Cours de statistiques, probabilités et prévisions
Licence 3 LSG IAE - Tours 2017- 2018

Adeline MBASSI

ATER en Sciences de Gestion / IAE-Tours

Thématiques de travail: Santé au travail ; styles de leadership; comportements organisationnels; bien-être au travail; capital psychologique

adeline.mbassi@univ-tours.fr
Laboratoire de recherche VALLOREM
Bâtiment A – Bureau A 247

Plan du cours (3 séances de CM et 12 séances de TD)

Partie I- Les statistiques (cas univariée)

Partie II- Les statistiques (cas bivariée)

Partie III- Lois de probabilités et prévisions

- Bibliographie

I- Les statistiques

1. Introduction

Chaque jour, le manager est amené à prendre des décisions pour gérer son entreprise face à une inflation d'informations. Pour cela, il doit se reposer sur **une démarche de questionnement rigoureuse** dans le but de définir l'objectif de l'enquête, recueillir les informations nécessaires , les analyser et les expliquer.

1.1. Définition et exemples d'application de la statistique

La statistique est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes **aléatoires**, c'est-à-dire dans lesquels le hasard intervient.

L'analyse des données est utilisée pour **décrire** les phénomènes étudiées , **faire des prévisions** et **prendre des décisions** à leur sujet. Les données étudiées peuvent être de diverses natures (qualitatives et /ou quantitatives).

I- Les statistiques

- Quelques exemples d'utilisation de la statistique dans divers domaines:

- **Economie, assurance , finance :** prévisions économétriques, analyse de la consommation des ménages, fixation des primes d'assurance et franchises, études quantitatives de marchés, gestion de portefeuille, évaluation d'actifs financiers, ...
- **Biologie, médecine:** essais thérapeutiques, épidémiologie, dynamique des populations, analyse du génome, ...
- **Sciences de l'information et de la communication:** traitement des images et des signaux, reconnaissance des formes et de la parole, analyse exploratoire des grandes bases de données, analyse des réseaux de communication, ...
- **Gestion d'entreprise :** prédiction des comportements des salariés au travail, des comportements d'achats des consommateurs ; fréquentation du point de vente d'une enseigne etc.

A retenir: *Les données statistiques sont entachées d'incertitudes et présentent des variations pour plusieurs raisons ; il y'a donc intervention **du hasard** et **des probabilités**. Et, le rôle de la statistique ici est de permettre de maîtriser au mieux cette incertitude pour extraire des informations utiles de données .*

I- Les statistiques

Dans le cadre de ce cours, nous ne nous attarderons pas sur les méthodes de collecte de données. Si l'on omet donc cette partie, les méthodes statistiques peuvent être regroupées en deux classes: les ***méthodes descriptives*** et les ***méthodes inférentielles***.

1.2. Statistique descriptive VS statistique inférentielle

❑ Statistique descriptive

La statistique descriptive , statistique exploratoire ou analyse de données a pour objectif de:

- permettre le résumé de l'information contenue dans les données de façon synthétique et efficace

- permettre de dégager les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour une étude ultérieure sophistiquée.

A retenir: *Ici les probabilités n'interviennent presque pas car, l'on mène l'étude sur tous les* ***individus*** *de la* ***population*** *étudiée.*

I- Les statistiques

❑ Statistique inférentielle

La statistique inférentielle ou statistique inductive regroupe les méthodes dont le but est de préciser un phénomène sur une population globale à partir de son observation sur **partie restreinte** de cette population (exemple des sondages).

Elle permet principalement :

- d'effectuer et de réaliser des inférences et des prédictions à partir des données recueillies
- de prendre des décisions au regard des observations

A retenir: *Notons qu'ici, il s'agit d'induire (ou encore d'inférer) du particulier au général. Le plus souvent, il faut pour cela proposer des modèles probabilistes du phénomène aléatoire étudié et savoir gérer les risques d'erreurs.* ***Les probabilités y ont donc un rôle majeur*** *.*

D'un point de vue méthodologique, on notera que la statistique descriptive ***précède en général*** *la statistique inférentielle dans une démarche de traitement de données : ces deux aspects de la statistique se complètent bien plus qu'ils ne s'opposent.*

I- Les statistiques

2. La statistique descriptive

Les statistiques descriptives servent à décrire et à présenter l'information contenue dans des données collectées sur un groupe d'individus. Les outils de la statistique descriptive varient selon que l'analyse concerne une, deux ou plusieurs variables (*analyse univariée, bivariée et multivariée*).

2.1. Notions de base et illustrations

La population (ou population statistique) est constituée de l'ensemble des individus objets de l'étude. C'est aussi, l'ensemble des éléments homogènes sur lesquels porte l'étude.

Un individu (ou unité statistique) désigne tout élément de la population considérée. C'est aussi, une unité de la population.

Illustration 1: si l'on s'intéresse aux notes d'un groupe d'étudiants, la population constitue ce groupe d'étudiants. Ici, l'individu est tout étudiant du groupe

Illustration 2: si l'étude porte sur des entreprises du BTP de moins de 50 salariés, la population constitue m'ensemble des entreprises du BTP de moins de 50 salariés de la région, ville concernée . L'individu étant l'une de ces entreprises de moins de 50 salariés

A retenir: *Une même étude peut amener à définir plusieurs populations. De ce fait, le terme population peut désigner aussi bien des humains que des objets, des villes, des pays, des entreprises, des logements etc.*

I- Les statistiques

2.1. Notions de base et illustrations

Echantillon: Dans une étude statistique, il est fréquent que l'on n'observe pas la population tout entière (par exemple, on n'observe pas tous les véhicules ayant circulé un jour donné dans Toulouse, mais seulement ceux étant passés dans certains points particuliers). Les observations du phénomène considéré sont donc réalisées sur une partie restreinte de la population, appelée échantillon. On appelle donc **échantillon** le sous-ensemble de la population sur lequel sont effectivement réalisées les observations. C'est aussi un groupe d'individus extrait de de la population.

Taille de l'échantillon: c'est le cardinal de l' échantillon, autrement dit c'est le nombre d'individus qu'il contient (échantillon de taille 800, de taille 1000…). En général, on note **n** la taille de l´échantillon considéré.

Enquête (statistique): c'est l' opération consistant à observer (ou mesurer, ou questionner…) l'ensemble des individus d'un échantillon (ou éventuellement, de la population complète).

I- Les statistiques

2.1. Notions de base et illustrations

Recensement: enquête dans laquelle l'échantillon observé est en fait **la population tout entière**(on parle aussi d'enquête exhaustive).

En France, on organise ainsi, de façon plus ou moins régulière, le recensement général de la population, le recensement général agricole...

Sondage: c'est, au contraire, une enquête dans laquelle l'échantillon observé est un sous-ensemble

strict de la population (on parle, dans ce cas, d'enquête non exhaustive). Les exemples de sondages dans la vie courante sont, de nos jours, légion.

A retenir: *Le principe des sondages est d'étendre à l'ensemble de la population les enseignements tirés de l'étude de l'échantillon. Pour que cela ait un sens, il faut que l' échantillon soit représentatif de la population.*

I- Les statistiques

2.1. Notions de base et illustrations

Variable statistique: c'est un ensemble de critères pertinents au regard de l'étude qui permettent de décrire chaque individu. C'est aussi, une caractéristique (salaire, âge, sexe…) définie sur la population et observée sur l'échantillon.

Illustration: Tableau portant sur la population, individu et variables statistiques

Population	Individu	Variables statistiques
Ensemble des abonnés d'un opérateur de téléphone	Un abonné	Age, sexe, services possédés, montant de la facture mensuelle, etc.
Entreprises de BTP	Une entreprise	Nombre de salariés, chiffre d'affaires, délais de paiement, etc.
Usagers d'un restaurant d'entreprise	Un usager	Emploi, âge, sexe, fréquentation, dépense, appréciation
Ensemble des pièces produites par une chaîne de production	Une pièce	Diamètre, poids, robustesse, etc.

2.2. Variables qualitatives et quantitatives

Une variable statistique peut être de différentes natures: qualitative ou quantitative; discrète ou continue; nominales ou ordinales.

- Une variable statistique est dite de **nature qualitative** si ses modalités ne sont pas mesurables. **Les modalités** correspondent aux valeurs possibles de la variable statistique.
- Une variable statistique est dite de **nature quantitative** si ses modalités sont mesurables. Les modalités d'une variable quantitative sont des nombres liés à l'unité choisie, qui doit toujours être précisée.

Illustration: Tableau de variables qualitatives ou quantitatives

Variable	Type
Catégorie socioprofessionnelle de l'usager	?
Chiffre d'affaires de l'entreprise	?
Nombre d'enfants	?
Année de naissance	?

I- Les statistiques

2.3. Variables qualitatives nominales ou ordinales

Ici, il faut noter que les réalisations possibles d'une variable qualitative s'appellent des modalités. Par exemple, la variable sexe présente deux modalités: femme / homme

- Une variable statistique qualitative est **dite nominale** si ses modalités ne sont pas naturellement ordonnées ou dont les modalités ne peuvent être classées selon un ordre préétabli.
- Une variable statistique qualitative est **dite ordinale** si l'ensemble de ses modalités peut être doté d'une relation d'ordre ou peuvent être classées.

Illustration: Tableau de variables qualitatives et de modalités

Variable	Modalités possibles	Variable qualitative nominale/ ordinale
CSP	Ouvrier qualifié, agent de maîtrise, ingénieur, etc.	?
Sexe	Femme; homme	?
Jugement	Très insatisfait, insatisfait, satisfait, très satisfait	?

2.3. Variables quantitatives discrètes ou continues

A noter que sur ces types de variables, on peut réaliser des opérations mathématiques telles que les calculs des moyennes etc.

- Une variable statistique quantitative est **dite discrète** si l'ensemble de ses modalités est un ensemble fini ou dénombrable . C'est aussi une variable qui prend un nombre limité de valeurs entières.
- Une variable statistique quantitative est **dite continue** si l'ensemble de ses modalités n'est pas dénombrable. Elle peut prendre toutes les valeurs dans un intervalle donné.

Illustration: Tableau de variables quantitatives et de valeurs possibles

Variable	Valeurs possibles	Variable quantitative discrète /continue
Nombre d'enfants	0,1,2, etc.	?
Note de satisfaction (sur 10)	0,1,2,…,9,10.	?
Prix payé pour un repas	5,5 £, 6,8 £, etc.	?

I- Les statistiques

3. Rappel des grandes notions statistiques et application

Dernièrement, on a vu qu'une variable statistique pouvait revêtir plusieurs formes: qualitative (nominale; ordinale), quantitative (discrète ; continue).

Lors d'une étude statistique, il est primordial de bien identifier la population étudiée et de préciser avec rigueur la ou les variables relevées sur chacun des individus de la population ou de l'échantillon la représentant.

Par ailleurs, les données brutes recueillies lors d'une enquête doivent faire l'objet d'une description détaillée et sont présentées sous forme de tableaux appelés **tableaux statistiques.**

Un tableau statistique ou tableau de distribution est donc une représentation chiffrée d'un fait social , économique etc. construit à partir d'une ou plusieurs variables (lignes, colonnes) chacune caractérisée par une ou plusieurs modalités (âge, sexe, année,…). Ce tableau révèle la distribution statistique en présentant les couples de **type (x_i;n_i)** où **x_i sont les modalités** et les **n_i leurs effectifs respectifs**, i entier de 1 à r, si r désigne le nombre de modalités du caractère.

<u>A retenir:</u> *Il faut noter qu'il est également possible de présenter la distribution des fréquences , c'est-à-dire les couples de type **(x_i; f_i).***

3. Rappel des grandes notions statistiques et application

Illustration: De la série brute à la présentation des statistiques

La liste suivante est composée de prénoms d'un groupe d'étudiants, suivis entre parenthèses du nombres de films que chacun d'entre eux a vus au cours du mois dernier:

Voici **la série statistique** y associée (ensemble de valeurs obtenues à l'observation d'un phénomène ou d'une population).

Pierre (3), Paul (2), Jacques (2), Ralph (3), Abdel (1), Sidonie (2),Henri (0), Paulette (1), Farida (2), Laure (2), Kevin (0), Carole (3), Marie-Claire (0), Jeanine (3), Julie (2),Ernest (3), Cindy (3), Vanessa (2), José (1), Aurélien (1).

Questions:

1) Déterminez:
   a. La population étudiée;
   b. La variable étudiée.
2) Précisez:
   a. La nature de la variable;
   b. Les modalités de la variable
3) Construisez le tableau statistique associé à la distribution des effectifs

I- Les statistiques

3. Rappel des grandes notions statistiques et application

Solution:

1) a. La population étudiée est le groupe d'étudiants

b. La variable étudiée est X= « nombre de films que chacun d'entre eux a vus au cours du mois dernier ».

2) a. La variable étudiée est quantitative discrète.

b. L'ensemble M des modalités est M = (0; 1; 2 ;3).

A retenir: *Les modalités correspondent aux valeurs possibles de la variable statistique. Une variable statistique définit une partition sur une population , chaque individu appartenant à une et une seule modalité.*

Si le nombre de modalités est noté r, l'ensemble des modalités de la variable X sera noté:

$$M = (x_1 \ ; \ x_2 \ ; \ \dots ; \ x_i \ ).$$

I- Les statistiques

3. Rappel des grandes notions statistiques et application

- **L'effectif** (aussi appelé fréquence absolue) de la modalité x_i est noté n_i et désigne le nombre d'individus de la population présentant la modalité x_i. C'est aussi le nombre de fois où la valeur apparait pour cette série.

- **L'effectif total** est le nombre de valeurs dans une série statistique et est représenté par **n**.

L'effectif total de la population n est alors:

Remarque 1 : Le symbole sigma. *Nous avons utilisé ci-dessus le symbole Σ (sigma majuscule). Il s'agit tout simplement d'une notation permettant de raccourcir certaines écritures. Ainsi, lorsqu'on fait la somme des valeurs indicées n_i (les effectifs de la série), au lien d'écrire $n = n_1 + \cdots + n_i + \cdots + n_r$, il est plus commode d'écrire $n = \sum_{i=1}^{r} n_i$. On peut également écrire :*

$$n = \sum_{i=1}^{r} n_i = \sum_{i=1}^{i=r} n_i = \sum_{1}^{r} n_i.$$

Bien entendu, toutes ces écritures représentent la même quantité.

3. Rappel des grandes notions statistiques et application

Construisez le tableau statistique associé à la distribution des effectifs

Illustration 1:

- La première colonne comporte les **modalités x_i de X**
- La seconde colonne comporte **les effectifs (n_i) associés à chacune de ces modalités**

Le tableau statistique associé à X est le suivant:

Modalités (x_i)	Effectifs (n_i)
0	3
1	4
2	7
3	6

L'effectif total est : **n = 20**

I- Les statistiques

3. Rappel des grandes notions statistiques et application

Illustration 2:

Prenons la série **2, 2, 4, 34, 11 , 4, 2**

- L'effectif total pour cette série est de **7** puisqu'il y a **7 valeurs** (des nombres ici) .
- L'effectif de la valeur 2 est de **3** puis que la valeur 2 apparaît **3 fois** dans la liste.
- Pour la valeur 4 c'est **2** , puisque le nombre 4 apparaît à **2 reprises**.

Illustration 3:

Prenons la série: **bleu, bleu, vert, noir, rouge, vert , bleu, noir, noir**

- L'effectif total pour cette série est de **9** puisqu'il y a **9 valeurs** (Ici, des couleurs).
- L'effectif de la valeur bleu est de **3** puisque le bleu apparaît **3 fois** dans la liste.
- Pour la valeur vert c'est **2** , puisque le vert apparaît à **2 reprises**.

I- Les statistiques

3. Rappel des grandes notions statistiques et application

La fréquence de la modalité xi est notée fi et est définie par :

$$f_i = \frac{n_i}{n} \quad \text{ou} \quad f_i = \frac{n_i}{N} \times 100.$$

Elle exprime la proportion d'individus présentant une modalité donnée. C'est aussi le quotient (division) de l'effectif de la valeur par l'effectif total. Elle peut s'exprimer sous la forme d'un nombre décimal (en général avec une précision de quatre chiffres après la virgule) ou sous la forme d'un pourcentage.

Illustration 1:

Prenons la série: **2, 2, 4, 34, 11, 4, 2, 1, 9, 9**

L'effectif de la valeur 2 est de **3** puisque le nombre 2 apparaît **3 fois** dans la liste.

L'effectif total pour cette série est de **10** puisqu'il y a **10 valeurs**.

La fréquence = (effectif de la valeur / effectif total) = **3 / 10**

I- Les statistiques

3. Rappel des grandes notions statistiques et application

Illustration 2:

Construisez le tableau statistique associé à la distribution des fréquences en sachant que :

Effectif total n= 20

Nombre de films vus	0	1	2	3
Effectifs	3	4	7	6
Fréquences	0,15	0,2	0,35	0,3
P_i (%)	15	20	35	30

3. Rappel des grandes notions statistiques et application

- L' étendue d'une série statistique: $x_r - x_1$

c'est la différence entre les valeurs extrêmes. Pour la calculer, il suffit de soustraire la plus grande valeur (maximum) et la plus petite valeur (minimum) de la série.

Illustration:

Prenons la série: **2, 2, 4, 11, 4, 2, 1, 9, 9**

- Les valeurs extrêmes sont **11** (plus grande valeur) et **1** (la plus petite)

- L'étendue pour cette série est de **11 – 1 =10** (Différence entre les valeurs extrêmes).

A retenir: *L'étendue donne des informations sur ce qu'on appelle la dispersion de la série statistique :*

*- Si **l'étendue est très petite**, alors il y a peu d'écart entre toutes les valeurs de la série. Celle-ci est homogène.*

*- Si **au contraire l'étendue est grande**, alors l'écart est important entre la plus petite et la plus grande valeur.*

I- Les statistiques

3. Rappel des grandes notions statistiques et application

Illustration 1:

Quelle est l'étendue de cette série: **3; 4; 7; 8: 12** ?

L'étendue de cette série est : ???

Illustration 2:

Nous avons une série de notes obtenues par la classe de 3eme C. Que puis- je dire si l'étendue de cette série est grande:

- Les élèves ont à peu près tous le même niveau
- On ne peut rien en déduire
- Le niveau est très différent suivant les élèves

Illustration 3:

Quelle est la fréquence de la valeur 4 pour la série: **4; 1; 2; 4; 5; 6; 4** ?

Quelle est l'étendue de cette série?

3. Rappel des grandes notions statistiques et application

❑ Effectifs cumulés: définition

Résultat de l'addition, de proche en proche, des effectifs d'une distribution observée, soit en commençant par le 1er (**effectifs cumulés croissants n_i CC**):

$N_1 = n_1$, $N_2 = n_1 + n_2$, ... , $N_i = n_1 + n_2 + ... + n_i$

Soit en commençant par le dernier **(effectifs cumulés décroissants n_i CD)**:

$N'_K = n_K$, $N'_{K-1} = n_K + n_{K-1}$, ... , $N_i' = n_K + n_{K-1} + ... + n_i$

Illustration 1

x_i	0	1	2	3
n_i	3	4	7	6
n_i CC	3	(3+4) = 7	(7+7)= 14	(14+6) = 20
n_i CD	20	(20- 3) = 17	(17-4) = 13	(13-7) = 6

3. Rappel des grandes notions statistiques et application

Illustration 2: **(cas d'une variable quantitative continue)**

Au cours d'une journée, la caisse centrale d'un hypermarché a enregistré le montant des dépenses effectuées.

1) Complétez le tableau en calculant les effectifs cumulés croissants et décroissants.
2) En utilisant le tableau dites combien de dépenses sont : a. Inférieures à 50 £… ; b. Supérieures ou égales à 30 euros…

Montant des dépenses (xi)	Effectif (ni)	ni CC	ni CD
[0; 20 [	35	35	125
[20; 40 [	41	76	90
[40; 60 [	30	106	49
[60; 100 [	12	118	19
[100; 140 [	5	123	7
[140; 200 [	2	125	2
Total	n = 125	-	-

3. Rappel des grandes notions statistiques et application

Illustration 2: **(cas d'une variable quantitative continue)**

Au cours d'une journée, la caisse centrale d'un hypermarché a enregistré le montant des dépenses effectuées.

1) Complétez le tableau en calculant les effectifs cumulés croissants et décroissants.
2) En utilisant le tableau dites combien de dépenses sont : a. Inférieures à 50 £… b. Supérieures ou égales à 30 euros…

Solution:

2.a. Les dépenses inférieures à 50£ équivaut à 106 puisque les effectifs cumulés des modalités suivantes: **[0; 20 [; [20; 40 [; [40; 60 [** correspondent à:

$$n_i\ CC = 35+41+30$$
$$= 106$$

2.b. Les dépenses supérieures ou égales à 30£ correspondent aux modalités suivantes:

[20; 40 [; [40; 60 [; [60; 100 [; [100; 140 [; [140; 200 [donc:

$$n_i\ CC = 41+30+12+5+2$$
$$=90$$

3. Rappel des grandes notions statistiques et application

Fréquences cumulées: définition

Résultat de l'addition, de proche en proche, des fréquences d'une distribution observée, soit en commençant par le 1er (**fréquences cumulées croissantes F_i *CC***):

$F_1 = f_1$, $F_2 = f_1 + f_2$, ... , $F_i = f_1 + f_2 + ... + f_i$

Soit en commençant par le dernier (**fréquences cumulées décroissantes F_i CD**):

$F'_K = f_K$, $F'_{K-1} = f_K + f_{K-1}$, ... , $F'_i = f_K + f_{K-1} + ... + f_i$

Illustration

3) En vous appuyant sur l'exercice précédent, complétez le tableau en calculant les fréquences cumulées croissantes et décroissantes.

3. Rappel des grandes notions statistiques et application

Solution: tableau des fréquences cumulées croissantes et décroissantes

Classe x_i	n_i	n_i CC	n_i CD	$f_{i\ (\%)}$	F_i CC	F_i CD
[0;20[	35	35	125	28	28	100
[20;40[	41	76	90	32,8	60,8	72
[40;60[	30	106	49	24	84,8	39,2
[60;100[	12	118	19	9,6	94,4	15,2
[100;140[	5	123	7	4	98,4	5,6
[140;200[	2	125	2	1,6	100	1,6
Total	125	-	-	100%	-	-

4. Représentation graphique des séries statistiques

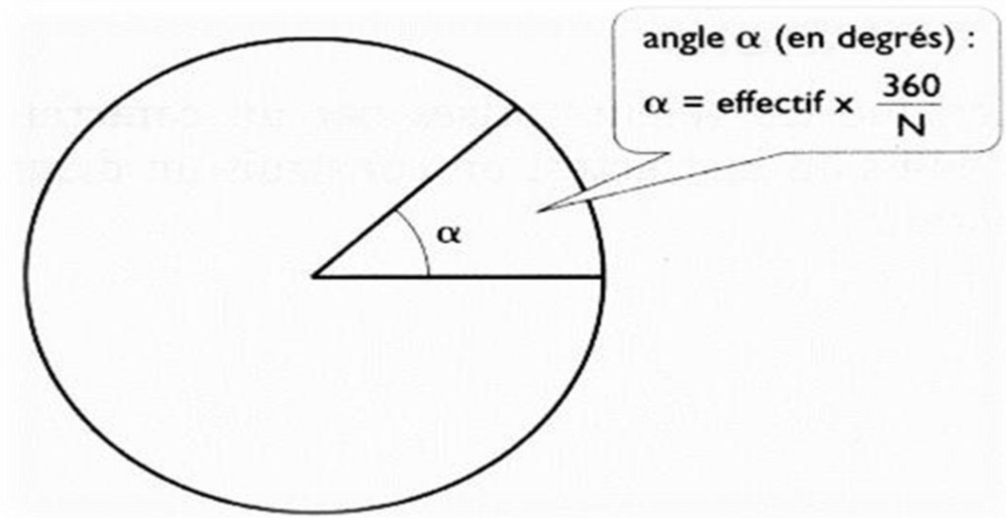
Les graphiques utilisés dépendent essentiellement de la nature de la variable. Ainsi, on peut résumer les différents graphiques aux variables qui leurs sont associées.

Nature de la série statistique	Représentation graphique
Variables qualitatives ordinale et nominale	Diagramme circulaire; diagramme en tuyaux d'orgue (en barres)
Variable quantitative discrète	Diagramme circulaire Diagramme à bâtons
Variable quantitative continue	Histogramme des fréquences; polygone des effectifs ou des fréquences cumulées;

4. Représentation graphique des séries statistiques

4.a. Variable qualitative

Le diagramme circulaire ou « camembert » est un graphique constitué d'un cercle divisé en secteurs dont **les angles au centre sont proportionnels aux effectifs** (ou aux fréquences). Ainsi, les aires des secteurs sont proportionnels aux effectifs. L'angle $\mathbf{a_i}$ est donné en degrés par:



A retenir:

- *Ce diagramme permet de représenter une situation de façon globale, en démontrant l'importance accordée à chacune des parties par rapport au autres.*

- *Il est recommandé de ne pas utiliser cette représentation lorsque la variable possède beaucoup de modalités.*
- *Cette représentation n'a pas de sens pour une variable ordinale (modalités ordonnées).*

I- Les statistiques

4. Représentation graphique des séries statistiques

4.a. Variable qualitative

Illustration : On souhaite étudier les raisons de l'utilisation d'internet au sein d'une population donnée. Les données suivantes vous sont présentées.

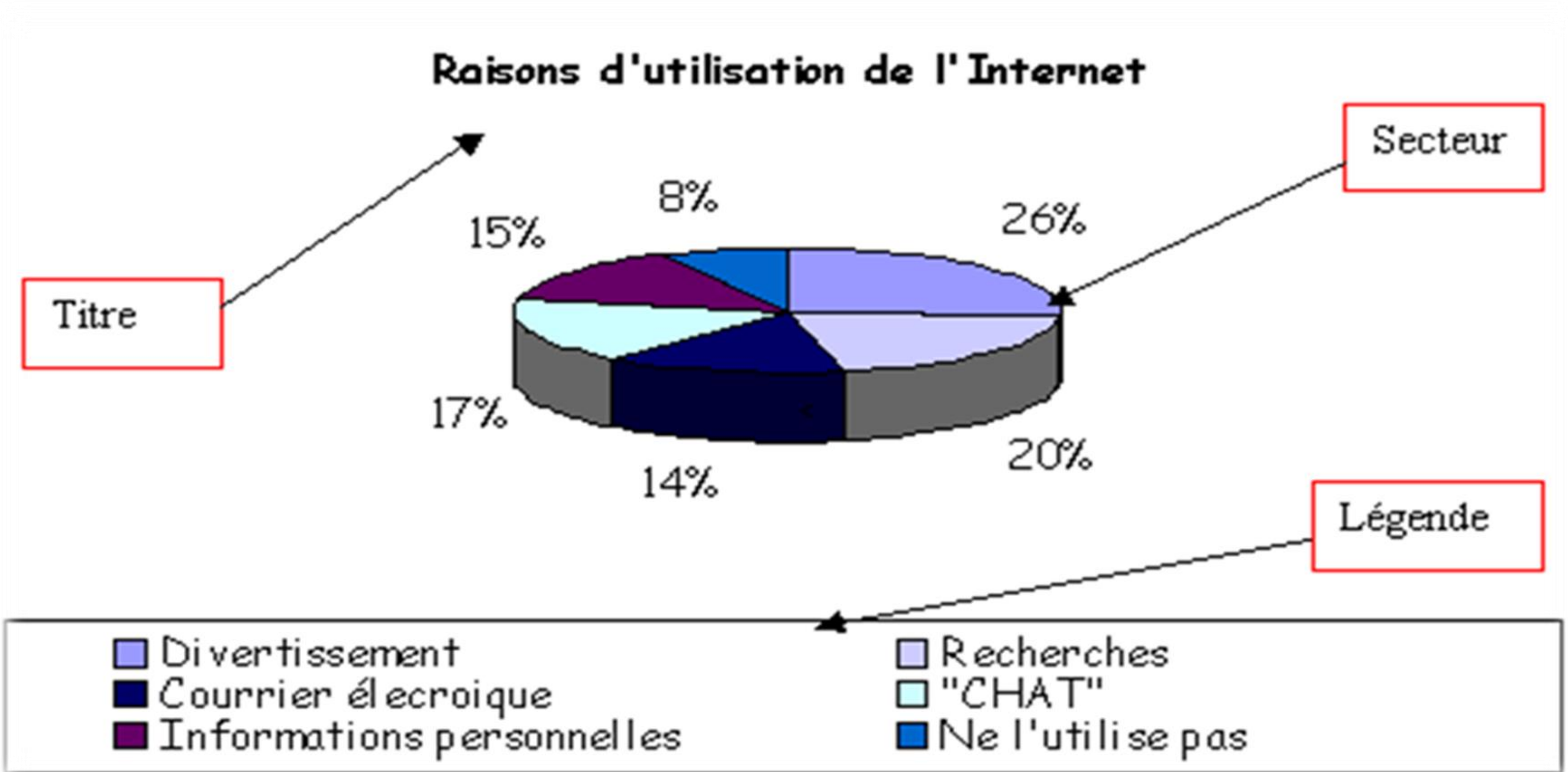
Raisons (x_i)	n_i	f_i	P_i (%)	Degrés a_i (°)
Divertissement	93	93/354	26, 3	94,7
Recherches	72	0, 203	20, 3	73,1
Courrier électronique	51	0,144	14, 4	51,8
« CHAT »	59	0,166	16,7	60,1
Informations personnelles	52	0,146	14,7	52,9
Ne l'utilise pas	27	0,076	7,6	27,4
Total	354	i=1	100%	360°

I- Les statistiques

4. Représentation graphique des séries statistiques

4.a. Variable qualitative

Solution : Représentation graphique du diagramme circulaire



4. Représentation graphique des séries statistiques

4.b.Variable quantitative discrète

Le diagramme en bâtons est un graphique qui à chaque modalité (x_{i}) d'une variable quantitative discrète associe un segment (bâton) dont **la hauteur est proportionnelle l'effectif (ou à la fréquence)**.

Illustration: A la bibliothèque d'une université, les étudiants peuvent emprunter 4 livres par semaine. Cette semaine, 250 étudiants ont emprunté des livres, comme l'indique le tableau suivant:

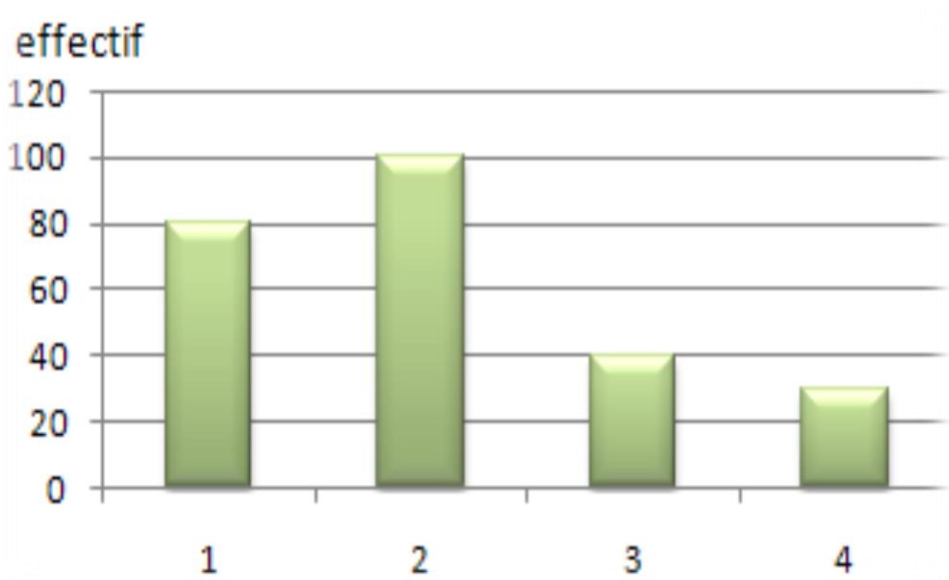
Nbre de livres empruntés (x_{i})	Nbre d'étudiants (n_{i})	f_i	$P_{i\ (\%)}$	Degrés ai (°)
1	80	0,32	32	115,2°
2	100	0,4	40	144°
3	40	0,16	16	57,6°
4	30	0;12	12	43,2°
Total	250	1	100%	360°

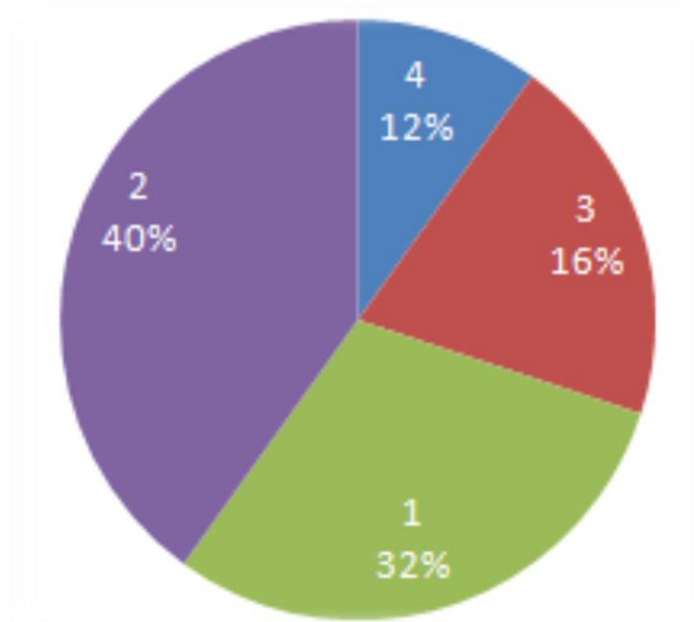
4. Représentation graphique des séries statistiques

4.b.Variable quantitative discrète

Le diagramme en bâtons est un graphique qui à chaque modalité (x_i) d'une variable quantitative discrète associe un segment (bâton) dont **la hauteur est proportionnelle l'effectif (ou à la fréquence)**.

Solution : Représentation graphique des diagrammes en bâtons et circulaire





4. Représentation graphique des séries statistiques

4.b. Variable quantitative continue

L'histogramme

-**Apparence** : L'histogramme est composé de colonnes accolées de hauteurs variables. L'ordonnée (axe Y - vertical) reçoit les valeurs et l'abscisse (axe X - horizontal) les catégories.

- Pour construire un histogramme, on porte les classes en abscisse et sur chacune d'elles pris comme base, on construit un rectangle dont l'aire (et non pas la hauteur) est proportionnelle à l'effectif ou à la fréquence de la classe correspondante.

-On ne laisse pas d'espace entre les colonnes d'un histogramme afin d'indiquer qu'il s'agit de l'étude d'une variable continue.

- Dans le cas d'une variable quantitative continue, on définit la densité d'effectif d_i ou hauteur h_i d'une classe d'effectif n_i et d'amplitude a_i par: $d_i = n_i / a_i$ ou, (dans le cas fréquences f_i / A_i).

A retenir: *Il faut toujours respecter la règle suivante: l'aire d'un rectangle est proportionnelle à l'effectif.*

4. Représentation graphique des séries statistiques

4.b.Variable quantitative continue

Illustration 1: Tracer l'histogramme

Une enquête a été réalisée auprès de 2500 personnes sur la question suivante :
" A quel âge avez-vous trouvé un emploi correspondant à votre qualification "sur l'âge de chaque habitant. Les résultats de l'enquête ont été reportés dans le tableau suivant :

Age	Effectif
[18 ; 22 [	100
[22 ; 26 [	200
[26 ; 30[	400
[30 ; 34 [	1100
[34 ; 38 [	700

4. Représentation graphique des séries statistiques

4.b.Variable quantitative continue

Illustration: Tracer l'histogramme

Une enquête a été réalisée auprès de 2500 personnes sur la question suivante :

" A quel âge avez-vous trouvé un emploi correspondant à votre qualification "sur l'âge de chaque habitant.

Solution:

Nous remarquons bien que les classes présentent des amplitudes constantes. Nous représentons ces résultats par un histogramme.

Protocole à tracer:

-Tracer l'axe des abscisses, celui avec les classes.

-Déterminer une échelle. On la représente par un rectangle dont la longueur correspond à l'amplitude des classes (ici on a choisi 1 cm pour une amplitude de 4).

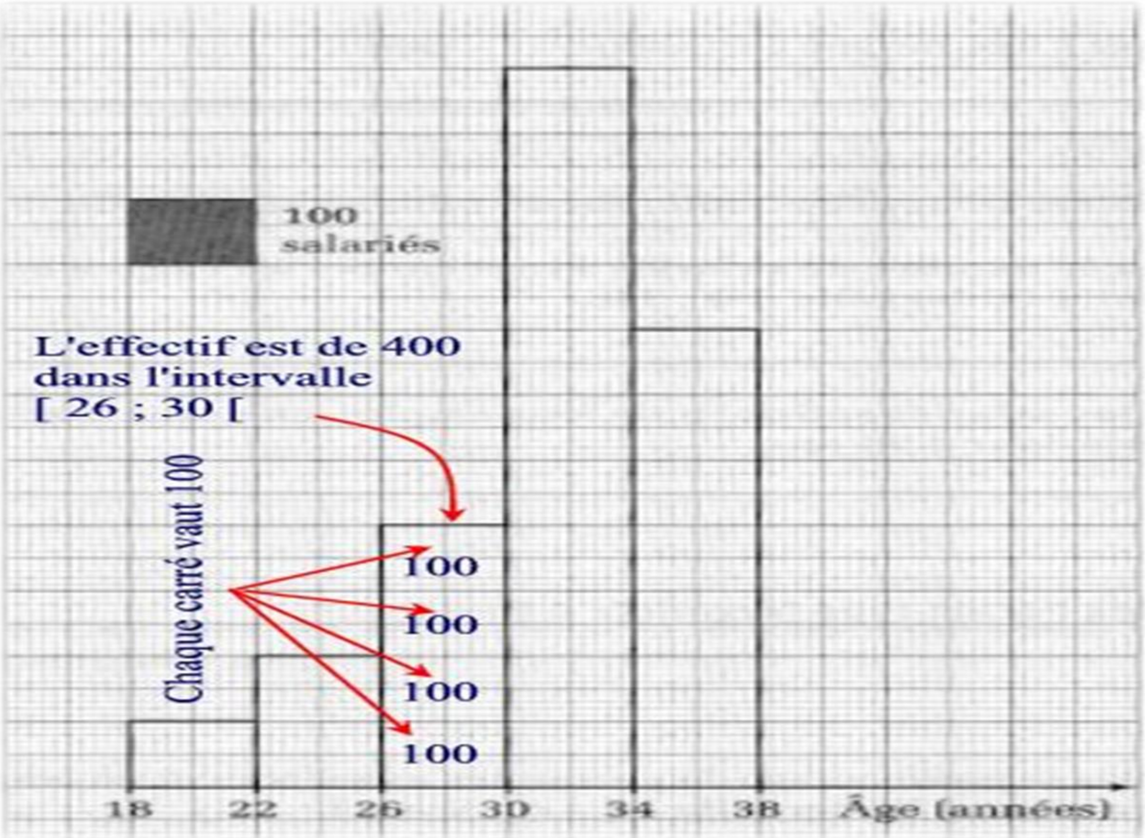
-La hauteur correspond à un effectif (ici on a choisi 1 cm pour 100 personnes).
En partant de cette échelle on construit un rectangle pour chaque classe en respectant l'effectif de chaque classe.

4. Représentation graphique des séries statistiques

4.b.Variable quantitative continue

Illustration: Tracer l'histogramme

Solution:



5. Indicateurs descriptifs (cas univarié)

✓ Indicateurs de tendance centrale

Un paramètre statistique est dit **de position** s'il s'agit d'un nombre clé permettant de préciser où se répartit une certaine fraction des observations.

A retenir: *il a pour objectif dans le cas d'un caractère quantitatif de caractériser l'ordre de grandeur des observations. Nous traitons le cas* ***des paramètres statistiques de tendance centrale*** *; un paramètre statistique est dit de tendance centrale s'il s'agit d'un nombre clé autour duquel les observations sont réparties comme le mode, la moyenne et la médiane.*

5.1. Le mode

Le Mode d'une série statistique est une valeur du caractère dont l'effectif associé est le plus grand.

- La détermination du mode s'applique uniquement aux variables qualitatives et aux variables quantitatives discrètes jamais aux variables quantitatives continues .

- Il est repérable sur le diagramme en bâtons (ou en barres) ou sur tableau des effectifs (ou fréquences).

- Une variable statistique peut avoir plusieurs modes (multimodales).

I- Les statistiques

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

Illustration : variable quantitative discrète

Voici une série statistique : 3 3 3 4 4 5 5 5 6 6 6 7 7 7 7 8 9

1) Déterminer le mode de cette série statistique.
2) Déterminer l'étendue de cette série statistique.

Solution:

1) Le mode de cette série est 7 car la valeur 7 a l'effectif (ici 4) le plus grand.
2) L'étendue pour cette série est de 9-3 =6 (Différence entre les valeurs extrêmes).

5. Indicateurs descriptifs (cas univarié)

✓ Indicateurs de tendance centrale

5.1. Le mode (cas d'une variable quantitative continue)

La classe modale d'une série statistique classée est une classe dont le rapport effectif/ amplitude de la classe est le plus grand. L'amplitude a de la classe i, $[e_i\ e_i+1[$ est la différence entre les deux bornes soit la différence entre l'extrémité supérieure et l'extrémité inférieure, donc: $a_i = e_{i+1} - e_i$.

A retenir: *Dans un histogramme à pas constant, le mode est la classe dont la hauteur est la plus grande.*

Illustration: Voici une série statistique

Déterminez la classe modale de cette série statistique

Valeurs (x_i)	[5;10[	[10;15[	[15;20[	[20;25[	[25;30[
n_i	2	5	7	8	3
(a_i)	10-5 =5	5	5	5	5
n_i / a_i	2/5 = 0,4	1	1,4	1,6	0,6

5. **Indicateurs descriptifs** (cas univarié)

✓ **Indicateurs de tendance centrale**

5.1. Le mode (cas d'une variable quantitative continue)

Solution:

Pour la classe [5 ; 10[, l'amplitude est de 5 , l'effectif de 2 et le rapport de 2/5 = 0,4

Pour la classe [10 ; 15[, l'amplitude est de 5, l'effectif de 5 et le rapport de 5/5 = 1

Pour la classe [15 ; 20[, l'amplitude est de 5, l'effectif de 7 et le rapport de 7/5 = 1,4

Pour la classe [20 ; 25[, l'amplitude est de 5, l'effectif de 8 et le rapport de 8/5 = 1,6

Pour la classe [25 ; 30], l'amplitude est de 5, l'effectif de 3 et le rapport de 3/5 = 0,6

La classe modale est donc la classe [20 ;25] car son rapport est le plus grand (1,6)

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.2. La médiane

La médiane d'une série statistique est un nombre tel que lorsque cette série est classée dans l'ordre croissant, il y a autant de données supérieures à lui que de données inférieures.

C'est aussi le nombre qui partage la série en deux parties ayant le même effectif.

Pour trouver la médiane on divise donc par deux l'effectif total, au besoin on arrondit à l'entier supérieur. La médiane est la valeur correspondant à ce rang.

Méthode de calcul de la médiane : Les N données numériques relatives à un caractère sont rangées dans l'ordre croissant :

-si N est impair, N = 2n + 1 alors la médiane est la donnée de rang n + 1

-si N est pair, N = 2n alors la médiane est la demi-somme des données de rang n et n + 1

A retenir: *S'il y a un nombre impair de données, la médiane est une donnée de la série statistique sinon ce n'est pas le cas.*

5. Indicateurs descriptifs (cas univarié)

✓ Indicateurs de tendance centrale

5.2. La médiane

Illustration 1:

Pour la série : 3, 7, 7, 7, 8, 9, 9, 9, 9, 10, 10, 10, 12, 12, 12, 14, 14, 15, 16, 17.

Il y a 20 valeurs. La médiane est donc la moyenne la 10ème valeur. Comme la dixième valeur vaut 10, la médiane est donc de 10.

A retenir: *Lorsque l'on dispose d'un tableau Valeur/effectif, l'utilisation des effectifs cumulés croissant peut aider à trouver plus vite la médiane .*

Illustration 2: Voici une série définie par un tableau

x_i	46	55	58	63	66	69
n_i	7	4	18	5	17	18
$n_{i\ Cc}$	7	11	29	34	51	69

L'effectif total **n = 69**; on fait 69/ 2= 34,5; la médiane Me sera égale à la 35 e valeur.

Par lecture du tableau, la 34ème valeur est 63, la 35ème valeur est 66 donc la médiane est de 66 .

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.2. La médiane

Illustration 3:

a) Quelle est la médiane des données suivantes : 12 ; 15 ; 5 ; 17 ; 25 ; 22 ; 16.

b) Même question pour les données suivantes : 5 ; 12 ; 17 : 6 ; 1 ; 25.

Solution:

Cas dans lequel **n (l'effectif total) est impair**

a) On range d'abord la série statistique dans l'ordre croissant puis on détermine l'effectif total de la série statistique.

- 5; 12; 15; 16; 17; 22; 25

-Ici, **n = 7;** on fait 7/2= 3,5; la médiane Me sera égale à la 4 eme valeur donc la médiane est de 16.

Cas dans lequel **n (l'effectif total) est pair**

b) On range d'abord la série statistique dans l'ordre croissant: 1; 5; 6; 12 ;17 ;25

L'effectif total **n = 6**; on fait 6/2= 3; la médiane Me est la demi-somme des 3e et 4e termes c'est-à-dire (6 + 12) : 2 = 9.

5. **Indicateurs descriptifs** (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La moyenne est une valeur qui se trouve au milieu de toutes les autres. C'est un indicateur de tendance centrale qui permet d'appréhender une population de manière globale.

A retenir: *La moyenne, comme les autres éléments du calcul statistique,* ***ne s'utilise que pour des caractères quantitatifs****. Par exemple, on pourrait essayer de calculer la couleur moyenne des voitures dans une entreprise, mais cela ne serait d'aucune utilité.*

❑Moyenne arithmétique simple

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{N}$$

Illustration:

Dans une entreprise, on s'intéresse à la production mensuelle d'une usine sur une période de 6mois.

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Production en unités	10000	50000	20000	30000	30000	40000

5. Indicateurs descriptifs (cas univarié)

✓ Indicateurs de tendance centrale

5.3. Moyenne

La Moyenne arithmétique simple

Illustration:

Dans une entreprise, on s'intéresse à la production mensuelle d'une usine sur une période de 6mois.

Mois	Janvier	Février	Mars	Avril	Mai	Juin
Production en unités	10000	50000	20000	30000	30000	40000

Solution:

On cherche à calculer une moyenne de production mensuelle en sachant que chaque mois a la même importance.

Moyenne = ∑ (10000 + 50000 + 20000+ 30000+ 30000+ 40000) / 6 = 30000

5. **Indicateurs descriptifs** (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique simple

Illustration: moyenne d'une série statistique

Siloé a eu les notes suivantes en mathématiques : 12 ; 11 ; 8 ; 7 ; 13.

a) Elle calcule sa moyenne et trouve 13,5. Sans faire de calcul comment peut-on être sûr qu'elle s'est trompée.

b) Calculer sa moyenne.

Solution:

a) **La moyenne de données statistiques est toujours comprise entre la valeur minimale et la valeur maximale de cette série**. Or 13,5 est supérieur à 13 qui est la plus grande valeur des données statistiques, ça ne peut donc pas être la moyenne.

b) (12 + 11 + 8 + 7 + 13)/5 = 10,2.

5. **Indicateurs descriptifs** (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée

Une moyenne pondérée est une moyenne dont certaines des valeurs sont affectées d'un poids. Pour calculer une moyenne pondérée, on effectue le calcul suivant :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N}$$

Illustration 1: cas d'une variable quantitative discrète

Lucas a eu 7 notes en français au cours du 1er trimestre :

-trois notes d'interrogation surprise : 14. Donc, 12 et 7 qui ont 1 de coefficient ;

-deux notes de devoir rédigé à la maison : 15 et 13 de coefficient 2 ;

- trois notes de contrôle : 12 ; 9 et 11 de coefficient 3. Quelle est sa moyenne du 1er trimestre ?

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N}$$

Solution:

$\bar{x}$ = 14 + (12 X 1) + (7 X 1) + (15 X 2) + (13 X 2) + (12 X 3) + (9 X 3) + (11 X 3) / (1 + 1 + 2 + 2 + 3 + 3 + 3)

= (14+12+7+30+26+36+27+33) / 15

= 185 / 15

= 12, 33

La moyenne est donc 12, 33

5. **Indicateurs descriptifs** (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée

Illustration 2: cas d'une variable quantitative continue

Une enquête sur les âges des enfants d'un centre de vacances a donné les résultats suivants :

Classes d'âges	Nombre d'enfants
[4; 6[	20
[8; 6[	30
[8; 10[	80
[10; 12[	70
[12; 14[	30
[14; 16[	10

Déterminez l'âge moyen de cette population

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée

Illustration 1: cas d'une variable quantitative continue

Une enquête sur les âges des enfants d'un centre de vacances a donné les résultats suivants :

Déterminez l'âge moyen de cette population

Solution:

Ici, on calcule le centre de la classe qui est le milieu de la classe. Par conséquent,

$$C_{i=} e_{i+1} + e_{i\,/\,2}$$

Par exemple, le centre de la classe [4; 6[est : (4+6) / 2 = 10/2 = 5

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée (Solution: cas d'une variable quantitative continue)

Classe d'âges	Nombre d'enfants(n_i)	Centre de classe (c_i)	$n_i\ x_i$
[4; 6[	20	(4+6) / 2= 5	(5 X 20) = 100
[6; 8[	30	7	210
[8; 10[	80	9	720
[10; 12[	70	11	770
[12; 14[	30	13	390
[14; 16[	10	15	150
Total	240	-	2350

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée (Solution: cas d'une variable quantitative continue)

Classe d'âges	Nombre d'enfants(n_i)	Centre de classe (x_i)	n_i x_i
[4; 6[	20	(4+6) / 2= 5	(5 X 20) = 100
[6; 8[	30	7	210
[8; 10[	80	9	720
[10; 12[	70	11	770
[12; 14[	30	13	390
[14; 16[	10	15	150
Total	240	-	2350

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de tendance centrale**

5.3. Moyenne

La Moyenne arithmétique pondérée

Solution:

L'âge moyen de cette population est :

$$\overline{x} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_p x_p}{N}$$

$\overline{x}$ = 2350 / 240 = 9,75 ~- 10 ans

L'âge moyen de cette population est de 10 ans.

5. Indicateurs descriptifs (cas univarié)

✓Indicateurs de dispersion

Un paramètre statistique est dit de dispersion s'il s'agit d'un nombre clé résumant la plus ou moins grande disparité des observations et leur variabilité de part et d'autre de la tendance centrale. Les indicateurs de dispersion les plus courantes sont: l'étendue, l'écart interquartile, l'écart- type, la variance, le coefficient de variation.

5.4. Quartiles

Les quartiles partagent la population ou l'échantillon en quatre groupes comprenant chacun 25% des observations. Ils sont au nombre de trois, ils se notent : Q_1, Q_2, et Q_3.

- Q_1 est le quantile d'ordre 0,25; au moins 25% des observations sont inférieures ou égales à Q_1 et au moins 75% supérieures ou égales à Q_1.

- Q_2 est le quantile d'ordre 0,50 ; au moins 50% des observations sont inférieures ou égales à Q_2 et au moins 50% supérieures ou égales à Q_2 (qui est égal à la médiane).

- Q_3 est le quantile d'ordre 0,75; au moins 75% des observations sont inférieures ou égales à Q_3 et au moins 25% supérieures ou égales à Q_3.

A retenir: *Le quartile est une valeur qui ne fait généralement pas partie de la distribution étudiée. Par contre, chacun des quatre sous-ensembles formés par les quartiles s'appelle quart.*

I- Les statistiques

5. Indicateurs descriptifs (cas univarié)

✓Indicateurs de dispersion

5.4. Quartiles

Illustration : Déterminez la valeur des trois quartiles Q1, Q2, Q3 de la distribution suivante:

60,32,87,98,56,75,35,68,86,90,75,59,61,84,64,48.

Solution:

1) Ranger les données en ordre croissant:

32,35,48,56,59,60,61,64, 68,75,75,84,86,87,90,98.

2) Déterminez la médiane

On sait que **n= 16, 16/2=8** ; l'effectif total N est pair donc la médiane est la demi-somme des données de rang n et n + 1 c'est-à-dire que Me se trouve entre le 8^e et le 9^e terme.

Me = (64 + 68) / 2 alors **Me = Q2 = 66**

32,35,48,56,59,60,61,64 66 68,75,75,84,86,87,90,98.

5. Indicateurs descriptifs (cas univarié)

✓Indicateurs de dispersion

5.4. Quartiles

Illustration : Déterminez la valeur des trois quartiles Q1, Q2, Q3 de la distribution suivante:

60,32,87,98,56,75,35,68,86,90,75,59,61,84,64,48.

Solution:

3) On sait que Q1, est le quantile d'ordre 0,25. En sachant que n= 16; on fait **(16 X 0,25) = 4**

32,35,48,56,59,60,61,64, **68,75,75,84,86,87,90,98.**

Ici on considère uniquement les 50% de la série statistique ci en jaune, et on sait que (16 X 0,25) = 4 donc N est pair alors la médiane est la demi-somme des données de rang n et n + 1 c'est-à-dire que Q1 se trouve entre le 4e et le 5e terme de ce sous-échantillon.

Q1 = (56 + 59) / 2 alors **Q1 = 57,5**

32,35,48,56,59,60,61,64 66 68,75,75,84,86,87,90,98. (Médiane)

32,35,48,56 57,5 59,60,61,64, **68,75,75,84,86,87,90,98.** (Q1)

5. Indicateurs descriptifs (cas univarié)

✓Indicateurs de dispersion

5.4. Quartiles

Illustration : Déterminez la valeur des trois quartiles Q1, Q2, Q3 de la distribution suivante:

60,32,87,98,56,75,35,68,86,90,75,59,61,84,64,48.

Solution:

4) On sait que Q3, est le quantile d'ordre 0,75. En sachant que n= 16; on fait **(16 X 0,75) = 12**

32,35,48,56,59,60,61,64, 68,75,75,84,86,87,90,98.

Ici on considère uniquement la deuxième moitié des 50% de la série statistique ci en jaune, et on sait que (16 X 0,75) = 12 donc N est pair alors la médiane est la demi-somme des données de rang n et n + 1 c'est-à-dire que Q3 se trouve entre le 12e et le 13e terme de ce sous-échantillon.

Q1 = (84 + 86) / 2 alors **Q3 = 85**

32,35,48,56,59,60,61,64 66 68,75,75,84,86,87,90,98. (Médiane)

32,35,48,56 57,5 59,60,61,64, **68,75,75,84,86,87,90,98.** (Q1)

32,35,48,56,59,60,61,64, 68,75,75,84 85 86,87,90,98.

5. Indicateurs descriptifs (cas univarié)

✓Indicateurs de dispersion

5.5. L'écart interquartile

En établissant la valeur des quartiles, il est possible de discuter de la dispersion des données. Graphiquement, la boîte d'un diagramme de quartiles illustre l'étendue interquartile. Concrètement, elle représente la dispersion du quart précédant la médiane et celui la succédant. Donc, cette boîte représente généralement 50 % des données. Pour obtenir la valeur de cette étendue, on effectue la soustraction suivante :

Ecart interquartile = 3^e quartile _ 1 er quartile = Q3 - Q1

A retenir: *Cette mesure de dispersion autour de la médiane élimine les 25% de valeurs les plus petites et les 25% les plus grandes, et permet ainsi d'écarter les valeurs extrêmes qui peuvent perturber les conclusions.*

Illustration : Déterminez l'écart interquartile de la série statistique précédente sachant que:
Q1= 57,5 ; Q2= 66; Q3 = 85

Solution: Ecart interquartile = Q3 – Q1 = 85 – 57,5 = 27,5

En d'autres mots, 50% des données sont regroupées dans un intervalle d'une longueur de **27,5** unités.

5. Indicateurs descriptifs (cas univarié)

✓ Indicateurs de dispersion

5.6. variance; écart-type et coefficient de variation

La variance est la moyenne des carrés des différences entre les observations x_i et leur moyenne $\bar{x}$. La variance correspond au carré de l'écart-type (noté σ), et les notations pour la désigner sont multiples. On trouvera par exemple:

V(X); Var(X); σ^2(X), ou S^2 pour l 'échantillon.

Elle peut être considérée comme une mesure servant à caractériser la dispersion d'une distribution statistique.

La formule est la suivante:

$$\begin{aligned}\mathrm{var}(X) = s_X^2 &= \frac{1}{n}\sum_{i=1}^{r} n_i(x_i - \overline{x})^2 \\ &= [\frac{1}{n}\sum_{i=1}^{r} n_i(x_i)^2] - (\overline{x})^2.\end{aligned}$$

I- Les statistiques

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de dispersion**

5.6. variance; écart-type et coefficient de variation

A retenir:

-La valeur de la variance est toujours positive puisque son unité est le carré de l'unité de la variable.

-La variance S^2 est nulle si et seulement si toutes les observations ont la même valeur (aucune dispersion).

- Si la valeur de la variance est faible, les observations sont proches les unes des autres de la moyenne. Si au contraire, la valeur de la variance est élevée, les observations sont très dispersées autour de la moyenne, plus elles sont hétérogènes et plus la variance S^2 s'accroit..

-L'unité dans laquelle s'exprime la variance vaut le carré de l'unité utilisée pour les valeurs observées. Ainsi, par exemple, une série de poids exprimés en kilos possède une variance qui, elle, doit s'interpréter en "kilos-carré".

-Tout comme la moyenne $\bar{x}$, la variance est très sensible aux valeurs extrêmes.

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de dispersion**

5.6. variance; écart-type et coefficient de variation

- L'écart-type de X, est la racine carrée positive de la variance, et sera donc notée $S(x)$ ou δ_X :

$$s(X) = \sqrt{\mathrm{var}(X)} \qquad s(X) = \sqrt{s^2(X)}$$

A retenir: *Dans l'absolu, la valeur de l'écart type ne donne aucune observation. Un écart type s'interprète toujours de façon relative, c'est-à-dire en le comparant à d'autres valeurs.*

- Le coefficient de variation est le rapport de l'écart-type à la moyenne. Il est noté CV(x) et défini par : $CV(x)= \delta_X / \bar{x}$. Ce coefficient s'exprime en pourcentage de la moyenne.

5. Indicateurs descriptifs (cas univarié)

✓ **Indicateurs de dispersion**

5.6. variance; écart-type et coefficient de variation

❑ Le coefficient de variation

A retenir: *Plus la valeur du coefficient de variation est élevée, plus la dispersion autour de la moyenne est grande. Il est généralement exprimé en pourcentage. Sans unité, il permet la comparaison de distributions de valeurs dont les échelles de mesure ne sont pas comparables. Il permet de mesurer l'homogénéité des données.*

Illustration (variable quantitative discrète et continue) ➡ Cf.TD

I- Les statistiques (Travaux dirigés 1)

Exercice 1: Types de variable, variable ou caractère

Mots-clefs: Variable quantitative discrète ; variable quantitative continue, caractère qualitatif ordinal , caractère qualitatif nominal

Énoncé: Quelle est la nature des caractères ci-dessous ?

1- Nombre d'actions vendues chaque jour à la bourse

2- Rémunérations des enseignants d'un lycée

3- Indicateur du moral des ménages

4- Écart de rémunération entre hommes et femmes

5- Les pays de l'Union européenne

6- Les niveaux de formation des salariés

7- Les formes de contrat de travail

8-Taux de croissance du PIB

9- Prix à la consommation

10-Solde commercial

11- Nombre de personnes par ménages

I- Les statistiques (Travaux dirigés 1)

Corrigé1: Types de variable, variable ou caractère

Mots-clefs: Variable quantitative discrète ; variable quantitative continue, caractère qualitatif ordinal , caractère qualitatif nominal

Énoncé: Quelle est la nature des caractères ci-dessous ?

1- Nombre d'actions vendues chaque jour à la bourse – variable quantitative discrète
2- Rémunérations des enseignants d'un lycée – Variable quantitative continue
3- Indicateur du moral des ménages – caractère qualitatif ordinal
4- Écart de rémunération entre hommes et femmes – Variable quantitative continue
5- Les pays de l'Union européenne – Caractère qualitatif nominal
6- Les niveaux de formation des salariés – caractère qualitatif ordinal
7- Les formes de contrat de travail – caractère qualitatif nominal
8-Taux de croissance du PIB – variable quantitative continue
9- Prix à la consommation - variable quantitative continue
10-Solde commercial - variable quantitative continue
11- Nombre de personnes par ménages - variable quantitative discrète

Une variable statistique ou caractère est une propriété étudiée sur chaque individu de la population. Lorsque les modalités de cette variable prennent des valeurs (possibilités de réponses) numériques : on parle de <u>variable quantitative</u>. Et lorsque, les modalités ne sont pas des nombres : on parle <u>de variable qualitative.</u>

I- Les statistiques (Travaux dirigés 1)

Exercice 2: Variance et écart-type

Un apprenti chercheur s'intéresse à 3 sous-groupes d'élèves du collège :

– G1 : les 3 élèves de ce sous-groupe ont le même âge (14, 14, 14 ans) ;

– G2 : les 3 élèves de ce sous-groupe ont 14 ans plus ou moins 1 an (14, 13, 15 ans) ;

– G3 : les 3 élèves de ce sous-groupe ont 14 ans plus ou moins 2 ans (14, 12, 16 ans).
Les 3 sous-groupes G1, G2 et G3 ont la même moyenne = 14 ans.

1) a- Quelle est la population étudiée?

b- Quel est le caractère étudié? Sa nature?

2) Déterminez la variance et l'écart-type pour chaque sous-groupe? Que peut-on conclure?

I- Les statistiques (Travaux dirigés 1)

Corrigé 2: Variance et écart-type

Un apprenti chercheur s'intéresse à 3 sous-groupes d'élèves du collège :

- G1 : les 3 élèves de ce sous-groupe ont le même âge (14, 14, 14 ans) ;
- G2 : les 3 élèves de ce sous-groupe ont 14 ans plus ou moins 1 an (14, 13, 15 ans) ;
- G3 : les 3 élèves de ce sous-groupe ont 14 ans plus ou moins 2 ans (14, 12, 16 ans).

Les 3 sous-groupes G1, G2 et G3 ont la même moyenne ($\bar{x}=14$)

1) a- La population étudiée est le sous-groupe d'élèves.

b- Le caractère étudié est **l'âge**. C'est une **variable quantitative discrète** parce que ce caractère ne peut prendre que des valeurs entières .

2) a- Calcul de la variance pour chaque sous-groupe

On sait que: $\mathbf{Var(X) = \sum (x - \bar{x})^2 / N}$ donc:

- $Var(G1) = (14-14)^2 + (14-14)^2 + (14-14)^2 / 3 = 0$
- $Var(G2) = (14-14)^2 + (13-14)^2 + (15-14)^2 / 3 = 0{,}67$
- $Var(G3) = (14-14)^2 + (12-14)^2 + (16-14)^2 / 3 = 2{,}66$

I- Les statistiques (Travaux dirigés 1)

Corrigé 2: Variance et écart-type

Un apprenti chercheur s'intéresse à 3 sous-groupes d'élèves du collège :

- G1 : les 3 élèves de ce sous-groupe ont le même âge (14, 14, 14 ans) ;
- G2 : les 3 élèves de ce sous-groupe ont 14 ans plus ou moins 1 an (14, 13, 15 ans) ;
- G3 : les 3 élèves de ce sous-groupe ont 14 ans plus ou moins 2 ans (14, 12, 16 ans).

Les 3 sous-groupes G1, G2 et G3 ont la même moyenne ($\bar{x}$=14)

2) b- Calcul de l'écart-type pour chaque sous-groupe

On sait que l'écart-type est la racine carré de la variance donc:

-**$\delta X = \sqrt{Var(X)}$** ; $\delta G1 = \sqrt{Var(G1)} = \sqrt{(0)} = 0$

-**$\delta X = \sqrt{Var(X)}$** ; $\delta G1 = \sqrt{Var(G2)} = \sqrt{(0{,}67)} = 0{,}82$

-**$\delta X = \sqrt{Var(X)}$** ; $\delta G1 = \sqrt{Var(G3)} = \sqrt{(2{,}66)} = 1{,}63$

2) c- Lorsque les valeurs d'une variable sont très centrées autour de la moyenne, la variance est faible comme c'est le cas pour Var(G2) = 0,67. La dispersion est nulle pour Var(G1)= 0 puisque toutes les valeurs sont égales à la moyenne $\bar{x}$ =14. A contrario, plus la variance s'éloigne de 0, plus les valeurs sont dispersées et s'éloignent de la moyenne comme dans le cas de Var(G3) = 2,66; Même conclusion pour les résultats de l'écart-type.

I- Les statistiques (Travaux dirigés 1)

Exercice 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

Dans un magasin de pièces détachées, sur un lot de 100 pièces vendues en une année, les prix s'échelonnent entre 200 £ et 800£ selon la répartition suivante:

Prix en £	[200;300[	[300;450[	[450;550[	[550;600[	[600;800[
Nombre de pièces vendues	15	35	25	10	15

1) Calculez les effectifs cumulés croissants et les fréquences cumulées croissantes de cette série statistique?

2)Déterminez le mode Mo par calcul et dites si on peut le déterminer graphiquement? Si oui, dans quel type de graphique ?

3) Calculez la médiane Me de cette série statistique en explicitant vos calculs. Donner son interprétation?

4) Déterminez Q1 et Q3 et déduire l'écart interquartile?

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

Dans un magasin de pièces détachées, sur un lot de 100 pièces vendues en une année, les prix s'échelonnent entre 200 £ et 800£ selon la répartition suivante:

1)Tableau de répartition des effectifs cumulés et fréquences cumulées

Prix £	Pièces (ni)	Ni cc	fi	Fi cc	hi =ni /ai ou di =ni /ai	
[200;300[	15	15	0,15	0,15	100	0,15
[300;450[	35	50	0,35	0,50	150	0,23
[450;550[	25	75	0,25	0,75	100	0,25
[550;600[	10	85	0,1	0,85	100	0,15
[600;800[	15	100	0,15	1	200	0,075
Total	N = 100	-	1	-		

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

2) La classe modale d'une série statistique de variable quantitative continue est la classe ayant la hauteur la plus élevée. Ici la classe modale est [450;550[parce qu'ayant la plus grande hauteur soit hi= 0,25. on peut le déterminer graphiquement sur un histogramme. D'ailleurs, sur un histogramme, le mode est la modalité ayant la plus grande hauteur.

Pour trouver le chiffre exact du mode dans l'intervalle [450;550[, on utilise la méthode d'interpolation linéaire dont la formule est la suivante:

$$M_o = e_{i-1} + \frac{a_i\left(n_i - n_{i-1}\right)}{\left(n_i - n_{i+1}\right)\left(n_i - n_{i-1}\right)}$$

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

2) Pour trouver le chiffre exact du mode dans l'intervalle [450;550[, on utilise la méthode d'interpolation linéaire dont la formule est la suivante:

$$M_o = e_{i-1} + \frac{a_i\,(n_i - n_{i-1})}{(n_i - n_{i+1})\,(n_i - n_{i-1})}$$

Avec:

$e_{i\,-1}$ *= 450 correspond à la borne inférieure de la classe modale* [450;550[

a_i = 100 correspond à l'amplitude de la classe modale soit 550 – 450 = 100

h_i = 0,25 correspond à la hauteur de la classe modale en rouge dans le tableau précédent

h_{i-1} = 0,23 est la hauteur de la classe modale supérieure n_{+1} en rouge dans le tableau

$h_{i}+_{1}$ = 0,2 correspond à la hauteur de la classe modale inférieure n_{-1} de la classe modale en rouge dans le tableau

*Il faut retenir que pour **<u>calculer le mode</u>**, on peut aussi bien utiliser <u>les effectifs cumulés, les fréquences cumulées, ou la hauteur</u> (parfois on parle de densité pour désigner la hauteur) selon les cas. Ici, on a privilégié la hauteur. La formule reste la même, on remplace juste les « n_i par h_i ».*

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

2) Pour trouver le chiffre exact du mode dans l'intervalle [450;550[on utilise la méthode d'interpolation linéaire dont la formule est la suivante:

$$M_o = e_{i-1} + \frac{a_i\,(n_i - n_{i-1})}{(n_i - n_{i+1})\,(n_i - n_{i-1})}$$

<u>Application:</u>

$$\text{Mo} = 450 + \frac{100(0{,}25 - 0{,}23)}{((0{,}25 - 0{,}23) + (0{,}25 - 0{,}2))}$$

= 450 +100 [(0,02 / (0,02+ 0,05))]
= 450 +100 (0,02 / 0,07)
= 450 +100 (0,285)
= 450 +28,5
= 478,5
Le mode de la classe modale [450;550[est : **<u>Mo= 478,5</u>**

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

3) Calculez la médiane Me de cette série statistique en explicitant vos calculs. Donner son interprétation?

Dans ce cas, comme la variable étudiée est quantitative continue, on utilise la méthode d'interpolation linéaire.

- **Q_2** est **le quantile d'ordre 0,50** ; au moins 50% des observations sont inférieures ou égales à **Q_2** et au moins 50% supérieures ou égales à **Q_2** (qui est égal à la médiane).

On sait que **N= 100** donc on fait **N/2** soit **100/2 =50**
Dans le tableau de répartition des effectifs cumulés, on constate que **50** appartient à **la classe médiane** **[300;450[**.

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

3) On sait que **N= 100** donc on fait **N/2** soit **100/2 =50**
Dans le tableau de répartition des effectifs cumulés, on constate que **50** appartient à **<u>la classe médiane</u> [300;450[**. Dans ce cas, comme la variable étudiée est quantitative continue, on utilise la méthode d'interpolation linéaire. On applique donc la formule suivante:

$$M_e = e_{i-1} + a_i \frac{\frac{n}{2} - N_{i-1}}{n_i}$$

Avec:
e_{i-1} = 300 correspond à la borne inférieure de la classe médiane [300;450[
a_i = 150 correspond à l'amplitude de la classe médiane soit 450 – 300 = 150
n_i = 35 correspond à l'effectif de la classe médiane
N_{i-1} = 15 est l'effectif cumulé de la classe avant la classe médiane
n = 100 est l'effectif total de la population étudiée

*Il faut retenir que pour **<u>calculer la médiane</u>**, on peut aussi bien utiliser <u>les fréquences cumulées</u> selon les cas. Ici, on a privilégié les effectifs cumulés. La formule reste la même, on remplace juste les « n_i par f_i ».*

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

3) On sait que **N= 100** donc on fait **N/2** soit **100/2 =50**
Dans le tableau de répartition des effectifs cumulés, on constate que **50** appartient à **la classe médiane** **[300;450**[, on applique la formule suivante;
Dans ce cas, comme la variable étudiée est quantitative continue, on utilise la méthode d'interpolation linéaire:

$$M_e = e_{i-1} + a_i \frac{\frac{n}{2} - N_{i-1}}{n_i}$$

Application:

Me= $300 + 150\left(\frac{\frac{100}{2} - 15}{35}\right)$
= 300 +150 ((50 -15) /35)
= 300 +150 (35/35)
= 300 +150 (1)
= 450
La médiane de la classe médiane [300;450[est : **Me= Q_2 = 450**
En guise d'interprétation, on dira que 50% des classes sont inférieures à Me= 450 et 50% sont supérieures à 450.

I- Les statistiques (Travaux dirigés 1)

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

4) Déterminez Q1 et Q3 et déduire l'écart interquartile?

- **Q_1** est le **quantile d'ordre 0,25**; au moins 25% des observations sont inférieures ou égales à

Q_1 et au moins 75% supérieures ou égales à **Q_1**. On sait que **N= 100**, on fait **100 X 0,25** = soit **25**. Dans le tableau de répartition des effectifs cumulés, on constate que **25 ne figure pas dans les effectifs cumulés croissants mais la valeur la plus proche de 25 est 50 qui** appartient à **la classe [300;450[**. Dans ce cas, on utilise la méthode d'interpolation linéaire. On applique donc la formule suivante:

$$Q_h = e_{h-1} + a_h \frac{\frac{hn}{4} - N_{h-1}}{n_h}$$

Avec:

e_{h-1}= *300 correspond à la borne inférieure de la classe contenant le quartile* [300;450[

a_h = 150 correspond à l'amplitude de la classe contenant le quartile 450 -300 = 150

N_{h-1} = 15 est l'effectif cumulé de la classe précédant celle contenant le quartile

n_h = 35 l'effectif de la classe contenant celle du quartile

n= h_h =100 est l'effectif total de la population

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

4) Déterminez Q1 et Q3 et déduire l'écart interquartile?

Q_1 et au moins 75% supérieures ou égales à **Q_1** On sait que **N= 100**, on fait **100 X 0,25** = soit **25.** Dans le tableau de répartition des effectifs cumulés, on constate que **25 ne figure pas dans les effectifs cumulés croissants mais la valeur la plus proche de 25 est 50 qui** appartient à **la classe [300;450**[. Dans ce cas, on utilise la méthode d'interpolation linéaire. On applique donc la formule suivante:

$$Q_h = e_{h-1} + a_h \frac{\frac{hn}{4} - N_{h-1}}{n_h}$$

Application:

$$\mathbf{Q_1} = 300 + 150 \left(\frac{\frac{100}{4} - 15}{35} \right)$$

= 300 +150 ((25 -15) /35)

= 300 +150 (10/35)

= 342, 857 ~ 342,86

Le quartile **Q_1** est : **<u>Q_1 = 342,86</u>**

En guise d'interprétation, on dira que 25% des classes sont inférieures à **Q_1** = **<u>342,86</u>** et 75 % sont supérieures à 342,86.

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

4) Déterminez Q1 et Q3 et déduire l'écart interquartile?

- Q_3 est le **quantile d'ordre 0,75**; au moins 75% des observations sont inférieures ou égales à **Q_3** et au moins 25% supérieures ou égales à **Q_3**.

On sait que **N= 100**, on fait **100 X 0,75** = soit **75**. Dans le tableau de répartition des effectifs cumulés, on constate que **75** appartient à **la classe [450;550[**. Dans ce cas, on utilise la méthode d'interpolation linéaire. On applique donc la formule suivante:

$$Q_h = e_{h-1} + a_h \frac{\frac{hn}{4} - N_{h-1}}{n_h}$$

Avec:

e_{h-1}= *450 correspond à la borne inférieure de la classe contenant le quartile* [450;550[

a_h = 100 correspond à l'amplitude de la classe contenant le quartile 550 - 450 = 100

N_{h-1} = 50 est l'effectif cumulé de la classe précédant celle contenant le quartile

n_h = 25 est l'effectif de la classe contenant celle du quartile

n= h_h =100 est l'effectif total de la population

I- Les statistiques (Travaux dirigés 1)

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

4) Déterminez Q1 et Q3 et déduire l'écart interquartile?

Q_3 est le **quantile d'ordre 0,75**; au moins 75% des observations sont inférieures ou égales à **Q_3** et au moins 25% supérieures ou égales à **Q_3**. On sait que **N= 100**, on fait **100 X 0,75** = soit **75**. Dans le tableau de répartition des effectifs cumulés, on constate que **75** appartient à **la classe [450;550[**. Dans ce cas, on utilise la méthode d'interpolation linéaire. On applique donc la formule suivante:

$$Q_h = e_{h-1} + a_h \frac{\frac{hn}{4} - N_{h-1}}{n_h}$$

Application:

Q1 = 450 + 100 ($\frac{\frac{100}{4} - 50}{25}$)

= 450+100 ((75 -50) /25)

= 450 +100 (25/25)

= 550

Le quartile Q3 est : <u>**Q3 = 550**</u>

En guise d'interprétation, on dira que 75% des classes sont inférieures à **Q3 = 550** et 25 % sont supérieures à 550.

Corrigé 3: Effectifs cumulés, fréquences cumulées, mode, médiane, quartiles

4) Déterminez Q1 et Q3 et déduire l'écart interquartile?

On sait que : Q1 =342, 86 ; Q2= M_e = 450 et Q3= 550;
Comme on a une variable quantitative continue, l'écart interquartile s'applique en suivant la formule suivante:

Ecart interquartile = Q3 – Q1/ Q2

Application:

Ecart interquartile = $\left(\frac{550 - 342{,}86}{450}\right)$
= 207,14 / 450
= 0,46

Q3 – Q1/Q2 = 0,46; en guise d'interprétation, on dira que la série est homogène car,
Q3 – Q1/Q2 = 0,46<1,5.

I- Les statistiques (Travaux dirigés 1)

Exercice 4: Statistiques à deux variables

Soit les tableaux statistiques suivants:

1)Tableau de distribution de la variable Emploi

2) Tableau de distribution de la variable Région d'habitation

Emploi	Effectif
Salarié d'une entreprise	68
Recherche d'emploi	46
Indépendant	56
Dirigeant d'entreprise	40
Total	210

Région	Effectif
Paris	104
Province	44
Europe hors France	62
Total	210

I- Les statistiques (Travaux dirigés 1)

Exercice 4: Statistiques à deux variables

3) On vous donne le tableau de contingence des deux tableaux:

Déterminez les types de variables étudiées? Leur nature? Et les différentes modalités pour chacune des variables?

Région/ Emploi	Paris	Province	Europe hors France	Total N_i
Salarié d'entreprise	34	10	24	68
Recherche d'emploi	38	3	5	46
Indépendant	23	31	2	56
Dirigeant d'entreprise	9	0	31	40
Total N_j	104	44	62	N=210

I- Les statistiques (Travaux dirigés 1)

Exercice 4: Statistiques à deux variables

4) a-Déterminez à partir du tableau le nombre total de la population étudiée résidant en province?

b- Le nombre de personnes résidant à Paris qui ont le statut d' indépendant?

5) Calculez le tableau des fréquences « profils-lignes (%) » et « profils-colonnes (%) »?

Région/ Emploi	Paris	Province	Europe hors France	Total N_i
Salarié d'entreprise	34	10	24	68
Recherche d'emploi	38	3	5	46
Indépendant	23	31	2	56
Dirigeant d'entreprise	9	0	31	40
Total N_j	104	44	62	N=210

I- Les statistiques

❑ Bibliographie

- Bressoud E.; Kahané J.-C, (2010), *Statistique descriptive: applications avec Excel et calculatrices*, 2e éd, Pearson Education France , Paris.

- Hahn C.; Macé S., (2012), *Méthodes statistiques appliquées au management*, Pearson France.

- Legros B. (2016), *Statistiques et probabilités en économie-gestion*, Dunod Paris.