf{f_{ij}} = \mathbf{n_{ij}} / \mathbf{n..}$ avec : $\Sigma f_{ij} = 1$

2. Notions fondamentales des distributions bivariées :

a) Effectif marginal :

Les *k* couples (x_i , $n_{i\bullet}$) forment la *distribution marginale* de la *variable X*.
Les *l* couples (y_j , $n_{\bullet j}$) forment la *distribution marginale* de la *variable Y*.

On écrit alors :

$\mathbf{n_{i.}} = \sum_{j=1}^{p} \boldsymbol{nij}$ et $\mathbf{n_{.j}} = \sum_{i=1}^{k} \boldsymbol{nij}$

Le point « **.** » désigne l'élément (la ligne i ou la colonne j) qui varie.
Disposant d'une distribution conjointe, on peut déduire les distributions marginales qui permettent d'étudier séparément chaque variable en représentant graphiquement sa distribution et s'il s'agit d'une variable quantitative, en calculant ses caractéristiques de tendance centrale, de dispersion, de forme …

[1] Bernard GOLDFARB, Catherine PARDOUX, **Introduction à la méthode statistique : Manuel et exercices corrigés**, 6e édition, DUNOD, Paris, France, 2011, P : 68.
[2] Bernard GOLDFARB, Catherine PARDOUX, op.cit., P : 68.

b) Distribution conditionnelle :

Elles sont déterminées ; soit selon X, soit selon Y.

▪ **Distribution conditionnelle de X selon Y :**

Elle signifie la distribution du caractère X selon, ou sous condition, que $y = y_j$, c'est-à-dire à la condition que y soit fixé à l'une de ses modalités. On obtient alors la distribution suivant les couples $(x_i ; n_{ij})$.

Par exemple, la première colonne du tableau ($y = y_1$) représente la distribution conditionnelle de x selon $y = y_1$ (ou sous condition $y = y_1$). On obtient alors la distribution suivant les couples $(x_i ; n_{i1})$, comme suit :

Tableau : Distribution conditionnelle de X selon $Y = y_1$

X	**Y_1**
x_1	n_{11}
x_2	n_{21}
.	.
.	.
.	.
.	.
.	.
x_i	n_{i1}
.	.
.	.
.	.
.	.
.	.
x_k	n_{k1}
$n_{.j}$	**$n_{.1}$**

▪ **Distribution conditionnelle de Y selon X**

Elle signifie la distribution du caractère Y selon, ou sous condition, que $x = xi$, c'est-à-dire à la condition que X soit fixé à l'une de ses modalités. On obtient alors la distribution suivant les couples $(y_j ; n_{ij})$.

Par exemple, la première ligne du tableau ($x = x_1$) représente la distribution conditionnelle de Y selon $x = x_2$ (ou sous condition que $x = x_1$). On obtient alors la distribution conditionnelle de Y suivant les couples (y_j ; n_{1j}).

Tableau : Distribution conditionnelle de Y selon $X = x_1$»

Y	y_1	y_2	.....	y_j	.....	y_p	$\mathbf{n_{i.}}$
$\mathbf{X_1}$	n_{11}	n_{12}	.....	n_{1j}	.....	n_{1p}	$\mathbf{n_{1.}}$

Au final, on aura donc autant de distributions conditionnelles de y selon x qu'il y a de modalités pour x, et autant de distributions conditionnelles de x selon y, qu'il y a de modalités pour y.

c) Notion de fréquence marginale :

On appelle fréquences marginales, notées « $\boldsymbol{f_{i\,.}}$ » ou « $\boldsymbol{f_{.j}}$ », les rapports des effectifs marginaux (« $n_{i.}$ » et/ou « $n_{.j}$ ») à l'effectif total (« n.. »). On écrit :

$\mathbf{f_{i.} = n_{i.}/n..}$ **et** $\mathbf{f_{.j} = n_{.j}/n..}$

Avec : $\mathbf{f_{i.}}$ et $\mathbf{f_{.j}} \in \mathbf{[0 ; 1]}$ **et** $\mathbf{\Sigma\, f_{i.} = \Sigma\, f_{.j} = 1.}$

▪ $\mathbf{f_{i.}}$ représente la proportion des individus présentant la modalité $x_{i,}$ par rapport à l'effectif total, quelque soit les modalités de y.

▪ $\mathbf{f_{.j}}$ représente la proportion des individus présentant la modalité y_j, par rapport à l'effectif total, quelque soit les modalités de x.

d) Notion de fréquence partielle sur effectif total :

On appelle fréquence partielle sur effectif total, notée « $\mathbf{f_{ij}}$ », le rapport de l'effectif partiel « $\mathbf{n_{ij}}$ » sur l'effectif total « **n..** ». On écrit :

$$\mathbf{f_{ij} = n_{ij} /n..} \qquad \text{avec } \Sigma f_{ij} = 1$$

Ainsi, « f_{ij} » représente la proportion des individus présentant simultanément la modalité x_i et la modalité y_j, par rapport à l'effectif total de la population étudiée.

e) Notion de fréquence conditionnelle :

On appelle fréquence conditionnelle « $\mathbf{f_{i/j}}$ » ou « $\mathbf{f_{j/i}}$ », le rapport de l'effectif partiel à l'effectif marginal correspondant. On écrit alors :

$\mathbf{f_{i/j} = n_{ij}/n_{.j}}$ et $\mathbf{f_{j/i} = n_{ij}/n_{i.}}$

▪ « $\mathbf{f_{i/j}}$ » se lit : *f_i si j*, soit ; *la fréquence conditionnelle de x = xi si y = y_j = constante*. Ou bien *fréquence conditionnelle de x = x_i si y est fixé (ou constant) à la modalité ou colonne « j », ou sous condition que y = y_j,* ou ; *par rapport à la colonne j »*. La lecture se fait dans ce cas à la verticale. Autrement dit, on croise toutes les lignes par la colonne « j ».

▪ $\mathbf{f_{j/i}}$ se lit « *f_j si i* », soit ; « *fréquence conditionnelle de y = y_j si x = x_i = constante* ». ou bien, « *fréquence conditionnelle de y = y_j sous condition que x = x_i* ou *par rapport à ligne i* ».

Section 2 : Caractéristiques des distributions à deux caractères

Comme pour les distributions à un seul caractère, on peut calculer sur les distributions à deux caractères plusieurs paramètres. Cependant, comme on l'a vu dans la section précédente, il y a plusieurs niveaux de calcul. Aussi, pour les besoins de notre cours, nous ne retiendrons que les moyennes arithmétiques et les variances qui nous serons utiles pour l'analyse de la relation entre les deux variables X et Y.

1. Caractéristiques des distributions marginales :

Il existe deux distributions marginales ; l'une suivant X, l'autre suivant Y. Aussi, nous déterminerons deux moyennes et deux variances.

a) Les moyennes marginales :

Suivant la distribution marginale de X, c'est-à-dire quelque soit Y, on obtient la moyenne arithmétique de la distribution à un seul caractère, suivant le caractère X, notée $\bar{X}$. Elle est représentée par les couples (x_i ; $n_{i.}$). On écrit alors :

$$\bar{\mathbf{X}} = \mathbf{1/n..} \sum_{i=1}^{k} \boldsymbol{xi} \,.\, \boldsymbol{ni.} = \mathbf{1/n..} \sum_{i=1}^{k} \sum_{j=1}^{p} \boldsymbol{nijxi} \quad (\text{Puisque } n_{..} = \sum_{j=1}^{p} n_{ij})$$

Suivant la distribution marginale de Y, c'est-à-dire quelque soit X, on obtient la moyenne arithmétique de la distribution à un seul caractère, suivant le caractère Y, notée Y. Elle est définie par les couples (y_j ; $n_{\bullet j}$). On écrit alors :

$$\bar{Y} = 1/n.. \sum_{j=1}^{p} yjn.j = 1/n.. \sum_{j=1}^{p} \sum_{i=1}^{k} nijyj$$

b) Les variances marginales

Suivant la variable X, on aura :

$$V(X) = 1/n.. \sum_{i=1}^{k} n_i.(x_i - \bar{X})^2 = 1/n.. \sum_{i=1}^{k} ni.(x_i^2) - (\bar{X})^2$$ (*formule développée*)

Suivant la variable Y, on aura :

$$V(Y) = 1/n.. \sum_{j=1}^{p} n_{\bullet j}(y_j - \bar{Y})^2 = 1/n.. \sum_{j=1}^{p} n_{\bullet j}(y_j)^2 - (\bar{Y})^2$$ (*formule développée)*

2. Caractéristiques des distributions conditionnelles :

Comme signalé plus haut, il y a autant de distributions conditionnelles qu'il y a de modalités pour chacun des deux caractères. On les résume en deux écritures : distribution conditionnelle de X selon Y (y = constante) et distribution conditionnelle de Y selon X (x = constante).

a) Distribution conditionnelle de X selon Y :

Dans ce cas on déterminera autant de moyennes et de variances conditionnelles qu'il y a de modalités de Y ou de colonnes (c'est-à-dire *p*), soit une moyenne et une variance conditionnelle par colonne. On écrit alors :

$$\bar{X}_j = 1/n._j \sum_{i=1}^{k} nijxi$$ (***j*** *indique la colonne ou la modalité de Y retenue*)

Les variances seront de type :

$$V(X)_j = 1/n._j \sum_{i=1}^{k} n_{ij}(x_i - \bar{x}_j)^2 = 1/n.j \sum_{i=1}^{k} n_{ij}(x_i^2) - (\bar{X}_j)^2$$

Ainsi, si j varie de 1 à *p*, on aura *p* moyennes conditionnelles et *p* variances conditionnelles possibles de X selon Y.

b) Distributions conditionnelles de Y selon X :

Dans ce cas on déterminera autant de moyennes et de variances qu'il y a de modalités de X ou de lignes, c'est-à-dire *k*. On écrit alors :

$$\bar{Y}_i = 1/ni.\sum_{j=1}^{p} nij(yj) \qquad (i \text{ indique la ligne ou la modalité } x_i \text{ retenue})$$

Les variances seront de type :

$$V(Y)_i = 1/n_i.\Sigma_{j=1}^{p} nij\ (yj - \bar{Y}_i)^2 \quad = \quad 1/n_i.\Sigma_{j=1}^{p} n_{ij}(y_j)^2 - (\bar{Y}_i)^2 \quad (\textit{formule développée})$$

Exemple :

L'observation de niveau de risque d'accident effectuée sur 1000 conducteurs a permis de déterminer les proportions de conducteurs suivants les distances parcourues en Km et 1'âge.

Age / Distance	**< 100**	**100 - 200**	**200 - 300**	**300 - 400**	**400 - 500**
< 26	4.4	1.6	--	--	--
26 – 32	7.2	8.2	4.0	2.6	--
32 – 38	2.4	7.2	13.6	14.4	4.4
38 – 42	--	--	2.4	11.6	6.0
42 - 48	--	--	--	4.4	5.6

1. Déterminer les distributions marginales.

2. Calculer les moyennes et les variances marginales. Interpréter.

3. Calculer la moyenne de l'âge des conducteurs sachant que la distance parcourue est de 300 à 400 Km. Interpréter.

La solution donnée à cet exercice doit se faire de la même manière que le précédent. Sauf que le tableau statistique présente dans ce cas les fréquences simultanées $\underline{f_{ij}}$ et non *marginales* ni *conditionnelles*. Ce qui représente en fait une *remarque de fond* dans ce cadre.

Nous avons :

▪ La population est composée par les 1000 conducteurs avec N = n..=1000.

▪ **X :** L'âge des conducteurs.

▪ **Y :** est la distance parcourue en Km.

1. Avant de déterminer les distributions marginales et afin de rendre le calcul et le raisonnement plus simples il faut reconstituer le tableau statistique en termes des effectifs et non les fréquences relatives.

On se propose donc pour répondre à cette question, de retrouver le tableau statistique en termes des effectifs en fonction du tableau initial tel que :

Age ~~Distance~~	< 100	100 - 200	200 - 300	300 - 400	400 - 500	Marge X
< 26	4.4	1.6	--	--	--	**6**
26 – 32	7.2	8.2	4.0	2.6	--	**22**
32 – 38	2.4	7.2	13.6	14.4	4.4	**42**
38 – 42	--	--	2.4	11.6	6.0	**20**
42 - 48	--	--	--	4.4	5.6	**10**
Marge Y	**14**	**17**	**20**	**33**	**16**	**100**

Par exemple, 4,4% de l'effectif total signifie que 44 sur les 1000 conducteurs observés ont un âge inférieur à 26 ans et parcourent une distance inférieure à 100 Km.

Tableau présentant les distributions marginales en termes des effectifs.

Age ~~Distance~~	0 - 100	100 - 200	200 - 300	300 - 400	400 - 500	Marge X
20 - 26	44	16	0	0	0	**60**
26 – 32	72	82	40	26	0	**220**
32 – 38	24	72	136	144	44	**420**
38 – 42	0	0	24	116	60	**200**
42 - 48	0	0	0	44	56	**100**
Marge Y	**140**	**170**	**200**	**330**	**160**	**1000**

NB. Les deux bornes de « 20 » pour X et « 0 » pour Y sont calculées en fonction des amplitudes fréquentes alors que les effectifs n_{ij} sont estimés en utilisant les fréquences relatives et l'effectif total tel que : $n_{ij} = f_{ij} \times n..$

1. Le calcul des moyennes et des variances marginales des deux variables (x; y) l'une indépendamment de l'autre est présenté successivement comme suit :

▪ Pour la variable X :

Age	$n_{i.}$	X_i	$X_i.n_{i.}$	$X_i^2.n_{i.}$
20 - 26	60	23	1380	31740
26 – 32	220	29	6380	185020
32 – 38	420	35	14700	514500
38 – 42	200	40	8000	320000
42 - 48	100	45	4500	202500
∑	**1000**	**-**	**34960**	**1253760**

On a donc :

$$\bar{X} = \frac{\sum X_i.n_{i.}}{n_{..}} = \frac{34960}{1000} = 34{,}96$$

Chaque conducteur observé dans cet échantillon de 1000 est âgé en moyen d'environ 35 ans et ce indépendamment de la distance parcourue.

$$V(X) = \frac{\sum X_i^2.n_{i.}}{n_{..}} - \bar{X}^2 = \frac{1253760}{1000} - (34{,}96)^2 = 31.55$$

$$\delta x = \sqrt{V(X)} = \sqrt{31{,}55} = 5.6$$

En moyenne et indépendamment de la distance parcourue, l'âge de chaque conducteur pris dans cet échantillon, s'écarte de l'âge moyen (égale à 34,96 ans) positivement ou négativement d'environ 6 ans.

▪ Pour la variable Y :

Distance	0 - 100	100 - 200	200 - 300	300 - 400	400 - 500	∑
n_j	140	170	200	330	160	**1000**
Y_j	50	150	250	350	450	**-**
$Y_j.n_{.j}$	7000	25500	50000	115500	72000	**270000**
$Y_j^2.n_{.j}$	350000	3825000	12500000	40425000	32400000	**89500000**

Ce qui donne :

$$\bar{Y} = \frac{\sum Y_j.n_{.j}}{n_{..}} = \frac{270000}{1000} = 270$$

Chaque conducteur observé dans cet échantillon de 1000 parcourt une distance moyenne d'environ 270 Km et ce indépendamment de l'âgé.

$$V(Y) = \frac{\sum Y_j^2.n_{.j}}{n_{..}} - \bar{Y}^2 = \frac{89500000}{1000} - (270)^2 = 16600$$

$$\delta y = \sqrt{V(Y)} = \sqrt{16600} = 128,8$$

En moyenne et indépendamment de l'âge, la distance parcourue par chaque conducteur observé dans cet échantillon, s'écarte de la distance moyenne (égale à 270 Km) positivement ou négativement d'environ 129 Km.

2. Calcul de la moyenne de l'âge des conducteurs sachant que la distance parcourue est de 300 à 400 Km :

Il s'agit de calculer la moyenne de la distribution représentée par « l'âge des conducteurs conditionné par la distance parcourue ». Si la distance est exactement comprise entre 300 et 400 Km, c.-à-d. la 4° modalité de Y, alors la distribution sera présentée par $X_{i/j=4}$

Soit donc le tableau suivant présentant les statistiques de la distribution conditionnelle de X :

$X_{i/j=4}$	n_{i4}	X_i	$X_i.n_{i4}$
20 - 26	0	23	0
26 – 32	26	29	754
32 – 38	144	35	5040
38 – 42	116	40	4640
42 - 48	44	45	1980
∑	**330**	**-**	**12414**

Ce qui donne :

$$\bar{X}_{i/j=4} = \frac{\sum X_i.n_{i4}}{n_{.4}} = \frac{12414}{330} = 37,6 = 38$$

Ce qui signifie que chaque conducteur pris parmi les 330 conducteurs ayant parcouru une distance comprise entre 300 et 400 Km, possède un âge moyen de 38 ans.

Section 3 : Analyse de la relation entre deux variables quantitatives

Dans le cas particulier où l'on a pu mettre en évidence l'existence d'une relation linéaire significative entre deux caractères quantitatifs continus X et Y, on peut chercher à formaliser la relation moyenne qui unit ces deux variables à l'aide d'une des trois équations suivantes :

- a.X + b.Y + c = 0 : équation de la droite moyenne liant les caractères X et Y ;
- Y = a.X + b : droite de régression de Y en fonction de X ;
- X = a.Y + b : droite de régression de X en fonction de Y.

Les trois équations proposées ci-dessus correspondent à trois droites différentes, trois résumés différents du nuage de points (X,Y). La différence entre les trois droites vient du fait que les trois équations proposées correspondent à trois objectifs différents.

1. La régression linéaire simple :

Les points (xi , yi) forment un nuage dont on cherche une approximation dans un but de simplification. Mais qui dit simplification dit déformation : nous voudrions qu'elle soit minimale ; encore faut-il préciser ce que l'on entend par là. Disons tout de suite que le choix du critère sera arbitraire même si l'on tente de le justifier par des considérations plus ou moins « intuitives ».

On peut vouloir par exemple :

- Préserver au mieux les distances entre points ;
- Réserver au mieux les angles des droites joignant les points…

Il n'existe pas de moyen de satisfaire à toutes ces exigences à la fois. Il nous faut donc choisir. Nous allons chercher la meilleure droite au sens des moindres carrés.

Le principe des moindres carrés stipule que la somme des carrés des écarts est minimum. Si on note nos écarts $\boldsymbol{e_i}$, on obtient :

$\boldsymbol{e_i} = (y_i - y_c)$, avec ; $\Sigma \boldsymbol{e_i}^2 \quad \Rightarrow \quad$ Min .

La forme mathématique la plus simple permettant de relier deux variables est la forme linéaire de type « $y = ax+b$ », c'est-à-dire une droite, d'où l'appellation linéaire.

Pour déterminer l'équation de la droite d'ajustement de y en x en utilisant le principe des moindres carrés, il faut trouver les paramètres **a** et **b** de la droite d'ajustement Y = aX + b.

Nous savons que :

$$a = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\bar{X}^2 - (\bar{X})^2}$$

et $b = \bar{Y} - a\,\bar{X}$

$$a = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\bar{X}^2 - (\bar{X})^2}$$

$$a = \frac{Cov(x ; y)}{V(x)} = \frac{(xi - \bar{X})(yi - \bar{Y})}{\bar{X}^2 - (\bar{X})^2}$$

et $Y = aX + b \Rightarrow \bar{Y} = a\,\bar{X} + b \Rightarrow b = \bar{Y} - a\,\bar{X}$

2. L'analyse de la Corrélation :

En général, on rencontre trois situations concernant la relation entre deux variables[1] :

a) Une relation totale, appelée aussi relation fonctionnelle, où les variations d'une variable sont expliquées exclusivement et totalement par les variations de l'autre. Dans ce cas les deux variables sont totalement dépendantes. Ce type de relation est très rare en pratique, notamment dans le domaine socio-économique ;

b) Absence de relation, signifiant que les variations d'une variable n'ont aucun effet sur les variations de l'autre. Les deux variables sont dans ce cas indépendantes ;

[1] Madjid HADJEM, **Polycopié Cours de Statistique I**, Semestre 1, 1ere Année Licence, 1 - Département des Sciences Commerciales, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université Mouloud MAMMERI de Tizi-Ouzou, 2023/2024, P : 100, https://dspace.ummto.dz/server/api/core/bitstreams/01256031-7855-41ea-ad3c-dbb7ab165058/content

c) Une relation relative, appelée aussi liaison ou dépendance partielle, où les variations d'une variable sont en grande partie, mais non en totalité, expliquées par celles de l'autre. C'est le cas le plus fréquent dans le domaine socio-économique. Une part des variations est laissée au hasard ou aux influences de l'environnement. La dépendance partielle suppose donc une partie certaine importante expliquée par les variations de l'une des variables (y = f(x)), et une partie incertaine ou aléatoire, peu importante et réduite non expliquée par les variations de l'une des variables mais par le hasard, on l'exprime souvent par le symbole Epsilon « **ε** » pour signifier son caractère négligeable dans la formulation mathématique (y = f(x) + ε).

Dans la relation linéaire simple, on parle d'analyse de la corrélation linéaire simple. Celle-ci mesure le degré ou l'intensité de la liaison entre deux variables. Elle consiste en la mesure d'un paramètre, appelé coefficient de corrélation linéaire de Pearson.

3. Le coefficient de corrélation :

Noté ***r***, il est le rapport de la covariance entre X et Y au produit des écarts-types de X et Y. C'est un nombre sans dimension. On écrit :

. $\boxed{r = Cov(x,y) / \delta_x.\delta_y}$

On peut également écrire que :

$$\boxed{r = \sqrt{a.a'}}$$

ou bien encore :

$$\boxed{r = a.(\delta x/\delta y)}$$

r varie entre -1 et 1: $r \in [-1 ; 1]$.

Si [1]:

a) r =1, les deux variables varient dans le même sens et la liaison est *totale* ou *fonctionnelle*. Cela signifie que la droite de régression (données estimées) s'ajuste parfaitement aux données réelles.

b) r = -1, les deux variables varient en sens inverse.

[1] Madjid HADJEM, op.cit., P : 101.

c) r = 0, pas de liaison entre les deux variables. La droite de régression est alors une droite de pente a = 0, soit parallèle à l'axe des abscisses, soit parallèle à l'axe des ordonnées : les variations d'une variable n'influence pas celles de l'autre.

d) r proche de 0, la relation entre les deux variables est faible.

e) r < 0, il existe une corrélation relative et inverse ou négative entre les deux variables, toute variation d'une variable entraine une variation en sens inverse de l'autre variable.

f) r > 0, il existe une corrélation *relative* ou *partielle* et positive entre les deux variables, toute variation d'une variable entraine une variation relative, dans le même sens, de l'autre variable. Dans ces deux derniers cas, une grande partie de la variation de Y est expliquée par les variations de X. L'autre partie, infime, est expliquée par le hasard.

g) r proche de ±1, la relation est forte.

Exemple :

Une agence de voyage a réalisé une étude pour savoir s'il existe une relation entre le prix moyen mensuel de billet (X en 1000 DA) et le nombre de clients qui ont réservé un voyage (Y). Les résultats sont donnes ci-après :

X	7	9	9.5	11	12.5	13	15	17	19	21
Y	90	110	130	150	170	190	210	230	250	270

1. Tracer le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet. Conclure quant à la nature de la relation.

2. Calculer un paramètre qui permet de mesurer l'intensité de la relation linéaire entre les deux variables étudiées. Interpréter votre résultat.

3. Déterminer l'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet.

4. Déterminer le nombre de clients qui ont réservé pour des prix moyens de billet qui sont respectivement de 8000 DA et 10000 DA.

Solution :

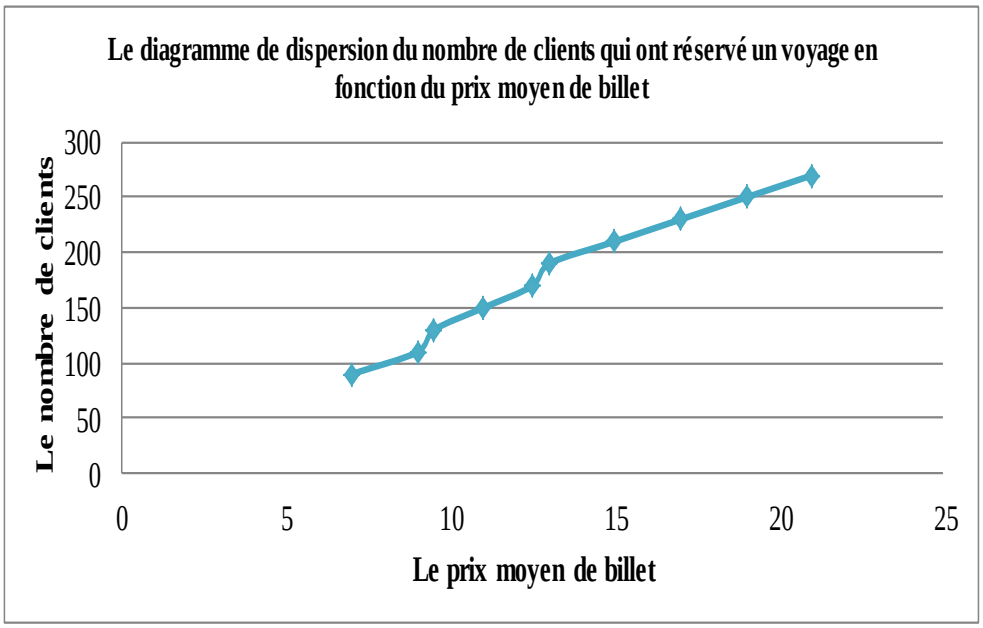
La population est représentée par dix (10) observations portant sur dix (10) clients d'une agence de voyage.

Alors que les variables utilisées sont définies par :

▪ **X** : « le prix moyen mensuel de billet en 1000 DA » comme variable exogène.

▪ **Y** : « le nombre de clients qui ont réservé un voyage suite à un prix affiché par l'agence » comme variable endogène.

1. Le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet est présenté comme suit :



On constate que le nombre de clients ayant réservé un voyage évolue positivement et linéairement au prix moyen de billet.

2. Le paramètre qui permet de **mesurer l'intensité de la relation** linéaire entre les deux variables étudiées est représenté par le **coefficient de corrélation** (r) définit comme suit :

$$r = \frac{Cov(x,y)}{\delta(x) * \delta(y)}$$

D'où le tableau et les résultats suivants :

	$\bar{X}$	$\bar{Y}$	$\bar{X}^2$	$\bar{Y}^2$	$\bar{X}Y$
	X	**Y**	**X^2**	**Y^2**	**X.Y**
	7	90	49	8100	630
	9	110	81	12100	990
	9.5	130	90.25	16900	1235
	11	150	121	22500	1650
	12.5	170	156.25	28900	2125
	13	190	169	36100	2470
	15	210	225	44100	3150
	17	230	289	52900	3910
	19	250	361	62500	4750
	21	270	441	72900	5670
Σ	**134**	**1800**	**1982.5**	**357000**	**26580**

Avec :

$$r = \frac{Cov(x,y)}{\delta(x) * \delta(y)}$$

$$r = \frac{XY - \bar{X} * \bar{Y}}{\sqrt{\bar{X}^2 - (\bar{X})^2} * \sqrt{\bar{Y}^2 - (\bar{Y})^2}}$$

▪ La moyenne arithmétique de X ($\bar{X}$) :

$$\bar{X} = \frac{\sum X_i}{N}$$ / Sachant que N = 10 (Nombre de situations)

$$\bar{X} = \frac{134}{10}$$

$$\bar{X} = 13,4$$

Le prix moyen des prix mensuels fixé pour les 10 clients (ou 10 derniers mois) pris dans cet échantillon est de 13400 DA.

▪ La moyenne arithmétique au carrée des X ($\bar{X})^2$:

$(\bar{X})^2 = (13,4)^2$

$(\bar{X})^2 = 179,56$

▪ La moyenne arithmétique des X au carrée ($\bar{X}^2$) :

$$\bar{X}^2 = \frac{\sum X_i^2}{N}$$

$$\bar{X}^2 = \frac{1982,5}{10}$$

$$\bar{X}^2 = 198,25$$

▪ La moyenne arithmétique de Y ($\bar{Y}$) :

$$\bar{Y} = \frac{\sum Y_i}{N}$$

$$\bar{Y} = \frac{1800}{10}$$

$$\bar{Y} = 180$$

Le nombre moyen des clients ayant réservé un voyage sur les dix derniers mois est de 180 clients.

▪ La moyenne arithmétique au carrée des Y ($\bar{Y}$)^2 :

$(\bar{Y})^2 = (180)^2$

$(\bar{Y})^2 = 32400$

▪ La moyenne arithmétique des Y au carrée ($\bar{Y}^2$) :

$$\bar{Y}^2 = \frac{\sum Y_i^2}{N}$$

$$\bar{Y}^2 = \frac{357000}{10}$$

$$\bar{Y}^2 = 35700$$

▪ La moyenne arithmétique de $\bar{X} * \bar{Y}$:

$\bar{X} * \bar{Y} = (13,4) * (180)$

$\bar{X} * \bar{Y} = 2412$

▪ La moyenne arithmétique de XY :

$$XY = \frac{\sum X_i Y_i}{N}$$

$XY = \frac{26580}{10}$

$XY = 2658$

- Cov(x,y) :

$Cov(x,y) = XY - \bar{X} * \bar{Y}$

$Cov(x,y) = 2658 – 2412$

$Cov(x,y) = 246$

- L'écart type δ(x) :

$\delta(x) = \sqrt{\bar{X}^2 – (\bar{X})^2}$

$\delta(x) = \sqrt{198,25 – 179 ,56}$

$\delta(x) = \sqrt{18,66}$

$\delta(x) = 4,32$

En moyenne, le prix de billet s'écarte du prix moyen calculé sur les 10 derniers mois d'environ 4320 DA.

- L'écart type δ(y) :

$\delta(y) = \sqrt{\bar{Y}^2 – (\bar{Y})^2}$

$\delta(y) = \sqrt{35700 – 32400}$

$\delta(y) = \sqrt{3300}$

$\delta(y) = 57,44 \approx 58$

En moyenne, le nombre des clients ayant réservé un voyage s'éloigne du nombre moyen des clients sur les dix derniers mois d'environ 58 clients.

- $r = \frac{XY - \bar{X} * \bar{Y}}{\sqrt{\bar{X}^2 – (\bar{X})^2} * \sqrt{\bar{Y}^2 – (\bar{Y})^2}}$

$r = \frac{2658 – 2412}{4,32 * 57,44}$

$r = \frac{246}{248,14}$

r = 0,99

Le prix de billet et le nombre de clients sont dépendants positivement.

Enfin, on trouve :

Les deux variables observées dans cette population sont donc ***fortement*** et ***positivement corrélées*** puisque jusqu'à ***99%*** de la variabilité du nombre de clients est expliquée par la variabilité du prix moyen fixé par mois. Ainsi, il y a seulement ***1%*** des changements susceptibles sur le nombre de clients de cette agence qui est expliqué par ***les autres variables non observées***.

3. L'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet est représentée par :

Y = aX + b $\Rightarrow$ Dont les valeurs des deux paramètres ***a*** et ***b*** sont estimées par :

- $a = \frac{Cov(x,y)}{V(X)}$

$$a = \frac{\overline{XY} - \overline{X} * \overline{Y}}{\overline{X^2} - (\overline{X})^2}$$

$$a = \frac{2658 - 2412}{198,25 - 179,56}$$

$$a = \frac{246}{18,69}$$

a = 13,16

Ce qui signifie que lorsque le prix moyen par mois ***augmente*** de ***1%***, le nombre de clients ***augmente aussi*** d'environ ***13%***.

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

b = 180 – (13,16) * (13,4)

b = 180 – 176,34

b = 3,65

Ce qui représente le nombre constant de clients qui peut réserver un voyage ***indépendamment*** du prix moyen par mois du billet.

Ce qui donne, enfin, l'équation de la droite d'ajustement donnée par :

$Y = aX + b$

$Y = 13{,}16X + 3{,}65$

4. Sur la base de la droite précédente on peut donner les prévisions suivantes :

▪ Si le prix moyen de billet par mois est fixé à 8000 DA alors le nombre de clients qui peut réserver sera prévu à :

$Y = (13{,}16) * (8) + 3{,}65$
$Y = 105{,}28 + 3{,}65$
$Y \approx 109$
Soit donc un nombre prévisionnel d'environ 109 clients.

▪ Si, par contre, le prix moyen de billet par mois est fixé à 10000 DA alors le nombre de clients qui peut réserver sera prévu à :

$Y = (13{,}16) * (10) + 3{,}65$
$Y = 131{,}6 + 3{,}65$
$Y = 135{,}25$
$Y \approx 135$

Soit donc un nombre prévisionnel d'environ 135 clients.

ⵜⴰⵙⴷⴰⵡⵉⵜ ⵎⵓⵍⵓⴷ ⵎⵎⴻⵔⵉ ⴰⵜ ⵜⵣⵉⵔⵉ
Ministry of Higher Education & Scientific Research
Mouloud MAMMERI University of Tizi Ouzou – Algeria
Faculty of Economics, Business & Management
Department of Economics

Academic Year : 2024 – 2025
1st Year LMD - Section E3
Module : Statistics I
Worksheet 6th
Lecturer : G. BELBACHIR
Email : gouraya.belbachir@ummto.dz

To access the online answer key, please scan the **QR** code using the anonymous login.

Exercise 1 :

Le tableau suivant donne la répartition des salariés d'une entreprise de bâtiment selon le nombre d'enfants à charge X et les dépenses mensuelles Y en milliers de DA.

Enfants à charge X / Dépenses Y	13 -20	20 - 25	25 - 30
1	23	18	26
2	16	22	24
3	31	16	13
4	21	14	08

1. Donner les deux distributions marginales.

2. Donner la dépense mensuelle moyenne pour salariés avec 4 enfants à charge.

3. Calculer la variance marginale de X.

Exercise 2 :

L'observation de niveau de risque d'accident effectuée sur 1000 conducteurs a permis de déterminer les proportions de conducteurs suivants les distances parcourues en Km et 1'âge.

Age / Distance	**< 100**	**100 - 200**	**200 - 300**	**300 - 400**	**400 - 500**
< 26	4.4	1.6	--	--	--
26 – 32	7.2	8.2	4.0	2.6	--
32 – 38	2.4	7.2	13.6	14.4	4.4
38 – 42	--	--	2.4	11.6	6.0
42 - 48	--	--	--	4.4	5.6

1. Déterminer les distributions marginales.

2. Calculer les moyennes et les variances marginales. Interpréter.

3. Calculer la moyenne de l'âge des conducteurs sachant que la distance parcourue est de 300 à 400 Km. Interpréter.

Exercise 3 :

Une agence de voyage a réalisé une étude pour savoir s'il existe une relation entre le prix moyen mensuel de billet (X en 1000 DA) et le nombre de clients qui ont réservé un voyage (Y). Les résultats sont donnes ci-après :

X	7	9	9.5	11	12.5	13	15	17	19	21
Y	90	110	130	150	170	190	210	230	250	270

1. Tracer le diagramme de dispersion du nombre de clients qui ont réservé un voyage en fonction du prix moyen de billet. Conclure quant à la nature de la relation.

2. Calculer un paramètre qui permet de mesurer l'intensité de la relation linéaire entre les deux variables étudiées. Interpréter votre résultat.

3. Déterminer l'équation de la droite donnant le nombre de client qui ont réservé un voyage en fonction du prix moyen de billet.

4. Déterminer le nombre de clients qui ont réservé pour des prix moyens de billet qui sont respectivement de 8000 DA et 10000 DA.

~ ... ~ ... ~

Conclusion générale

Conclusion générale :

Le cours débute par une introduction au domaine de la statistique descriptive, en exposant les définitions et les notions de base, puis vers la présentation des données avant de passer au calcul des paramètres de mesure des variations des données. Suite à cela on élargit le champ d'analyse en étudiant les nombres indices et, en dernier lieu, les distributions à deux caractères et l'analyse de la relation entre elles. Nous sommes ainsi passés du simple au complexe.

Les chapitres exposés de cette façon, c'est à dire enchaînée et suivant un raisonnement cohérent et logique, répondant largement aux principes de la statistique descriptive, ont pour but de faciliter à l'étudiant l'assimilation des connaissances proposées et de l'imprégner du raisonnement de la statistique qui est, on ne peut plus, rationnel et scientifique. Cela permet d'accompagner l'étudiant dans sa transition de l'enseignement du secondaire à l'enseignement supérieur. Au terme de ces chapitres, l'étudiant peut passer à l'étape plus complexe de l'analyse statistique, à savoir ; les probabilités et les variables aléatoires. C'est l'objet de l'enseignement dispensé en Statistique 2 au second semestre.

Bibliographie

Bibliographie :

1. ANDERSON D. R., SWEENEY D. J., CAMM J. D., WILLIAMS T. A., COCHRAN J. J, **Statistiques pour L'Economie et la Gestion**, De Boeck, Bruxelles, 2015.

2. BRESSOUD Etienne, KAHANE Jean Claude, **Statistique Descriptive : avec Excel et la Calculatrice**, Collection Synthex, Pearson Education France, 2009.

3. CHRISTOPHE Hurlin, **Statistique et Probabilités en Economie-Gestion**, Dunod, 2015.

4. DELMAS Jean-François, **Introduction au Calcul des Probabilités et à la Statistique**, Les presses de l'ENSTA, Paris, 2010.

5. DESMARIS Christian, **Informatique & Statistique**, Tome 2, Conférence de Méthode, Institut d'Études Politiques de Lyon, Université Lumière, Lyon II, France, 2003-2004.

6. DOANE P. David, SEWARD E. Lori, **Applied Statistics in Business and Economics**, Fifth Edition, McGraw-Hill Education, 2016.

7. FLOYD John E., **Statistics for Economists : A Beginning**, University of Toronto, July 2, 2010.

8. FLUX Jamie, **Business Statistics All in One Skills Practice Workbook: With Full Step by Step Solutions**, Broché, ISBN-13 979-8340780041, Septembre 2024.

9. FOUCAN L., **Probabilités et Statistiques**, PAES, 2012 – 2013.

10. GOLDFARB B., PARDOUX C., **Introduction à la Méthode Statistique**, 6° Edition, Dunod, 2011.

11. GOLDFARB Bernard, PARDOUX Catherine, **Introduction à la méthode statistique : Manuel et exercices corrigés**, 6e édition, DUNOD, Paris, France, 2011.

12. HADJEM Madjid, **Polycopié Cours de Statistique I**, Semestre 1, 1ere Année Licence, 1 - Département des Sciences Commerciales, Faculté des Sciences Economique, Commerciale et des Sciences de Gestion, Université Mouloud MAMMERI de Tizi-Ouzou, 2023/2024, https://dspace.ummto.dz/server/api/core/bitstreams/01256031-7855-41ea-ad3c-dbb7ab165058/content

13. HAMDANI Hocine, **Statistique Descriptive**, OPU, Alger, 2005.

14. LANE David, **Introduction to Statistics**, Rice University, 2003.

15. LEBOUCHER Lucien, VOISIN Marie-José, **Introduction à la Statistique Descriptive : Cours et Exercices avec Tableur**, Cepaduès-Editions, Toulouse, France, 2011.

16. LECOUTRE J. P., **Statistique et Probabilités**, Dunod, 2002.

17. Partie 2 : Les chroniques, chapitre 2 : **construction des indices**.

18. POSIERE Jean-Pierre, **Mathématiques Appliquées à la Gestion**, Gualino, Paris, 2005.

19. ROSS Sheldon, **A First Course in Probability**, MA : Pearson, Boston, 2019.

20. SAPORTA G., **Probabilités, Analyse des Données et Statistique**, Editions Technip, 2006.

21. VEYSSEYRE Renée, **Aide-mémoire Statistique et probabilités pour l'ingénieur**, 2e édition, Dunod, Paris, France, 2006.

22. VORNETTI Patrica, **Statistique & Informatique**, L1, Sc Eco, PPT, 2016.