Statistique descriptive
Cours et exercices

Destiné aux Etudiants de la Première Année
de la Formation Préparatoire en Sciences et Technologies

Ecole Nationale Polytechnique d'Oran Maurice Audin

Dr Rafik Medjati

2023

Introduction

La statistique descriptive comprend :

- Des outils arithmétiques et graphiques pour décrire et visualiser de grandes séries de données.
- Des méthodes pour voir l'évolution d'une population statistique.
- Des techniques pour tester les liaisons entre les phénomènes.
- Des méthodes d'extrapolation des données historiques pour mieux entrevoir les évolutions futures.

Historiquement dans l'ancienne Babylone, 3000 ans avant Jésus, en Chine, plus de 2000 ans avant Jésus, et en Egypte, vers 1700 ans avant Jésus, la statistique rassemblait des renseignements concernant des populations telles que la connaissance du nombre d'habitants d'un pays, leur répartition par sexe, par âge, par catégorie socio-professionnelle et en économie telle que l'évaluation des ressources de l'état des stocks, …

Les méthodes statistiques sont aujourd'hui utilisées également dans :

- le domaine de la santé tel que l'évaluation de l'efficacité d'un médicament, de l'état sanitaire d'une population, …
- En agronomie telle que la recherche d'engrais précis.
- En sociologie telle que les enquêtes et les sondages.
- Dans l'industrie telle que l'organisation scientifique du travail, le contrôle de qualité, et la gestion des stocks, …
- Et bien d'autres domaines tel que le sport.

Ce manuel est conçu spécialement aux étudiants de première année universitaire, ces objectifs pédagogiques sont de favoriser l'apprentissage et développer l'autonomie de tous les étudiants, il est très utile pour l'apprentissage et pour l'entrainement à l'examen, il prend en compte tous les besoins en statistiques des étudiants de première année universitaire, notamment ceux de première année CPST des écoles nationales polytechniques d'Algérie.

Ce manuel de statistique descriptive présente de façon synthétique, structurée et illustrer l'ensemble des connaissances nécessaires est développé en trois chapitres.

Un premier chapitre introductif, dans lequel le vocabulaire des statistiques est exposé.

Le deuxième chapitre associe à chaque série statistique un ensemble de nombres réels qui sont appelés paramètres, d'une part des nombres qui se positionnent au milieu des valeurs dans un rangement croissant, d'autre part d'autres nombres qui nous donnent la dispersion au tour de la moyenne arithmétique.

Le troisième chapitre, l'étude statistique se fait sur deux variables mesurées simultanément sur les mêmes individus. On obtient donc deux mesures, la série statistique est alors une suite de couples des valeurs prises par les deux variables sur chaque individu.

Chacune des deux variables peut être, soit quantitative, soit qualitative. Par la suite une examinassions de la liaison entre les deux variables est établi ce qui permet de trouver un modèle d'ajustement en utilisant la méthode des moindres carrés ordinaires.

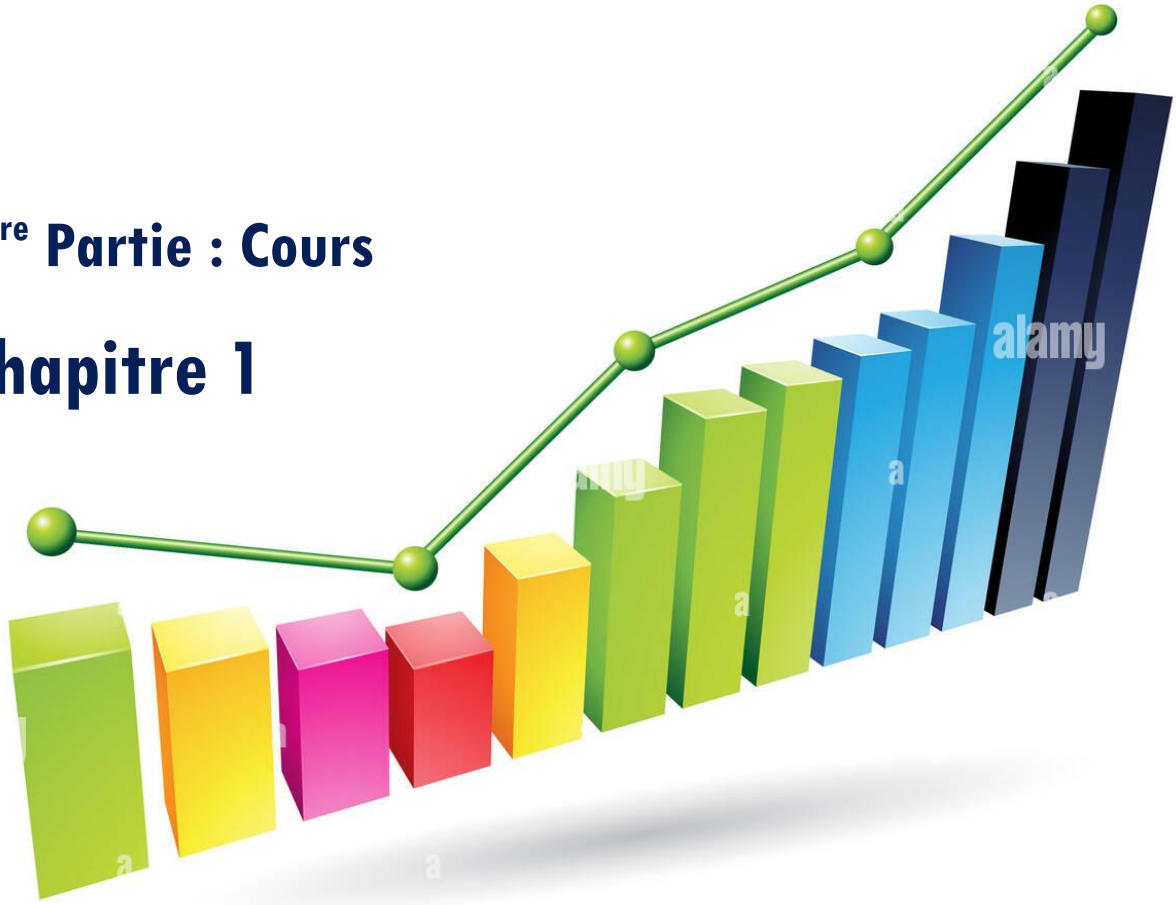
Ce manuel est développé avec un langage simple et facile à comprendre, englobe :

Des cours, des exemples et exercices variés, classés par thèmes.

Table des matières

1ère Partie : Cours

Chapitre 1



Vocabulaire de la statistique descriptive

Ce chapitre est introductif, consacré à la définition de la statistique descriptive ainsi que des différents termes qui en constituent le vocabulaire de base.

En fait la statistique est une discipline qui concerne la quantification de certains phénomènes et l'élaboration de procédures et de règles pour étudier ces données quantitatives. Les méthodes statistiques donnent des procédures et des règles qui sont très utiles à l'analyse numérique.

L'intérêt qui est porté à cette discipline est démontré par son utilisation intense dans de nombreux domaines tels que les sciences expérimentales (biologie, physique, agriculture, agronomie, médecine, industrie,...), les sciences humaines, l'économie, etc.

D'après Bernard PY, dans son livre Statistique descriptive, comprendre et réussir (éditions Economica) : «La statistique descriptive est un ensemble de méthodes permettant de décrireet d'analyser, de façon quantifiée, des phénomènes repérés par des éléments nombreux, de même nature, susceptibles d'être dénombrés et classés».

1.1 Définitions fondamentales de la statistique

Pour un groupe d'individus ou d'objets la statistique est l'étude de :

1. La collecte de données.
2. Leur analyse, leur traitement et l'interprétation des résultats.
3. Leur présentation afin de rendre les données compréhensibles par tous.

Remarque 1.1.1 D'une manière générale, la méthode statistique est basée sur quatre concepts : la population, les variables (les caractères), les observations et les données.

1.1.1 Population statistique

Une population statistique est l'ensemble d'objets ou de personnes sur lequel on effectue des observations.

Exemples 1.1.1

1. Ensemble de personnes interrogées pour une enquête.
2. Ensemble de pays pour lesquels on dispose de données géographiques ou économiques.

1.1.2 Individus (ou unités statistiques)

Les individus sont les éléments de la population statistique étudiée. Pour chaque individu, on dispose d'une ou plusieurs observations.

Exemples 1.1.2

1. Chacune des personnes interrogées pour une enquête.
2. Chaque pays pour lequel on étudie des données socio-économiques, . . .
3. Chaque jour de l'année pour lequel on dispose de données météorologiques, ...

Remarque 1.1.2 Lorsqu'on observe qu'une partie de la population, on parle de sondage. La partie de la population étudiée est appelée échantillon et on cherche toujours à généraliserles résultats obtenus sur l'échantillon à toute la population.

Exemples 1.1.3

1. Etudier les notes du module d'analyse sur un échantillon de 200 étudiants est une expérience statistique. Les unités statistiques ou bien les individus correspondent aux étudiants. La population est l'ensemble des étudiants.

2. L'étude du chômage sur un échantillon des étudiants récemment diplômés du master 2, la population pourrait être l'ensemble des étudiants diplômés du master. Les unités statistiques ou bien les individus correspondent aux diplômés.

1.1.3 Caractère statistique (ou variable statistique)

Chaque individu d'une population est décrit par un ensemble de caractéristiques appelées variables ou caractères. On les notes souvent par des lettres majuscules : $X, Y \ldots$, donc c'est ce qui est observé ou mesuré sur les individus d'une population statistique.

Exemples 1.1.4

1. Chacune des personnes interrogées pour une enquête.
2. Chaque pays pour lequel on étudie des données socio-économiques, ...
3. Chaque jour de l'année pour lequel on dispose de données météorologiques, ...

Remarque 1.1.3 Une variable doit donc présenter au minimum deux observations.

Exemples 1.1.5

1. La variables "sexe" a deux observations : masculin ou féminin.
2. Les observations de la variables "âge des ouvriers d'une entreprise", peuvent être : $[25, 30[\,; [30, 40[\,; [40, 45[$ et $[45, 60[$.

1.1.4 Modalité

Les modalités d'un caractère sont les différents résultats de l'observation (nombres ou propriétés).

Exemples 1.1.5

1. Cas qualitatif :
Les modalités de la variable X = « situation familiale » sont :
M ={ célibataire, marié, veuf, divorcé }.
2. Cas quantitatif discret :
Les modalités du la variable X = « Note à un examen » sont :
$M = \{7, 9, 14, 16, 5\}$.
3. Cas quantitatif continue :
Les modalités de la variable X = « Taille » sont les valeurs appartenant aux intervalles $[150, 165[$, $[165, 180[$, etc. . .

Remarque 1.1.4 Les variables statistiques peuvent être classées selon leurs natures, d'où il y'a deux types variables qualitative ou quantitative.

1.2 Variable quantitative

Une variable statistique est quantitative si ses valeurs sont des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne, ... ont un sens.

Remarque 1.2.1 Les variables quantitatives peuvent être discrètes ou continues.

1.2.1 Variable quantitative discrète

C'est une variable quantitative pouvant prendre par nature un nombre fini (ou dénombrable) de valeurs.

Exemples 1.2.1
1. Nombre d'enfants par famille.
2. Nombre de pièces d'un appartement.

1.2.2 Variable quantitative continue

C'est une variable quantitative pouvant prendre par nature une infinité de valeurs, généralement tout un intervalle réel.

Exemples 1.2.2

Tailles, poids, salaires, surfaces cultivées, température.

Remarque 1.2.2 Dans ce cas on utilise des intervalles $[a_i, b_i[$ au lieu de x_i.

1.3 Variable qualitative

Une variable statistique est qualitative si ses valeurs, ou modalités, s'expriment de façon littérale ou par un codage (ie une observation qui n'est pas mesurable).

Exemples 1.3.1
1. Sexe, situation familiale, . . .
2. Etat du temps constaté à un endroit donné chaque jour (pluvieux, neigeux, beau, venteux, ...)

Remarque 1.3.1 Il y'a deux types de variables statistiques qualitatives : nominale ou ordinale.

1.3.1 Variable qualitative nominale

La variable est dite qualitative nominale quand les modalités ne peuvent pas être ordonnées (ne peuvent pas être classées).

Exemples 1.3.2

La variable X = « situation familiale » avec les modalités notées **c** (célibataire), **m** (marié), **v** (veuve), **d** (divorcé).

1.3.2 Variable qualitative ordinale

La variable est dite qualitative ordinale quand les modalités peuvent être ordonnées.

Si $M = \{x_1, x_2, \ldots, x_r\}$ désigne l'ensemble des valeurs observées, ces valeurs sont ordonnées, c'est-à-dire :

$$x_1 \prec x_2 \prec \cdots \prec x_r.$$

La notation $x_1 \prec x_2$ se lit x_1 précède x_2.

Exemples 1.3.3

1. Un questionnaire de satisfaction demande aux consommateurs d'évaluer une prestation en cochant l'une des six catégories suivantes :
   (a) nulle, (b) médiocre, (c) moyenne, (d) assez bonne, (e) très bonne, (f) excellente.
   La variable X = « Niveau d'instruction ».

1.4 Effectif et fréquence d'une modalité (ou d'une classe)

1.4.1 Effectif

- L'effectif n_i d'une modalité (ou d'une classe) est le nombre de fois où la modalité (respectivement la classe) n° i a été observée.
- L'effectif total N est le nombre total d'individus observés.

$$N = n_1 + n_2 + \cdots + n_r = \sum_{i=1}^{r} n_i$$

1.4.2 Fréquence relative

- La fréquence (ou fréquence relative) f_i d'une modalité est le rapport de l'effectif n_i à l'effectif total N $\quad f_i = \dfrac{n_i}{N} = \dfrac{n_i}{\sum_{i=1}^{r} n_i}$

Remarque 1.4.1 Les fréquences relatives peuvent être exprimées en pourcentages, et on a le résultat suivant :

$$\sum_{i=1}^{r} f_i = 1$$

Car $\sum_{i=1}^{r} f_i = \sum_{i=1}^{r} \frac{n_i}{N} = \frac{1}{N}\sum_{i=1}^{r} n_i = \frac{N}{N} = 1$

Exemple 1.4.1

Sur 200 familles, 50 ont 2 enfants, on dira que la fréquence f_i correspondant à la valeur $x_i = 2$ de la variable « nombre d'enfants », est :

$$f_i = \frac{n_i}{N} = \frac{50}{200} = \frac{1}{4} = 0{,}25 \text{ soit } 25\%$$

1.5 Présentation dans un tableau statistique

1.5.1 Cas qualitatif nominale

Pour une variable statistique qualitative nominale, si l'ensemble $M = \{x_1, x_2, \ldots, x_r\}$ désigne l'ensemble des modalités, alors le tableau statistique associé à ce caractère est

Modalités (numérotés) x_i	*Effectifs* n_i	*Fréquences* f_i
x_1	n_1	f_1
x_2	n_2	f_2
.	.	.
.	.	.
.	.	.
x_r	n_r	f_r
Total	N	1

Exemple 1.5.1

On s'intéresse aux valeurs de la variable X = «situation familiale» prises sur 20 personnes dont la codification est ;
c : célibataire, ***m*** : marié(e), ***v*** : veuf(ve), ***d*** : divorcé(e). Donc le domaine de la variable X est $M = \{x_1, x_2, \ldots, x_r\}$.

Considérons les résultats suivants :

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
m	m	d	c	c	m	c	c	c	m	c	m	v	m	v	d	c	c	c	m

et on obtient le tableau suivant :

x_i	Effectifs n_i	Fréquences f_i
c	9	0,45
m	7	0,35
v	2	0,10
d	2	0,10
Total	20	1

Remarque 1.5.1 Avant d'aborder les autres cas on définit ce qui suit.

1.5.2 Fréquences cumulées croissantes F_i (f_{ic})

Pour $i = 1$, la fréquence cumulée croissante d'ordre 1 est $F_1 = f_1$.

Pour $i = 2$, la fréquence cumulée croissante d'ordre 2 est $F_2 = f_1 + f_2$

En général pour i quelconque la fréquence cumulée croissante d'ordre i est :

$$F_i = f_{ic} = f_1 + f_2 + \cdots + f_i = \sum_{p=1}^{i} f_p$$

1.5.3 Fréquences cumulées décroissantes F'_i (f_{id})

Pour $i = 1$, la fréquence cumulée croissante d'ordre 1 est $F'_r = f_r$

Pour $i = 2$, la fréquence cumulée croissante d'ordre 2 est $F'_{r-1} = f_r + f_{r-1}$.

En général pour i quelconque la fréquence cumulée décroissante d'ordre i est :

$$F'_i = f_{id} = f_r + f_{r-1} + \cdots + f_i = \sum_{p=i}^{r} f_p$$

Remarque 1.5.2 De la même manière on définit les effectifs cumulés croissants N_i (n_{ic}) et les effectifs cumulés décroissants N'_r (n_{id}).

$$N_i = n_{ic} = n_1 + n_2 + \cdots + n_i = \sum_{p=1}^{i} n_p$$

et

$$N'_i = n_{id} = n_r + n_{r-1} + \cdots + n_i = \sum_{p=i}^{r} n_p$$

Remarque 1.5.3 Les Fréquences cumulées et les effectifs cumulés sont utilisés pour les deux types de la variable quantitatif et la variable qualitatif ordinale.

1.5.4 Cas qualitatif ordinale et quantitatif discret

Si $M = \{x_1, x_2, \ldots, x_r\}$ désigne l'ensemble des modalités d'une variable qualitative ordinale, c'est à dire ces valeurs sont ordonnées : $x_1 \prec x_2 \prec \cdots \prec x_r$.
Avec $x_1 \prec x_2$ se lit x_1 précède x_2, ainsi le tableau associé est

x_i	*Effectifs* n_i	*Effectifs cumulés croissants* N_i	*Fréquences* f_i	*Fréquences cumulées croissantes* F_i
x_1	n_1	N_1	f_1	F_1
x_2	n_2	N_2	f_2	F_2
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
x_r	n_r	$N_r = N$	f_r	$F_r = 1$
Total	N		1	

Exemple 1.5.2
20 chemises sont classées par taille :
x_1 = S, x_2 = M, x_3 = L, x_4 = XL et x_5 = XXL. Le tableau associé est

x_i	*Effectifs* n_i	*Effectifs cumulés croissants* N_i	*Fréquences* f_i	*Fréquences cumulées croissantes* F_i
x_1	4	4	0,20	0,20
x_2	2	6	0,10	0,30
x_3	5	11	0,25	0,55
x_4	8	19	0,40	0,90
x_5	1	20	0,05	1,00
Total	20		1,00	

Remarque 1.5.4 Le cas quantitatif discret se fait de la même manière que le cas qualitatif ordinal, et on obtient un tableau statistique semblable à celui du cas qualitatif ordinal.

1.5.5 Cas quantitatif continue

Dans le cas quantitatif continu on aura le tableau suivant

Classes $[b_{i-1}, b_i[$	*Centres* c_i	n_i	N_i	f_i	F_i
$[b_0, b_1[$	c_1	n_1	N_1	f_1	F_1
$[b_1, b_2[$	c_2	n_2	N_2	f_2	F_2
.	.	.	.	.	.
.	.	.	.	.	.
$[b_{r-1}, b_r[$	c_r	n_r	$N_r = N$	f_r	$F_r = 1$
Total		N		1	

avec le centre d'une classe est $c_i = \frac{b_i - b_{i-1}}{2}$.

et l'amplitude d'une classe est $a_i = b_i - b_{i-1}$

Exemple 1.5.3

La répartition de 100 ménages selon leurs dépenses de consommation mensuelles exprimées en milliers dinars se présente comme suit :

Classes de dépenses	*Nombre de ménages*
[20, 40[	15
[40, 60[	20
[60, 100[	20
[100, 200[	45

et le tableau associé est :

Classes $[b_{i-1}, b_i[$	*Centres* c_i	n_i	N_i	f_i	F_i
[20, 40[	30	15	15	0,15	0,15
[40, 60[	50	20	35	0,20	0,35
[60, 100[	80	20	55	0,20	0,55
[100, 200[	150	45	100	0,45	1,00
Total		100		1,00	

Exemple 1.5.4 : (le calcul de F'_i et N'_i)

On calcul les effectifs cumulés décroissants et les fréquences cumulées croissantes comme apparait dans le tableau

Classes $[b_{i-1}, b_i[$	*Centres* c_i	n_i	N'_i	f_i	F'_i
[20, 40[	30	15	100	0,15	1,00
[40, 60[	50	20	85	0,20	0,85
[60, 100[	80	20	65	0,20	0,65
[100, 200[	150	45	45	0,45	0,45
Total		100		1,00	

Remarque 1.5.5 L'ensemble des couples

1. $\{(x_i, n_i)\}$ ou encore $\{(x_i, f_i)\}$ si la variable est discrète.
2. $\{([b_{i-1}, b_i[, n_i)\}$, ou $\{([b_{i-1}, b_i[, f_i\}$ si la variable est continue.

Est appelé série statistique de la variable.

1.6 Séries statistiques et leurs représentations graphiques

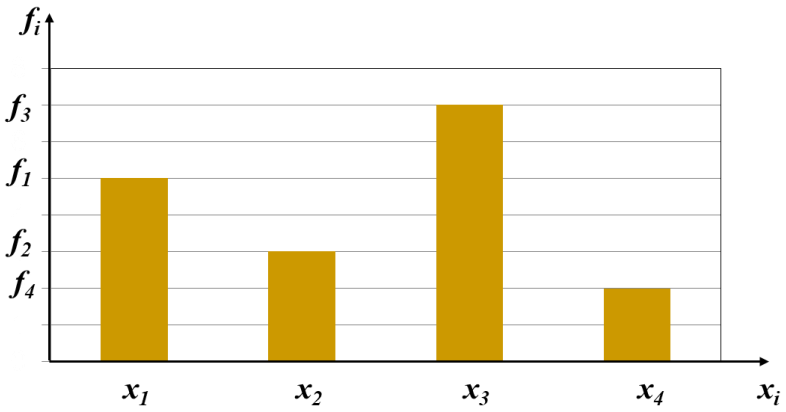
Un tableau statistique contient toute l'information prélevée sur un échantillon mais il est très utile de présenter ces information par un graphique, afin d'avoir un aperçu de l'évolution de la variable. On utilise des différents types de représentations graphiques suivant la nature de la variable étudiée.

1.6.1 Cas qualitatif

- **Diagramme à bandes**

C'est un repère cartésien tel que :

à chaque modalité x_i on associe un rectangle de base constante dont la hauteur est l'effectif n_i (la fréquence f_i) comme est présenté dans la figure suivante

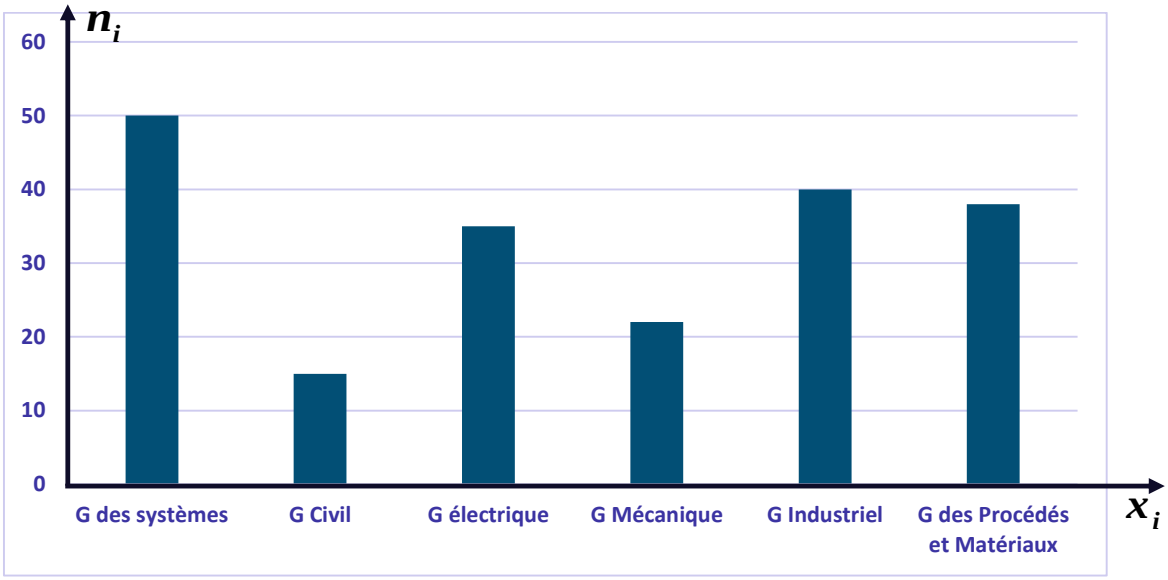


Remarque 1.6.1 Pour l'axe des effectifs (fréquences), on choisit une échelle arithmétique.

Exemple 1.6.1 D'après une étude faite à l'ENPO-Oran, la répartition de 200 étudiants de deuxième année des CPST admis en troisième année, sur les spécialités préférées a fourni les résultats suivants

Spécialité préférée x_i	*Nombre d'élèves* n_i
Génie des systèmes	50
Génie Civil	15
Génie électrique	35
Génie Mécanique	22
Génie Industriel	40
Génie des Procédés et Matériaux	38
Total	200

Ces résultats peuvent être traduits par le diagramme à bandes suivant

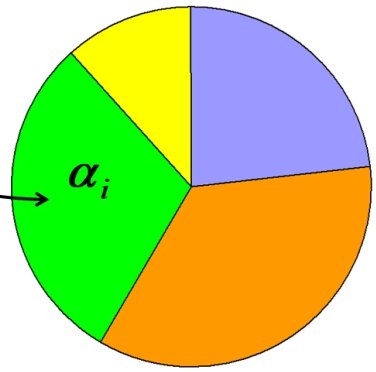


- **Diagramme à secteurs (circulaires)**

C'est un graphique où, les modalités sont représentées par des portions de disque proportionnelles à leurs effectifs, ou à leurs fréquences.

En effet, pour une modalité x_i, d'effectif n_i, l'angle au centre α_i correspondant est donné (en degré) par :

$$\alpha_i = f_i \times 360° = \frac{n_i}{N} \times 360°$$



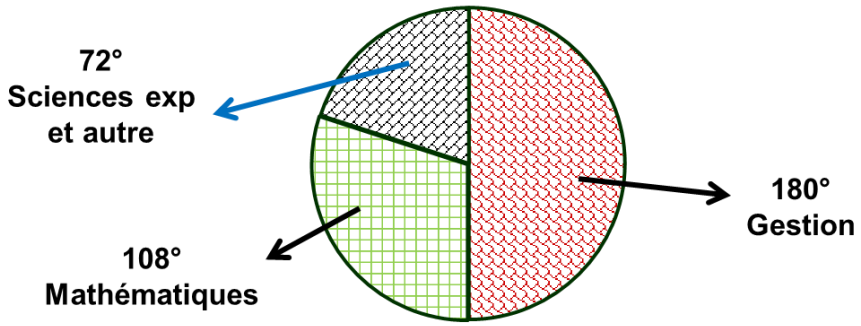
Remarque 1.6.2 Les diagrammes à bandes et circulaire peuvent être utilisés dans le cas quantitatif.

Exemple 1.6.2

D'après une étude faite à l'école de commerce d'Oran, la répartition de 50 étudiants selon la branche du bac est reportée dans le tableau suivant :

Section du bac x_i	*Effectifs n_i*	f_i	α_i
Gestion	25	0,50	180°
Mathématiques	15	0,30	108°
Sciences exp et autres	10	0,20	72°
Total	50	1,00	360°

Ces données peuvent être traduits par le diagramme à secteurs suivant



1.6.2 Cas quantitatif discret (Diagrammes en bâtons)

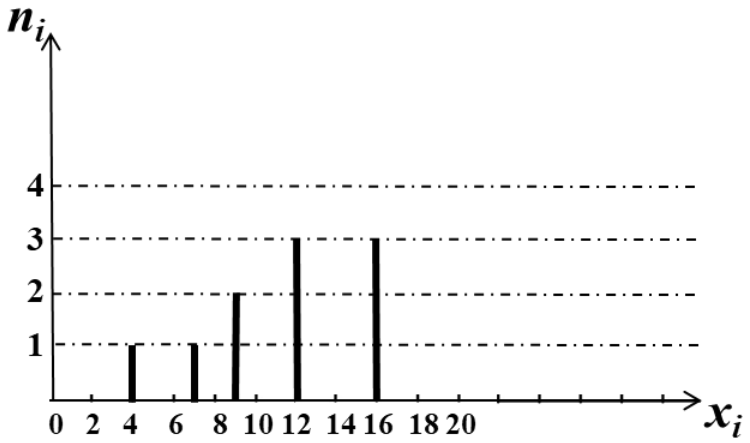
C'est un repère cartésien tel que les valeurs sont placées en abscisse, les effectifs (ou fréquences) en ordonnée, et à chaque point $(x_i, 0)$ on associe un segment vertical dont la longueur est l'effectif n_i (la fréquence f_i), voir les deux exemples suivants.

Exemples 1.6.3

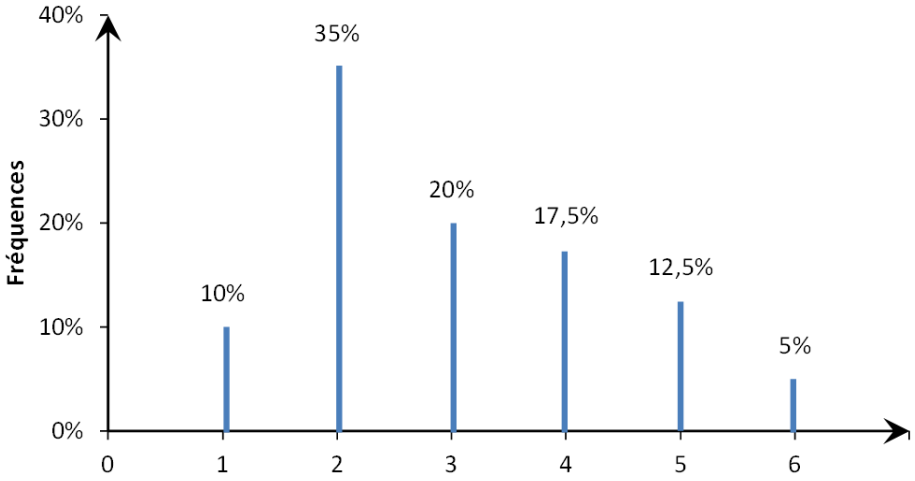
La répartition des notes d'un contrôle noté sur 20 pour une classe de 10 élèves est reportée dans le tableau suivant :

x_i	*Effectifs n_i*
4	1
7	1
9	2
12	3
16	3
Total	10

donne



La répartition d'une population de familles selon le nombre d'enfant de chaque famille est représentée sous le diagramme en bâtons suivant

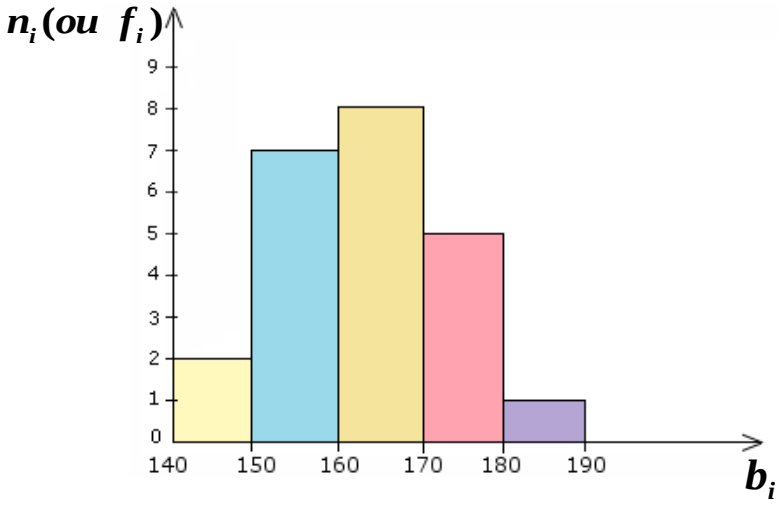


A partir de ce diagramme on peut construire le tableau des données suivant

x_i	*Fréquences en* (%) f_i
1	10
2	35
3	20
4	17,5
5	12,5
6	5
Total	100 %

1.6.3 Cas quantitatif continu (Histogramme)

On représente une série statistique continue par **un histogramme**. Il s'agit d'une figure obtenue sur un repère cartésien en représentant pour chaque classe $[\, b_{i-1} , b_i \,[$ **un rectangle de surface** S_i proportionnelle à l'effectif n_i ou à la fréquence f_i. Les rectangles de l'histogramme **sont voisins**, comme est présenté dans la figure suivante



- **Principe de construction de l'histogramme**

Il y'a deux (02) cas

1er cas : Cas d'amplitudes égales

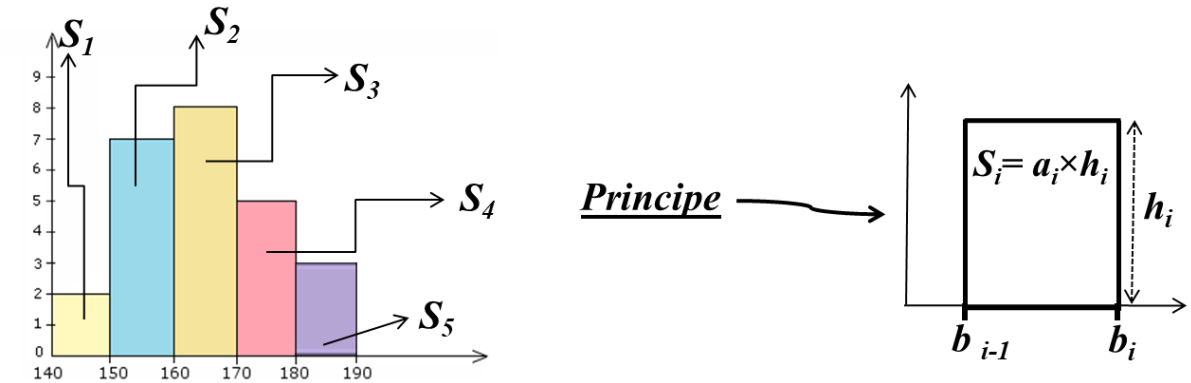
Si les classes sont de même amplitude a_i (ie $a_i = a_j$), on place en ordonnée les effectifs n_i (ou les fréquences f_i).

Exemple 1.6.4

La répartition de 20 personnes selon leurs tailles est donnée dans le tableau suivant :

$[b_{i-1}, b_i[$	n_i	a_i	f_i
[140 , 150[	2	10	0,08
[150 , 160[	7	10	0,28
[160 , 170[	8	10	0,32
[170 , 180[	5	10	0,20
[180 , 190[	3	10	0,12

Ces données peuvent être représentées par l'histogramme suivant



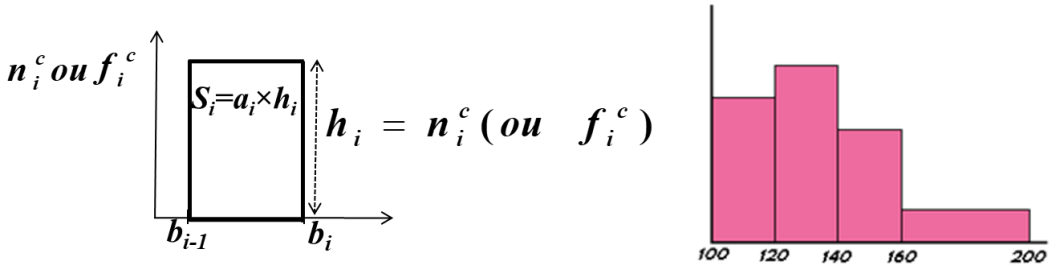
2eme cas : Cas d'amplitudes inégales

Si les classes sont d'amplitude inégales a_i (ie $a_i \neq a_j$), on définit

1. **La densité** d'une classe par : $d_i = \dfrac{n_i}{a_i}$

   et on pose $h_i = \dfrac{n_i}{a_i} \times a^* = d_i \times a^* = n_i^c$

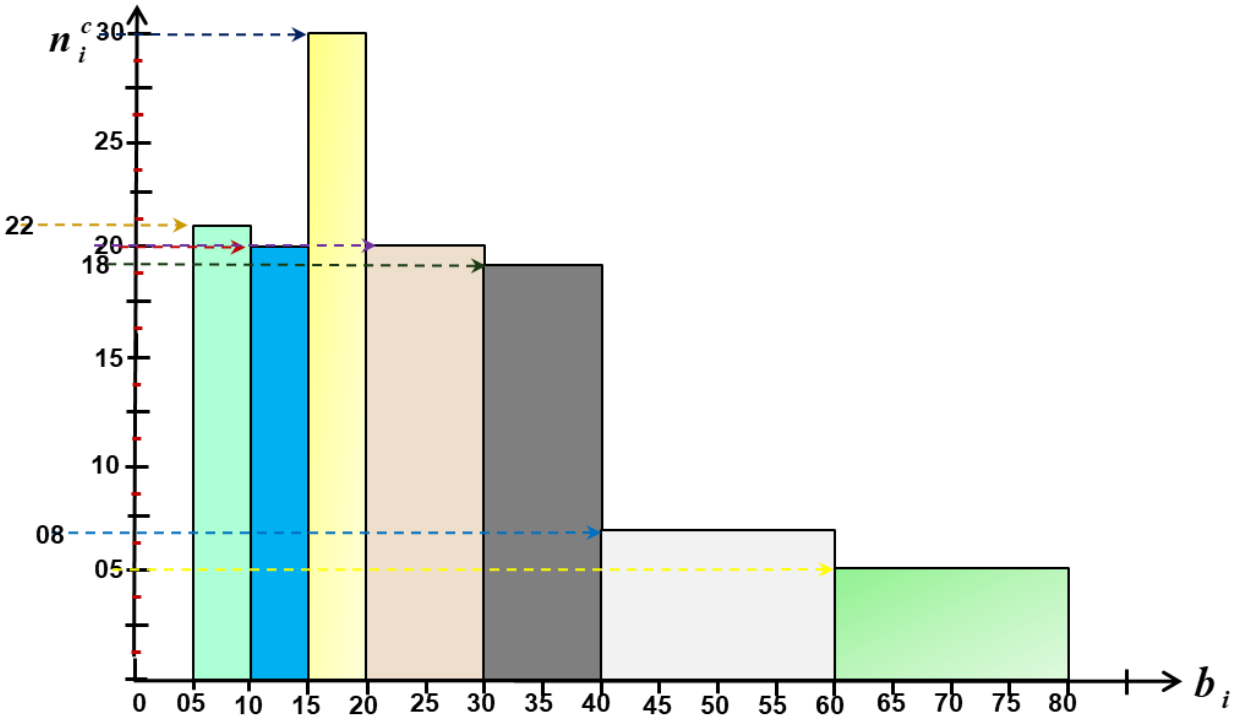
2. avec a^* est appelée **amplitude de référence**. Elle est choisie arbitrairement de manière à faciliter la représentation graphique (valeurs sur l'axe des ordonnées).
3. h_i est dans ce cas est appelée **effectif corrigé** qu'on note n_i^c .
   Ainsi les rectangles S_i seront de la manière suivante :



Exemple 1.6.5 La répartition de 100 individus par classes d'âges est donnée par le tableau suivant :

Classe	n_i	a_i	$d_i = \dfrac{n_i}{a_i}$	$n_i^c = d_i \times a^*$	f_i	$f_i^c = \dfrac{n_i^c}{N} = \dfrac{f_i}{a_i} \times a^*$
[05 , 10[	11	5	2,2	22	0,11	0,22
[10 , 15[	10	5	2,0	20	0,10	0,20
[15 , 20[	15	5	3,0	30	0,30	0,30
[20 , 30[	20	10	2,0	20	0,20	0,20
[30 , 40[	18	10	1,8	18	0,18	0,18
[40 , 60[	16	20	0,8	08	0,16	0,08
[60 , 80[	10	20	0,5	05	0,10	0,05
Total	100				1,00	

Si on prend l'amplitude de référence $a^* = 10$ on obtient l'histogramme des effectifs suivant :



Remarque 1.6.3

1. Il existe deux types d'histogramme (histogramme des effectifs et histogramme des fréquences).
2. On associe à un histogramme des fréquences un polygone des fréquences qui se définit comme suit :

- **Polygone des fréquences**

Il s'agit d'une **ligne brisée** reliant :

1. **les milieux des sommets des rectangles** de l'histogramme.
2. **La fermeture** se fait par **deux points sur l'axe des abscisses** situés respectivement à **un demi-intervalle de La borne inférieure de la première classe et de la borne supérieure de la dernière classe.**

Pour bien voir un cas d'un histogramme avec son polygone des fréquences faisant l'exercice suivant.

Exercice d'application

Une association de course à pieds a une équipe féminine. La liste suivante est composée des prénoms d'athlètes suivis entre parenthèses du derniers temps aux 10 Km.

Aicha(51), Ahlem(49), Amel(50), Badra(58), Bouchra(55), Dalia(64), Fadia(60), Fahima(61), Fatiha(46), Fatima(56), Fouzia(50), Hajera(42), Houria(54), Ikram(48), Ilham(45), Imane(57), Jamila(59), Khadija(54), Lamia(54), Leila(46), Meriem(46),

Nabila(41), Samia(39), Samira(37), Wafaa(50), Yamina(47), Yasmine(50), Zahira(44), Zakia(51), Zoulikha(59).

Le responsable de l'association décide de créer dans un ordre croissant cinq (05) équipes (classes) de niveau d'athlètes équivalents telles que :

la 1ere équipe contient 3 athlètes, la 2eme équipe contient 3 athlètes, la 3eme équipe contient 6 athlètes, la 4eme équipe contient 9 athlètes, et la 5eme équipe contient 9 athlètes.

1. Constituer les équipes. (Faire un tableau donnant les temps minimums et maximums pour chacune des équipes).
2. Donner une représentation graphique des fréquences sous forme d'un histogramme (l'amplitude de référence $a^* = 1000$).
3. Dessiner le polygone des fréquences.

Solution de l'exercice d'application

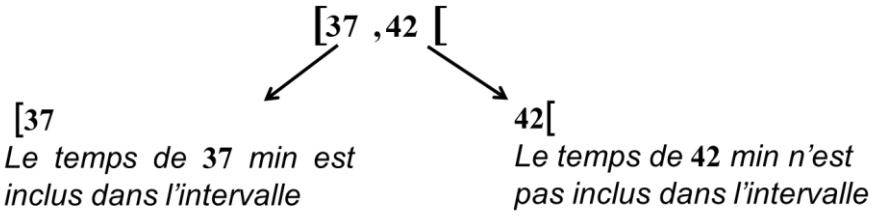
1. **Constitution des équipes** : La liste des athlètes est ;

Aicha(51), Ahlem(49), Amel(50), Badra(58), Bouchra(55), Dalia(64), Fadia(60), Fahima(61), Fatiha (46), Fatima(56), Fouzia(50), Hajera(42), Houria(54), Ikram(48), Ilham(45), Imane(57), Jamila(59), Khadija(54), Lamia(54), Leila(46), Meriem(46), Nabila(41), Samia(39), Samira(37), Wafaa(50), Yamina(47), Yasmine(50), Zahira(44), Zakia(51), Zoulikha(59).

Et on va ranger par ordre croissant des temps : Les 3 sportives dont le temps est le plus petit (les meilleures) constitueront l'équipe 1 ;

Equipe 1	Equipe 2	Equipe 3	Equipe 4	Equipe 5
Samira (37)	Hajera (42)	Leila (46)	Amel (50)	Bouchra (55)
Samia (39)	Zahira (44)	Fatiha (46)	Wafaa (50)	Fatima (56)
Nabila (41)	Ilham (45)	Meriem (46)	Yasmine (50)	Imane (57)
		Yamina (47)	Fouzia (50)	Badra (58)
		Ikram (48)	Aicha (51)	Jamila (59)
		Ahlem (49)	Zakia (51)	Zoulikha (59)
			Khadija (54)	Fadia (60)
			Houria (54)	Fahima (61)
			Lamia (54)	Dalia (64)

On a créé ainsi des « classes ». On les écrit sous forme d'intervalle, par exemple l'intervalle de temps de l'équipe 1 est :



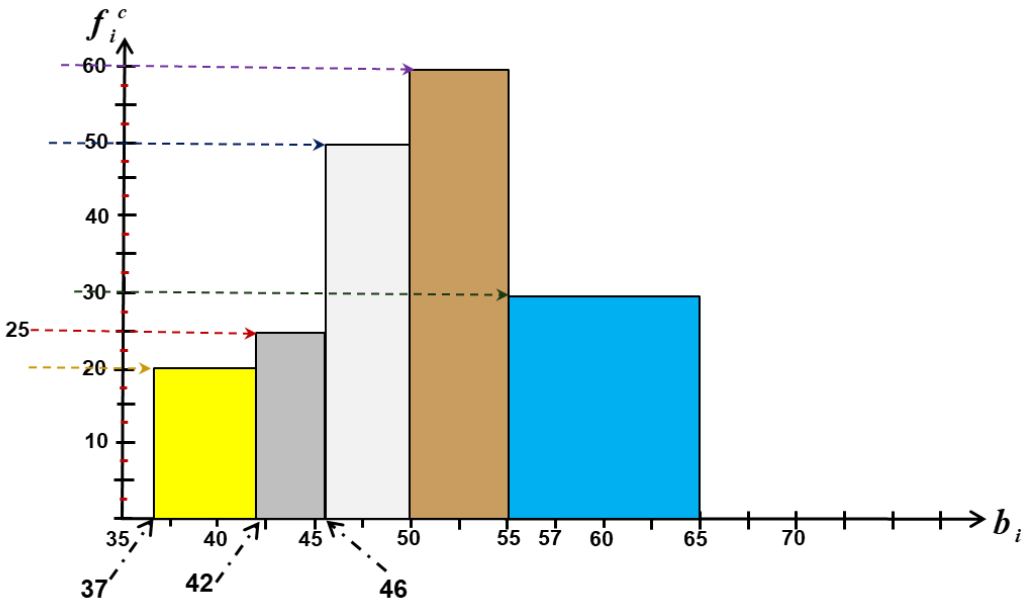
On peut ainsi construire un nouveau tableau :

$[b_{i-1}, b_i[$	n_i
[37, 42[	3
[42 , 46[	3
[46 , 50[	6
[50 , 55[	9
[55, 65[	9

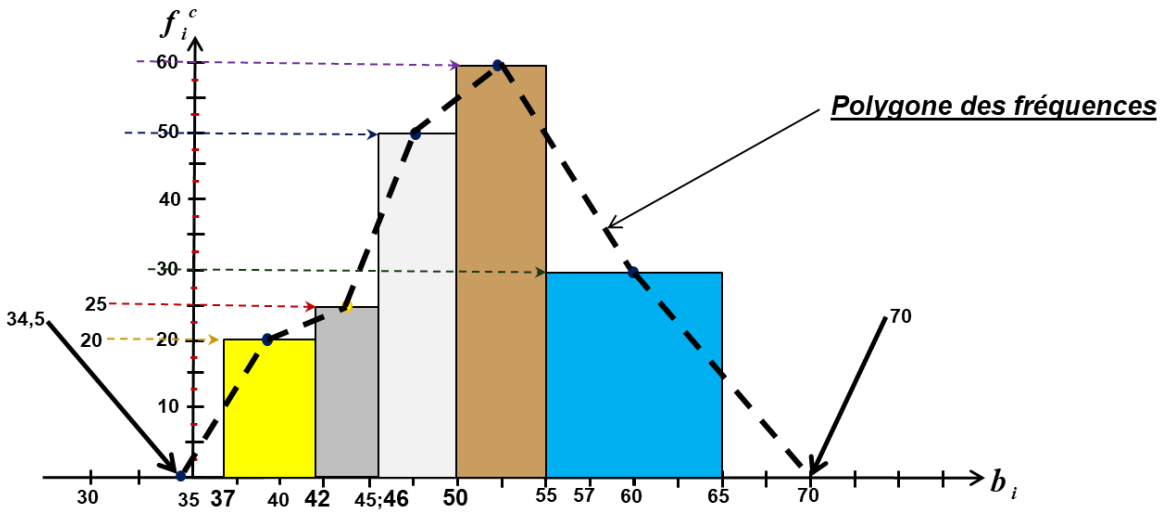
2. La représentation graphique des fréquences sous forme d'un histogramme.
On va calculer les différentes amplitudes :

Classe	n_i	a_i	$d_i = \frac{n_i}{a_i}$	$n_i^c = d_i \times a^*$	f_i	$f_i^c = \frac{n_i^c}{N} = \frac{f_i}{a_i} \times a^*$
[37, 42[	3	5	0,60	600	0,10	20
[42, 46[	3	4	0,75	750	0,10	25
[46, 50[	6	4	1,50	1500	0,20	50
[50, 55[	9	5	1,80	1800	0,30	60
[55, 65[	9	10	0,90	900	0,30	30
Total	30				1,00	

Donc on obtient l'histogramme des fréquences suivant



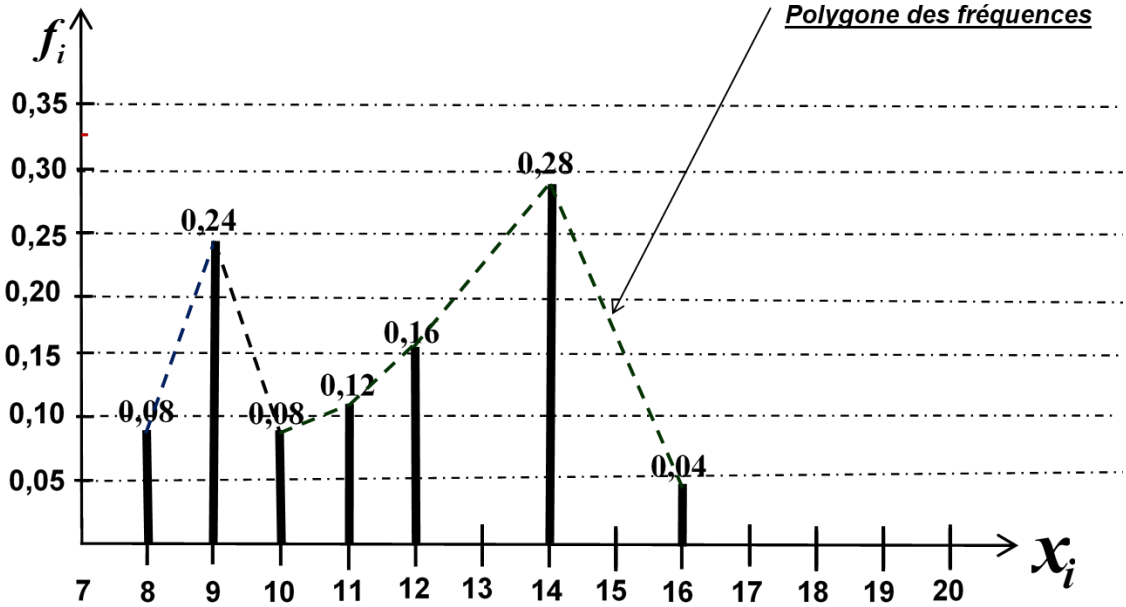
Ainsi le polygone des fréquences apparait sur cet histogramme comme suit



Remarque 1.6.4

1. **Le polygone des effectifs** se définit de la même manière en l'associant à un histogramme des effectifs.
2. Pour le cas d'une variable statistique discrète **Le polygone des effectifs** respectivement des fréquences) se dessine sur le diagramme en bâton des effectifs (respectivement des fréquences) en reliant les sommets des bâtons.

Exemple 1.6.6 La répartition des notes d'un contrôle noté sur 20 des élèves d'une classe est reportée dans diagramme en bâton des fréquences suivant avec la représentation du polygone des fréquences.



1.7 Fonction de répartition et diagramme cumulatif

On appelle fonction de répartition d'une variable statistique quantitative toute application définie par :

$$F : \mathbb{R} \to [0, 1]$$
$$x \to F(x) = P(X \leq x)$$

$F(x)$ proportion des individus dont la valeur de la variable est strictement inférieure ou égale à x, c'est-à-dire $X \leq x$.

1.7.1 Cas de la variable statistique quantitative discrète

$F(x)$ = fréquence de $(X \leq x) = f_1 + f_2 + \ldots + f_p = F_p$ tel que : $f_1, f_2, \ldots, f_p$ sont les fréquences des valeurs de la variable $\leq x$, si non $F(x) = 0$. Donc

$$F(x) = \begin{cases} 0 & si \quad x < x_1 \\ F_i & si \quad x_i \leq x < x_{i+1} \\ 1 & si \quad x_r \leq x \end{cases}$$

tel que r désigne l'ordre de la **dernière valeur (modalité)**.

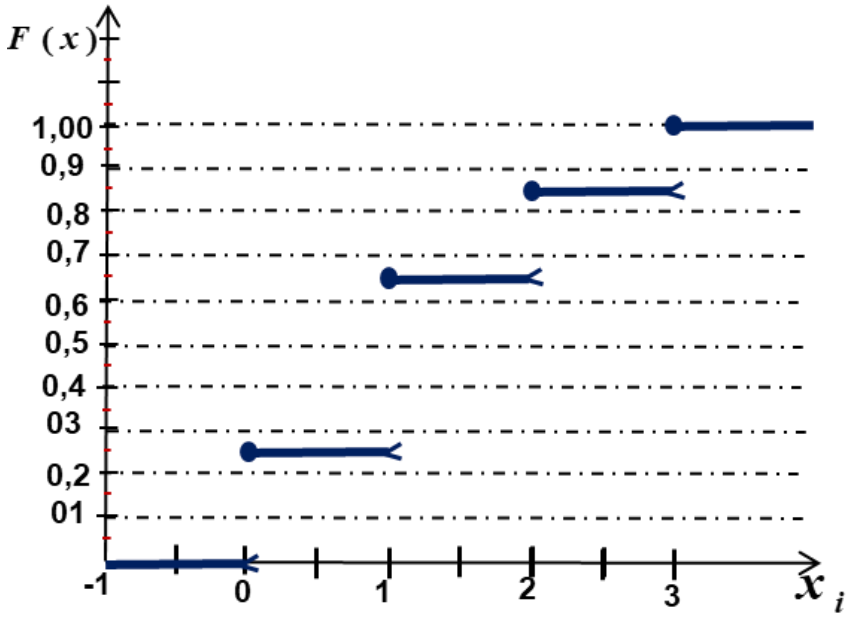
Exemple 1.7.1 Le tableau suivant, donne le nombre d'absences des étudiants au module d'analyse.

Nbre d'absences x_i	*Effectifs* n_i	f_i	F_i
0	5	0,25	0,25
1	8	0,40	0,65
2	4	0,20	0,85
3	3	0,15	1,00
Total	20	1,00	

Donc la fonction de répartition correspondante est

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 0{,}25 & si \quad 0 \leq x < 1 \\ 0{,}65 & si \quad 1 \leq x < 2 \\ 0{,}85 & si \quad 2 \leq x < 3 \\ 1 & si \quad 3 \leq x \end{cases}$$

Ainsi on obtient la représentation de la fonction de répartition, appelée **diagramme cumulatif** ou **diagramme intégral**



Remarque 1.7.1 Dans le cas discret on a une fonction en escalier.

1.7.2 Cas de la variable statistique quantitative continu

Dans ce cas on commence par la technique d'obtention de la courbe de la fonction de répartition qui est appelée **courbe cumulative** qui est une courbe approximative, en utilisant l'interpolation linéaire entre deux points dans le plan.

En fait : L'**interpolation linéaire** est la méthode la plus simple pour estimer la valeur prise par une fonction continue entre deux points déterminés (interpolation). Elle consiste à utiliser pour cela la fonction affine (de la forme $f(x) = m.x + b$) passant par les deux points déterminés.

Donc la fonction de répartition dans le cas quantitative continu est définit de la même façon que dans le cas quantitatif discret :

$$F : \mathbb{R} \to [\,0\,,\,1\,]$$
$$x \to F(x) = P(X \leq x)$$

$F(x)$ proportion des individus dont la valeur de la variable est strictement inférieure ou égale à x, c'est-à-dire $X \leq x$.

La courbe cumulative, est une ligne brisée obtenue en joignant :

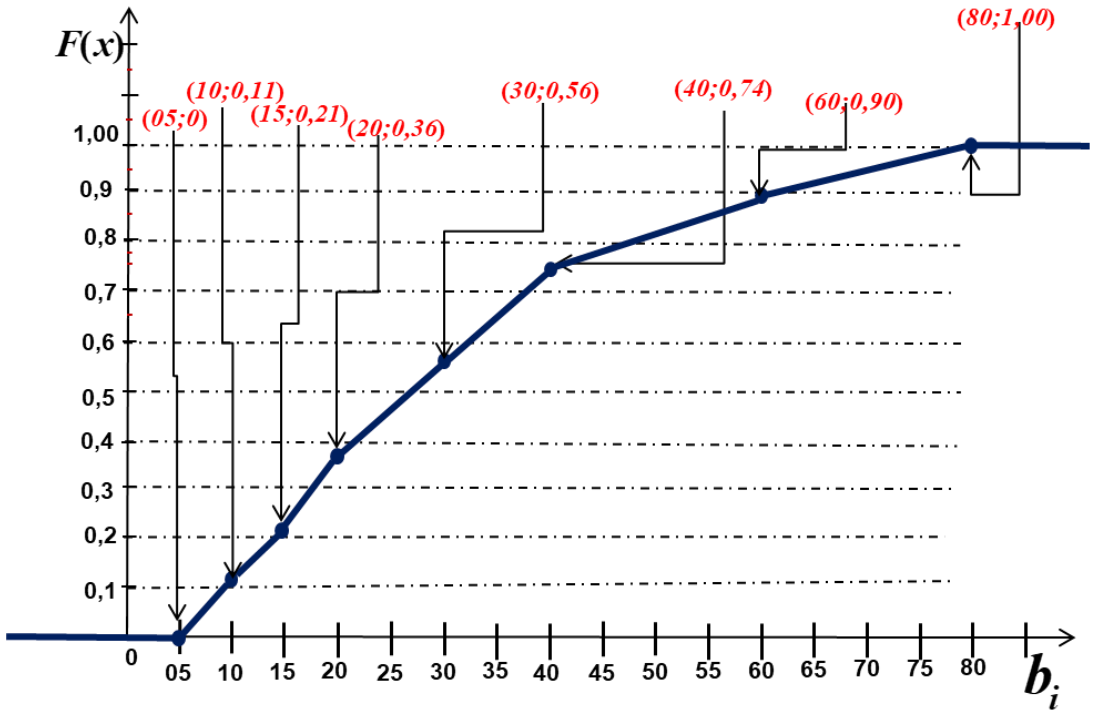
1. **Les différents points de coordonnés (b_i, F_i) dans l'ordre croissant avec $F_0 = 0$.**
2. Et en joignant du côté gauche du point $(b_0\,,\,F_0)$ la ½ droite $y = 0$ et du coté droit du

point $(b_r\ ,\ F_r)$ la ½ droite $y = 1$.

Exemple 1.7.2 On reprend l'exemple de la page 16.

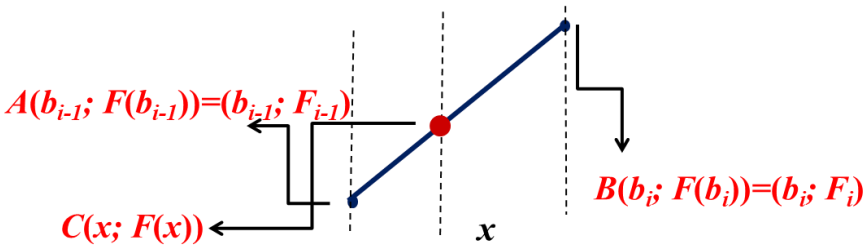
$[b_{i-1}\ ,\ b_i[$	n_i	N_i	f_i	F_i
[5 , 10[	11	11	0 ,11	0,11
[10 , 15[	10	21	0 ,10	0,21
[15, 20[	15	36	0 ,15	0,36
[20, 30[	20	56	0 ,20	0,56
[30 , 40[	18	74	0 ,18	0,74
[40 , 60[	16	90	0 ,16	0,90
[60, 80[	10	100	0 ,10	1,00

Ainsi la courbe de la fonction de répartition, appelée **courbe cumulative** se dessine comme suit :



- **Technique de calcul d'une valeur de $F(x)$ pour $x \in \mathbb{R}$:**

Pour calculer $F(x)$, $\forall\ x \in \mathbb{R}$ on va réaliser **une interpolation linéaire** entre les points $A(b_{i-1}\ ;\ F(b_{i-1})) = (b_{i-1}\ ;\ F_{i-1})$ et $B(b_i\ ;\ F(b_i)) = (b_i\ ;\ F_i)$ tels que $x \in [b_{i-1}\ ,\ b_i[$



L'équation de la droite (AB) est de la forme $y = mx + p$ tel que :

$$m = \frac{y_B - y_A}{x_B - x_A} = \frac{y_C - y_A}{x_C - x_A} = \frac{F(b_i) - F(b_{i-1})}{b_i - b_{i-1}} = \frac{F_i - F_{i-1}}{b_i - b_{i-1}}$$

$$\Rightarrow F(x) = F_{i-1} + m(x - b_{i-1}) = F_{i-1} + \frac{f_i}{b_i - b_{i-1}}(x - b_{i-1})$$

Remarque 1.7.2 Dans un cas contraire on peut calculer x si on a la valeur de $F(x)$, en utilisant toujours **une interpolation linéaire** entre les points $A(b_{i-1}\,;\, F(b_{i-1})) = (b_{i-1}\,;\, F_{i-1})$ et $B(b_i\,;\, F(b_i)) = (b_i\,;\, F_i)$ tels que $x \in [b_{i-1}, b_i[$ et $(F(x) \in [F_{i-1}, F_i\,[\,)$, en effet :

$$\frac{F(b_i) - F(b_{i-1})}{b_i - b_{i-1}} = \frac{F_i - F_{i-1}}{b_i - b_{i-1}} = \frac{F(x) - F_{i-1}}{x - b_{i-1}}$$

$$\Rightarrow x = (F(x) - F_{i-1})\left(\frac{b_i - b_{i-1}}{F_i - F_{i-1}}\right) + b_{i-1} \Rightarrow x = b_{i-1} + a_i\left(\frac{F(x) - F_{i-1}}{F_i - F_{i-1}}\right)$$

Ainsi on peut formuler la fonction de répartition d'une façon générale comme suit

$$F(x) = \begin{cases} 0 & si \quad x < b_0 \\ F_{i-1} + \dfrac{f_i}{b_i - b_{i-1}}(x - b_{i-1}) & si \quad b_{i-1} \le x < b_i \\ 1 & si \quad b_r \le x \end{cases} \qquad \text{avec } F_0 = 0.$$

Et la fonction de répartition pour la série statistique dans l'exemple cité ci-dessus est

$$F(x) = \begin{cases} 0 & si \quad x < 5 \\ 0 + \dfrac{0{,}11}{5}(x-5) & si \quad 5 \le x < 10 \\ 0{,}11 + \dfrac{0{,}10}{5}(x-10) & si \quad 10 \le x < 15 \\ 0{,}21 + \dfrac{0{,}15}{5}(x-15) & si \quad 15 \le x < 20 \\ 0{,}36 + \dfrac{0{,}20}{10}(x-20) & si \quad 20 \le x < 30 \end{cases} \quad \text{et} \quad F(x) = \begin{cases} 0{,}56 + \dfrac{0{,}18}{10}(x-30) & si \quad 30 \le x < 40 \\ 0{,}74 + \dfrac{0{,}16}{20}(x-40) & si \quad 40 \le x < 60 \\ 0{,}90 + \dfrac{0{,}10}{20}(x-60) & si \quad 60 \le x < 80 \\ 1 & si \quad 80 \le x \end{cases}$$

1.8 Exercices du chapitre 1

Exercice 1

La liste suivante est composée de prénoms d'un groupe de personnes, suivis entre parenthèses du nombre d'enfants que chacun d'entre eux avait au 31 décembre 2016 :

Abdelkader (6), Ahmed (3), Dahmane (3), Djilali (3), Farida (1), Fatima (4), Halim (3), Houari (4), Kadour (3), Mohamed (2), Mounir (4), Nadir (0), Nacer (2), Omar (1), Sabri (2), Sidahmed (2), Sofiane (0), Zakaria (5), Zohir (1), Zoubir (2).

1. Déterminez la population étudiée et la variable étudiée.
2. Précisez la nature de la variable ainsi que ces modalités.
3. Représentez la distribution des fréquences par un diagramme en bâtons.
4. Calculez les effectifs cumulés croissants et cumulés décroissants.
5. Calculez les fréquences cumulées croissantes, cumulées décroissantes.

Exercice 2

On interroge 50 personnes sur leur dernier diplôme obtenu. La codification a été faite selon le tableau suivant :

Dernier diplôme obtenu	*Sans diplôme*	*Primaire*	*Secondaire*	*Supérieur non-universitaire*	*Universitaire*
x_i	*Sd*	*P*	*Se*	*Su*	*U*

Le tableau suivant indique la répartition des 50 personnes selon la codification :

1. Déterminez : la population étudiée ; la variable étudiée.
2. Précisez : la nature de la variable ; les modalités de la variable.
3. Construisez le tableau statistique complet associé à cette série.
4. Représentez la distribution des fréquences par un diagramme à bandes.

x_i	*Nombre de personnes*
Sd	04
P	11
Se	14
Su	09
U	12

Exercice 3

Le tableau suivant indique la répartition des familles d'une ville selon le nombre de pièces de leurs appartements.

1. Déterminez la population et la variable étudiée, la nature et les modalités de la variable.
2. Représentez la distribution par diagramme circulaire.
3. Représentez la fonction de répartition. Combien d'appartement de cette ville sont composées d'au moins 3 pièces ? au plus 4 pièces.

Nombre de pièces de l'appartement	*Nombre de familles*
1	25125
2	46290
3	75453
4	61767
5	91365

Exercice 4

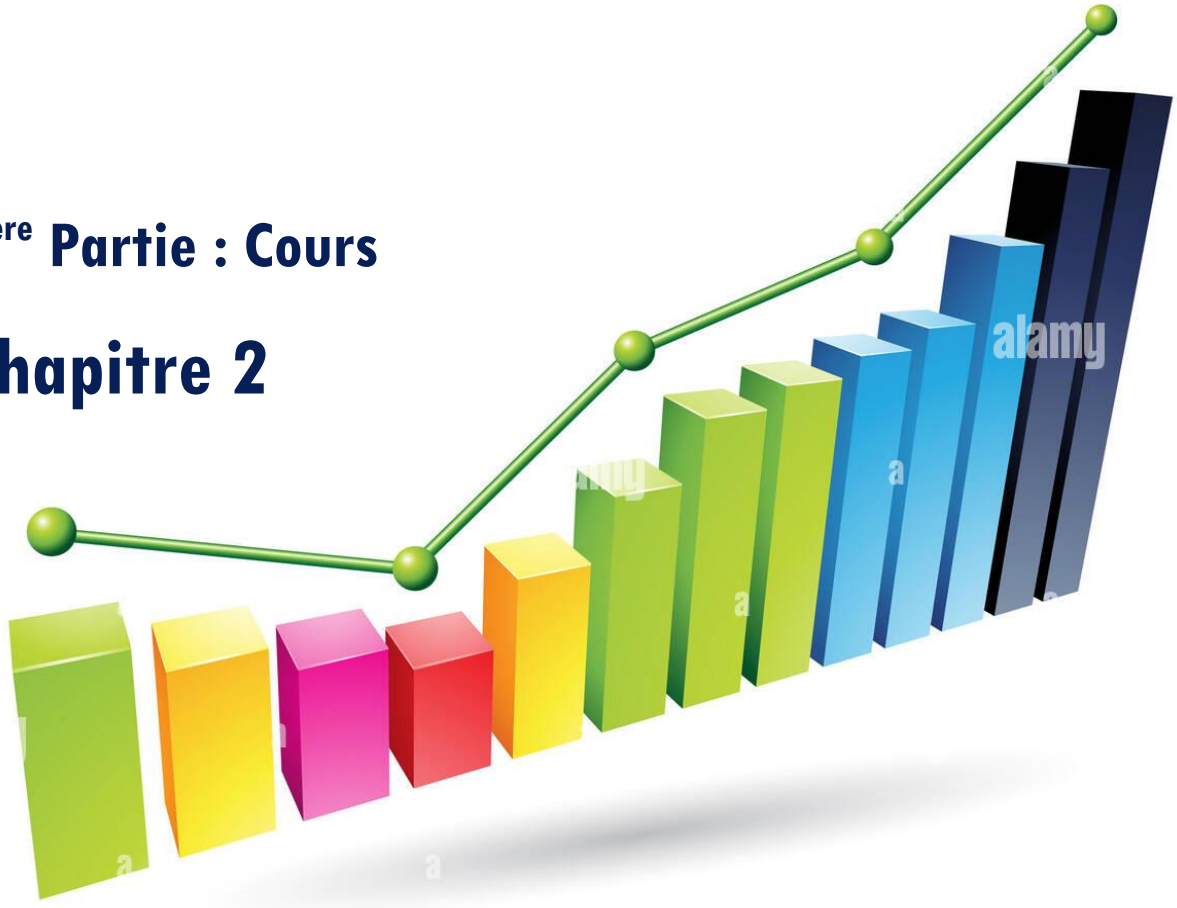
On a effectué avec les mêmes athlètes un sondage sur le temps d'entrainement par semaine, exprimé en heures, Les résultats sont donnés ci-dessous :

Temps d'entrainement	*Nombre des athlètes* n_i
[0 ,10 [	13
[10 , 15 [	36
[15 , 20 [	05
[20 , 30 [	46

1. Déterminez la population et la variable étudiée, la nature et les modalités de la variable.
2. Construisez le tableau statistique des effectifs associé à cette série statistique.
3. Dessinez l'histogramme de la distribution.
4. Dessiner le polygone des effectifs.
5. Représenter la courbe cumulative de la série statistique.

1ère Partie : Cours

Chapitre 2



Statistique descriptive univariée

Dans ce chapitre nous allons associer à chaque série statistique un ensemble de nombres réels qui sont appelés paramètres, d'une part des nombres qui se positionnent au milieu des valeurs dans un rangement croissant, d'autre part d'autres des nombres qui nous donnent la dispersion au tour d'un paramètre appelé moyenne arithmétique.

Ainsi on considère deux types de paramètres.

- Paramètres de positions
- Paramètres de dispersion.

2.1 Paramètres de position

2.1.1 Moyenne arithmétique

- Pour une variable quantitative discrète la moyenne arithmétique notée est donnée par :

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \cdots + n_r x_r}{N} = \sum_{i=1}^{i=r} \frac{n_i x_i}{N} = \sum_{i=1}^{i=r} f_i x_i$$

- Dans le cas d'une variable quantitative continue on a :

$$\bar{x} = \frac{n_1 c_1 + n_2 c_2 + \cdots + n_r c_r}{N} = \sum_{i=1}^{i=r} \frac{n_i c_i}{N} = \sum_{i=1}^{i=r} f_i c_i$$

avec $c_i = \frac{b_{i-1} + b_i}{2}$, le centre de la classe $[b_{i-1}, b_i[$.

Exemple 2.1.1
On reprend l'exemple de la page 10.

Classes	Centres c_i	Effectifs n_i	Fréquences f_i	$n_i c_i$
[20-40[	30	15	0,15	450
[40-60[	50	20	0,20	1000
[60-100[	80	20	0,20	1600
[100-200[	150	45	0,45	6750
Total		100	1,00	9800

$$\Rightarrow \bar{x} = \sum_{i=1}^{i=r} \frac{n_i c_i}{N} = \frac{9800}{100} = 98 \Rightarrow \bar{x} \in [60-100[$$

Remarque 2.1.1 La somme des écarts à la moyenne est nulle, c'est-à-dire

$$\sum_{i=1}^{i=r} n_i (x_i - \bar{x}) = 0$$

En effet

$$\sum_{i=1}^{i=r} n_i (x_i - \bar{x}) = \sum_{i=1}^{i=r} (n_i x_i - n_i \bar{x}) = \sum_{i=1}^{i=r} n_i x_i - \bar{x} \sum_{i=1}^{i=r} n_i = \bar{x}N - \bar{x}N = 0$$

2.1.2 Moyenne géométrique

Notée G et donnée par

$$G = \sqrt[N]{x_1^{n_1} . x_2^{n_2} \cdots x_r^{n_r}}$$

qui peut être exprimée en fonction des fréquences f_i :

$$G = \sqrt[N]{x_1^{n_1} . x_2^{n_2} \cdots x_r^{n_r}} = x_1^{\frac{n_1}{N}} . x_2^{\frac{n_2}{N}} \cdots x_r^{\frac{n_r}{N}} = x_1^{f_1} . x_2^{f_2} \cdots x_r^{f_r}$$

Remarque 2.1.2 La moyenne géométrique G vérifie

$$\ln G = \frac{1}{N} \sum_{i=1}^{i=r} n_i \ln x_i .$$

2.1.3 La moyenne harmonique

Notée H et donnée par

$$H = \frac{1}{\frac{1}{N}\sum_{i=1}^{i=r}\left(\frac{n_i}{x_i}\right)} = \frac{1}{\sum_{i=1}^{i=r}\left(\frac{f_i}{x_i}\right)} = \frac{N}{\sum_{i=1}^{i=r}\left(\frac{n_i}{x_i}\right)}$$

2.1.4 La moyenne quadratique

Notée Q et donnée par $Q = \sqrt{\frac{1}{N}\sum_{i=1}^{i=r} n_i {x_i}^2}$

Remarque 2.1.3 $\bar{x}$, G, H, Q vérifient

$$x_{\min} \le H \le G \le \bar{x} \le Q \le x_{\max}$$

Exemple 2.1.2

Le calcul de $\bar{x}$, G, H, Q de la série : 2, 5, 11, 18.

$$\Rightarrow \bar{x} = \frac{2+5+11+18}{4} = 9,\ G = \sqrt[4]{2\times5\times11\times18} = 6{,}67,$$

$$Q = \sqrt{\frac{1}{4}\left(2^2+5^2+11^2+18^2\right)} = 10{,}88 \text{ et } H = \frac{4}{\frac{1}{2}+\frac{1}{5}+\frac{1}{11}+\frac{1}{18}} = 4{,}72$$

Et on vérifie que

$$x_{\min} = 2 \le H = 4{,}72 \le G = 6{,}67 \le \bar{x} = 9 \le Q = 10{,}88 \le x_{\max} = 18$$

Remarque 2.1.4 Avec deux valeurs x_1, x_2 distinctes ($n_1 = n_2 = 1$) on a :

$$G = G(x_1, x_2) = G(\bar{x}, H)$$

En effet

$$H = \frac{1}{\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)} = \frac{2}{\left(\frac{x_1+x_2}{x_1x_2}\right)} = \frac{x_1x_2}{\left(\frac{x_1+x_2}{2}\right)}$$

$$\Rightarrow G^2 = \bar{x}.H \Rightarrow G = \sqrt{\bar{x}.H} \Rightarrow G(x_1, x_2) = G(\bar{x}, H)$$

Remarque 2.1.5 Avec deux valeurs x_1, x_2 distinctes ($n_1 = n_2 = 1$) si on pose :

$$y_1 = (H - x_1) \quad et \quad y_2 = (H - x_2)$$

on a

$$H = H(x_1, x_2) = H(y_1, y_2)$$

2.1.5 Le mode

C'est la valeur de la variable ayant le plus grand effectif (ou la fréquence la plus élevée). On note le mode M_o.

- **Cas quantitatif discret**

Exemple 2.1.3

On considère les notes obtenues en statistique par un groupe de 20 étudiants :

7, 13, 5, 15, 12, 9, 7, 8, 14, 16, 13, 6, 13, 10, 13, 12, 10, 7, 12, 13.

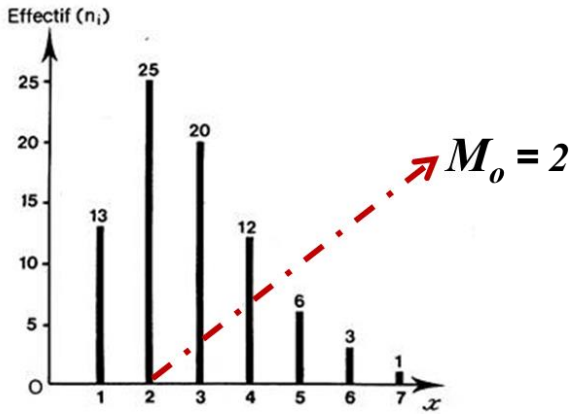
$\Rightarrow$

x_i	5	6	7	8	9	10	12	13	14	15	16
n_i	1	1	3	1	1	2	3	5	1	1	1

Le mode de cette série est $M_o = 13$, valeur qui apparaît cinq fois.
L'interprétation : Est que **la note la plus fréquente est 13**.

Remarque 2.1.6 Graphiquement, dans un diagramme en bâton le mode correspond à **l'abscisse du bâton le plus élevé**.

C'est-à-dire :



- **Cas quantitatif continu**

Dans ce cas, on parle plutôt de **classe modale.** On a deux cas.

1er cas : Cas d'amplitudes égales

Si les classes sont de même amplitude a_i (ie $a_i = a_j \; \forall \; i \neq j$), la classe modale est la classe d'effectif n_i le plus élevé, soit $[b_{i-1}, b_i[$, avec le mode $M_o \in [b_{i-1}, b_i[$ alors :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right) \text{ avec}$$

b_{i-1} : la borne inférieure de la classe modale.

b_i : la borne supérieure de la classe modale.

a_i : l'amplitude de la classe modale.

$m_1 = n_i - n_{i-1}$ et $m_2 = n_i - n_{i+1}$.

Exemple 2.1.4

Soit la distribution de la population de 20 ménages selon le revenu (en centaines de DA) des deux parents

Classes en *CDA*	a_i	n_i	f_i
[200-300[	100	40	0,20
[300-400[	100	60	0,30
[400-500[	100	30	0,15
[500-600[	100	50	0,25
[600-700[	100	20	0,10
Total		200	1,00

Comme $n_2 = 60$ est le plus grand effectif, donc la classe modale est [300 - 400[. Le mode est calculé par :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right) = 300 + 100\left(\frac{60 - 40}{(60 - 40) + (60 - 30)}\right)$$

$$\Rightarrow M_O = 340 \ \ CDA$$

Interprétation : On dit que le salaire le plus fréquent est de 340 CDA.

2eme cas : Cas d'amplitudes inégales

Si les classes sont d'amplitude inégales a_i (ie $a_i \neq a_j$), **la classe modale** est la classe d'effectif corrigé n_i^c le plus élevé (ou encore la fréquence corrigée f_i^c la plus élevée), soit $[b_{i-1}, b_i[$, avec le Mode $M_o \in [b_{i-1}, b_i[$ est tel que :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right) \text{ avec}$$

b_{i-1} : la borne inférieure de la classe modale.

b_i : la borne supérieure de la classe modale.

a_i : l'amplitude de la classe modale.

$m_1 = h_i - h_{i-1} = n_i^c - n_{i-1}^c$ et $m_2 = h_i - h_{i+1} = n_i^c - n_{i+1}^c$.

Où h_i, h_{i-1} et h_{i+1} sont les effectifs corrigés.

Remarque 2.1.7 Dans les 2 cas on peut calculer le mode M_o en utilisant les fréquences à la place des effectifs, en prenant

1. $m_1 = f_i - f_{i-1}$ et $m_2 = f_i - f_{i+1}$ si $(a_i = a_j \ \forall \ i \neq j)$
2. $m_1 = f_i^c - f_{i-1}^c$ et $m_2 = f_i^c - f_{i+1}^c$ si $(a_i \neq a_j)$.

Exemple 2.1.5

Soit la répartition de 100 personnes selon leur âge, on prend $a^* = 100$

Classes	a_i	n_i	d_i	n_i^c
[20 , 30[	10	20	2,00	200
[30 , 40[	10	25	2,50	250
[40 , 60[	20	35	1,75	175
[60 , 80[	20	20	1,00	100

Comme $n_2^c = 250$ est le plus grand effectif corrigé, donc la classe modale est $[30, 40[$. Le mode est calculé par :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right) = 30 + 10\left(\frac{250 - 200}{(250 - 200) + (250 - 175)}\right)$$

$$\Rightarrow M_O = 34.$$

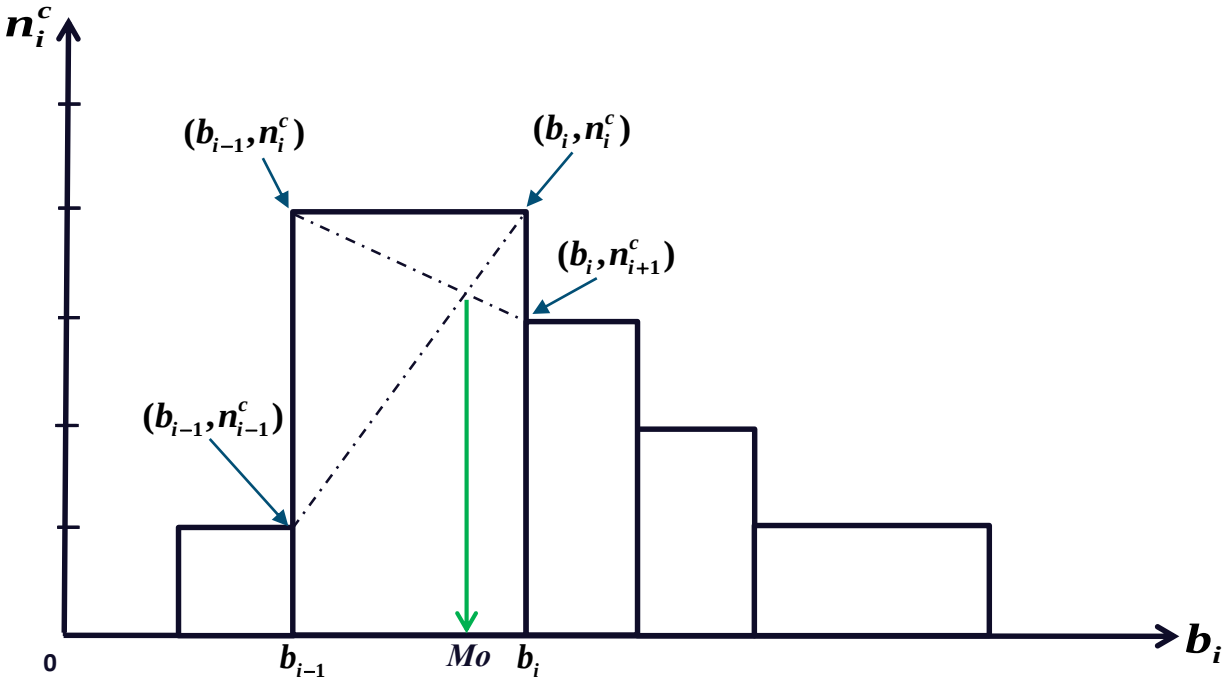
Interprétation : L'âge le plus fréquent est de 34 ans.

Détermination graphique du mode

Si **la classe modale** est $[b_{i-1}, b_i[$, avec le mode $M_o \in [b_{i-1}, b_i[$ alors :

$$M_O = b_{i-1} + a_i\left(\frac{m_1}{m_1 + m_2}\right)$$

Et graphiquement le mode $M_o \in [b_{i-1}, b_i[$ sur un histogramme est le point d'intersection des deux segments $[(b_{i-1}, h_i); (b_i, h_{i+1})]$ et $[(b_{i-1}, h_{i-1}); (b_i, h_i)]$, voir figure suivante :



2.1.6 La médiane

Pour une série statistique rangée par ordre croissant la médiane $Mé$ est la valeur de la variable qui partage la population en deux groupes d'effectifs égaux.

- **Cas quantitatif discret**

Pour une série statistique rangée par ordre croissant c'est-à-dire :

$v_1 \leq v_2 \leq \cdots \leq v_N$ la médiane $Mé$ est la valeur du milieu qui dépendra de l'effectif total N.

1. Si N est impair ($N = 2k+1$), alors $Mé = v_{k+1}$.

2. Si N est pair ($N = 2k$), alors $Mé = \frac{v_k + v_{k+1}}{2}$.

Exemples 2.1.6

1. Soit la répartition de 9 ménages selon le nombre d'enfants

x_i	0	1	2	3	4
n_i	2	2	1	3	1

$\Rightarrow$

Nombre d'enfant par ménage v_i	0	0	1	1	2	3	3	3	4
(ordre croissant) des individus	1	2	3	4	5	6	7	8	9
	4 observations				v_5	4 observation			

On a $N = 9 = 2\times4+1 \Rightarrow$ $Mé = v_{k+1} = v_5 = 2.$

2. Soit la répartition de 10 ménages selon le nombre d'enfants

x_i	0	1	2	3	4
n_i	2	2	1	3	2

On a $N = 10 = 2\times5$ (pair) $\Rightarrow Mé = \dfrac{v_k + v_{k+1}}{2} = \dfrac{2+3}{2} = 2{,}5.$

3. Une répartition avec effectif total grand (voir le tableau ci-dessous avec $N = 50$)

x_i	n_i	N_i	F_i
60	3	3	0,06
65	13	16	0,32
70	9	25	0,50
75	18	43	0,86
85	7	50	1,00
Total	*50*		

L'effectif total est $N = 50 = 2\times25$ qui est un nombre pair donc :

$$Me = \frac{v_{25} + v_{26}}{2}$$

Comme $N_2 = 16$, $N_3 = 25$ et $N_4 = 43$, alors $v_{25} = x_3$ et $v_{26} = x_4$

$$\Rightarrow Me = \frac{x_3 + x_4}{2} = \frac{70+75}{2} = 72{,}5$$

- **Cas quantitatif continu**

On suit les étapes suivantes :

1. Détermination de la classe médiane $[b_{i-1}, b_i[$, En cherchant la classe qui contient l'individu d'ordre $k+1$ (resp k) si $N = 2k+1$ (resp $N = 2k$).

2. Par interpolation linéaire, on peut calculer la médiane à l'intérieur de la classe médiane qui est donnée par :

$$Mé = b_{i-1} + a_i \left(\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}} \right) \text{ avec}$$

N_i : l'effectif cumulé croissant de la classe médiane,

N_{i-1} : l'effectif cumulé croissant de la classe avant la classe médiane,

N : l'effectif total.

Remarque 2.1.8 On peut déterminer la médiane de la même manière en utilisant les fréquences cumulées croissantes.
Et on aura la formule :

$$Mé = b_{i-1} + a_i\left(\frac{0{,}5 - F_{i-1}}{F_i - F_{i-1}}\right) \text{ avec}$$

F_i : la fréquence cumulée croissante de la classe médiane,

F_{i-1} : la fréquence cumulée croissante de la classe qui précède la classe médiane et N est l'effectif total.

Exemple 2.1.7 On reprend l'exemple de la page 16.

$[b_{i-1}, b_i[$	n_i	N_i	f_i	F_i
[5 , 10[	11	11	0 ,11	0,11
[10 , 15[	10	21	0 ,10	0,21
[15, 20[	15	36	0 ,15	0,36
[20, 30[	20	56	0 ,20	0,56
[30 , 40[	18	74	0 ,18	0,74
[40 , 60[	16	90	0 ,16	0,90
[60, 80[	10	100	0 ,10	1,00

L'effectif total est $N = 100 = 2\times50$ qui est un nombre pair, donc
comme $N_3 = 36$ et $N_4 = 56$ alors la classe médiane est $[20, 30[$ et on aura :

$$Mé = b_{i-1} + a_i\left(\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}}\right)$$

$$\Rightarrow Mé = 20 + 10\left(\frac{50-36}{56-36}\right) = 20 + 10\left(\frac{14}{20}\right) \Rightarrow Mé = 27 \quad ans$$

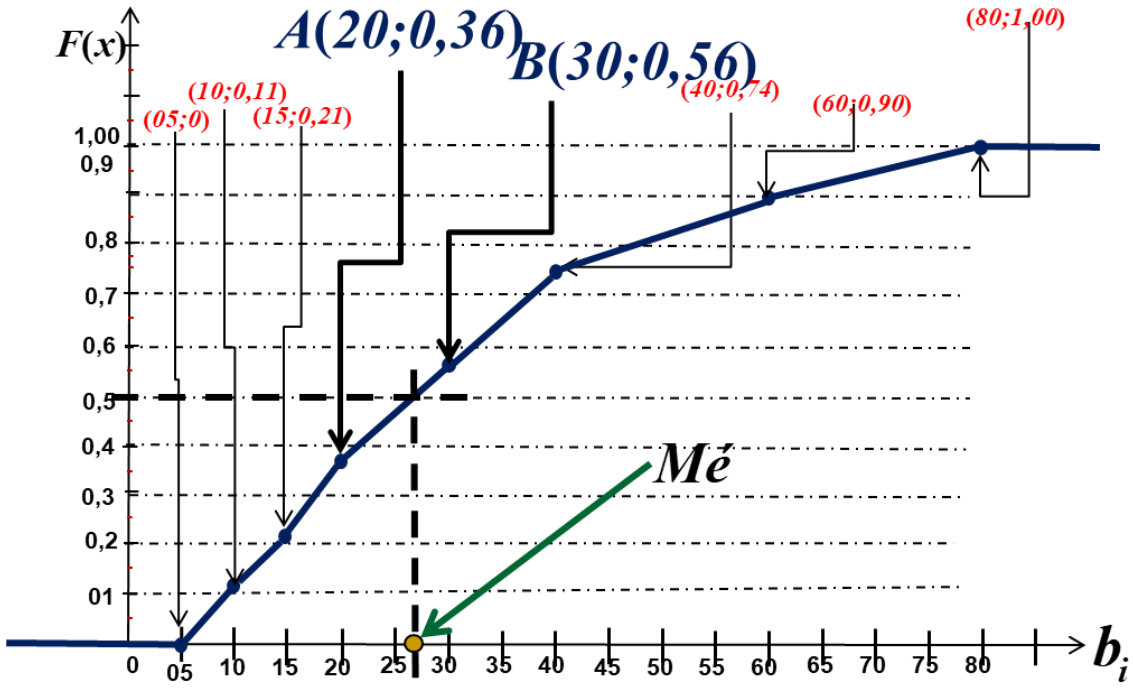
Remarque 2.1.9 La médiane peut être définie comme l'inverse de la fonction de répartition pour la valeur $x = 0{,}5$ c'est à dire $Mé = F^{-1}(0{,}5)$.
On dit que l'ordre de la médiane est $p = F(Mé) = 0{,}5$.

Et on peut calculer la médiane graphiquement à partir de la courbe cumulative.

Exemple 2.1.8 On reprend l'exemple de la page 16 sur la répartition des 100 individus selon leur âge.

$[b_{i-1}, b_i[$	n_i	N_i	f_i	F_i
$[5, 10[$	11	11	0,11	0,11
$[10, 15[$	10	21	0,10	0,21
$[15, 20[$	15	36	0,15	0,36
$[20, 30[$	20	56	0,20	0,56
$[30, 40[$	18	74	0,18	0,74
$[40, 60[$	16	90	0,16	0,90
$[60, 80[$	10	100	0,10	1,00

La classe médiane est [20, 30[et la médiane $Mé$ est l'abscisse d'ordre $p = F(Mé) = 0{,}5$ c'est à dire $Mé = F^{-1}(0{,}5)$, donc



Et on peut soit avoir directement une valeur approchée de la médiane, ou faire une interpolation linéaire entre les deux point A, B. Ainsi l'équation de la droite (AB) est de la forme $y = mx + p$ tel que :

$$m = \frac{y_B - y_A}{x_B - x_A} = \frac{F_i - F_{i-1}}{b_i - b_{i-1}} = \frac{F(Mé) - F_{i-1}}{Mé - b_{i-1}}$$

$$\Rightarrow Mé = b_{i-1} + a_i\left(\frac{F(Mé) - F_{i-1}}{F_i - F_{i-1}}\right) = 20 + 10\left(\frac{0{,}50 - 0{,}36}{0{,}56 - 0{,}36}\right)$$

$$\Rightarrow Mé = 27 \quad ans$$

2.1.7 Les Quartiles

Les quartiles Q_1, Q_2, Q_3 divisent une série statistique en 4 parties d'effectifs égaux : 25% des valeurs sont $\leq Q_1$, 25% comprises entre Q_1 et Q_2; 25% entre Q_2 et Q_3 , et 25% supérieures à Q_3.

Remarque 2.1.10 Q_1 , Q_2 , Q_3 sont respectivement l'abscisse des points d'ordonnées 0,25 ; 0,5 ; 0,75 sur la courbe cumulative croissante. Q_2 est égal à la médiane.

C'est-à-dire :

1. L'ordre de Q_1 est $p = F(Q_1) = 0{,}25$.
2. L'ordre de Q_2 est $p = F(Q_2) = F(Mé) = 0{,}50$.
3. L'ordre de Q_3 est $p = F(Q_3) = 0{,}75$.

- **Calcul des quartiles (Cas quantitatif discret)**

On n'a p l'ordre du quartile Q_i , avec $i = 1, 2, 3$ alors :

- Si $(N \times p)$ est un nombre entier, on a $Q_i = \frac{v_{(N\times p)} + v_{(N\times p)+1}}{2}$.

- Si $(N \times p)$ n'est pas un nombre entier, on a $Q_i = v_{\lceil N\times p \rceil}$.

Où $\lceil N \times p \rceil$ représente le plus petit nombre entier supérieur ou égal à $N \times p$ (qui est appelée **partie entière avec excès**).

Exemples 2.1.9

1. Soit la répartition de 12 ménages selon le nombre d'enfants

x_i	0	1	2	3	4
n_i	2	2	1	5	2

Le premier quartile Q_1 : Comme $(N \times p) = 12\times 0{,}25 = 3$ est un nombre entier, on a :

$$Q_1 = \frac{v_{(N\times p)} + v_{(N\times p)+1}}{2} = \frac{v_3 + v_4}{2} = \frac{1+1}{2} \Rightarrow Q_1 = 1$$

Le Deuxième quartile $Mé = Q_2$: Comme $(N \times p) = 12\times 0{,}50 = 6$ est un nombre entier, on a :

$$Q_2 = Mé = \frac{v_{(N\times p)} + v_{(N\times p)+1}}{2} = \frac{v_6 + v_7}{2} = \frac{3+3}{2} \Rightarrow Q_2 = 3$$

Le troisième quartile Q_3 : Comme $(N \times p) = 12 \times 0{,}75 = 9$ est un nombre entier, on a :

$$Q_3 = \frac{v_{(N\times p)} + v_{(N\times p)+1}}{2} = \frac{v_9 + v_{10}}{2} = \frac{3+3}{2} \Rightarrow Q_3 = 3$$

2. Soit la répartition de 9 ménages selon le nombre d'enfants

x_i	0	1	2	3	4
n_i	2	2	1	2	2

Le premier quartile Q_1 : Comme $(N \times p) = 9 \times 0{,}25 = 2{,}25$ n'est pas un nombre entier, on a : $Q_1 = v_{\lceil 2{,}25 \rceil} = v_3 = 1.$

Le Deuxième quartile $Mé = Q_2$: Comme $(N \times p) = 9 \times 0{,}50 = 4{,}50$ n'est pas un nombre entier, on a : $Q_2 = v_{\lceil 4{,}50 \rceil} = v_5 = 2.$

Le troisième quartile Q_3 : Comme $(N \times p) = 9 \times 0{,}75 = = 6{,}75$ n'est pas un nombre entier, on a : $Q_3 = v_{\lceil 6{,}75 \rceil} = v_7 = 3.$

3. Une répartition avec effectif total grand (voir le tableau ci-dessous avec $N = 50$)

x_i	n_i	N_i	F_i
60	3	3	0,06
65	13	16	0,32
70	9	25	0,50
75	18	43	0,86
85	7	50	1,00
Total	*50*		

Le premier quartile Q_1 $N \times p = 50 \times \frac{1}{4} = 12{,}5$ qui n'est pas un nombre entier, donc :

$Q_1 = v_{\lceil 12{,}5 \rceil} = v_{13} = x_2 = 65$ car $N_1 = 3$ et $N_2 = 16.$

Le troisième quartile Q_3 $N \times p = 50 \times \frac{3}{4} = 37{,}5$ qui n'est pas un nombre entier, donc :

$Q_3 = v_{\lceil 37{,}5 \rceil} = v_{38} = x_4 = 75$ car $N_3 = 25$ et $N_4 = 43.$

- **Calcul des quartiles (Cas quantitatif continu)**

Pour le calcul de Q_1, Q_2, Q_3 : On suit les étapes suivantes

1. **Détermination de la classe** $[b_{i-1}, b_i[$ de $Q_1 \in [b_{i-1}, b_i]$, En cherchant la classe qui contient l'individu d'ordre $\lceil N \times p \rceil = \lceil N/4 \rceil$.

2. Si N_i : l'effectif cumulé croissant de la classe de Q_1 ,

N_{i-1} : l'effectif cumulé croissant de la classe qui précède la classe de Q_1.

N : l'effectif total.

F_i : la fréquence cumulée croissante de la classe de Q_1 ,

F_{i-1} : la fréquence cumulée croissante de la classe qui précède la classe de Q_1.

Alors on aura

$$Q_1 = b_{i-1} + a_i\left(\frac{N/4 - N_{i-1}}{N_i - N_{i-1}}\right) = b_{i-1} + a_i\left(\frac{0{,}25 - F_{i-1}}{F_i - F_{i-1}}\right)$$

N_i : l'effectif cumulé croissant de la classe médiane,

N_{i-1} : l'effectif cumulé croissant de la classe avant la classe médiane,

N : l'effectif total.

Remarque 2.1.11 Le calcul de Q_2 et Q_3 se fait de la même manière tel que

$$Q_2 = Mé = b_{i-1} + a_i\left(\frac{N/2 - N_{i-1}}{N_i - N_{i-1}}\right) = b_{i-1} + a_i\left(\frac{0{,}5 - F_{i-1}}{F_i - F_{i-1}}\right)$$

$$Q_3 = b_{i-1} + a_i\left(\frac{N(3/4) - N_{i-1}}{N_i - N_{i-1}}\right) = b_{i-1} + a_i\left(\frac{0{,}75 - F_{i-1}}{F_i - F_{i-1}}\right)$$

Remarque 2.1.12 L'unité des paramètres de position est l'unité de la variable.

2.2 Paramètres de dispersion

Remarque 2.2.1 Les quartiles déjà vu comme paramètres de positions peuvent être considérés comme paramètres de dispersion.

2.2.1 L'étendue

L'étendue noté E est simplement la différence entre la plus grande et la plus petite valeur observée $E = x_{max} - x_{min}$.

2.2.2 L'écart interquartile

Il s'agit de l'écart entre le premier et le dernier quartile. C'est-à-dire $IQ = Q_3 - Q_1$.

Remarque 2.2.2 L'écart interquartile mesure l'étendue des 50% de valeurs situées au milieu d'une série de données classées.

Exemple 2.2.1 On reprend l'exemple de la page 16 sur la répartition des 100 individus selon leur âge.

$[b_{i-1}, b_i[$	n_i	N_i	f_i	F_i
$[5, 10[$	11	11	0,11	0,11
$[10, 15[$	10	21	0,10	0,21
$[15, 20[$	15	36	0,15	0,36
$[20, 30[$	20	56	0,20	0,56
$[30, 40[$	18	74	0,18	0,74
$[40, 60[$	16	90	0,16	0,90
$[60, 80[$	10	100	0,10	1,00

Calculons les quartiles Q_1, Q_3 et l' écart interquartile.

On a : $\left\lceil \frac{N}{4} \right\rceil = 25$, $\left\lceil \frac{3N}{4} \right\rceil = 75$, donc la classe de Q_1 est $[15, 20[$, celle de Q_3 est $[40, 60[$:

$$\Rightarrow Q_1 = b_{i-1} + a_i\left(\frac{0,25 - F_{i-1}}{F_i - F_{i-1}}\right) \Rightarrow Q_1 = 15 + 5\left(\frac{0,25 - 0,21}{0,36 - 0,21}\right) = 16,33 \ \ ans$$

Ce qui signifie que 25 % des individus sont âgés de moins de 16 ans et 4 mois ($0,33 \times 12 = 3,96 \approx 4$). Et pour Q_3 on aura :

$$Q_3 = b_{i-1} + a_i\left(\frac{0,75 - F_{i-1}}{F_i - F_{i-1}}\right) \Rightarrow Q_3 = 40 + 20\left(\frac{0,75 - 0,74}{0,90 - 0,74}\right) = 41,25 \ \ ans$$

Ce qui signifie que 75 % des individus sont âgés de moins de 41 ans et 3 mois ($0,25 \times 12 = 3$). Donc l'écart interquartile est :

$$\Rightarrow IQ = Q_3 - Q_1 = 24,92 \ \ ans$$

Ce qui signifie que la différence d'âge entre Q_1 et Q_3 est de 24 ans, 11 mois et 12 jours ($0,92 \times 12 = 11,04$ et $0,4 \times 30 = 12$).

Remarque 2.2.3 Si $N \times p = N_i$, alors le quantile $x_p = b_i$ malgré que $b_i \notin [b_{i-1}, b_i[$ et la classe du quantile est $[b_{i-1}, b_i[$.

Exemple 2.2.2 Soit la répartition de 100 personnes selon leur âge

Classes	n_i	N_i	f_i	F_i
[20 , 30[	25	25	0,25	0,25
[30 , 40[	20	45	0,20	0,45
[40 , 60[	35	80	0,35	0,80
[60 , 80[	20	100	1,00	1,00

Calcul du 1er quartile Q_1 :

On a l'ordre du 1er quartile Q_1 est $p = 0{,}25$.

Comme $\lceil N \times p \rceil = \lceil 100 \times 0{,}25 \rceil = \lceil 25 \rceil = 25$ et $N_1 = 25$, alors :

La casse de Q_1 est $[20, 30[$, c'est à dire $Q_1 \in [20, 30]$. Donc :

$$Q_1 = b_0 + a_1\left(\frac{N(1/4) - N_0}{N_1 - N_0}\right) = 20 + 10\left(\frac{25-0}{25-0}\right) = 30 \in [20,30]$$

Remarque 2.2.4 On peut obtenir des valeurs approximatives des quartiles graphiquement à partir de l**a courbe cumulative.**

2.2.3 La variance

La variance d'une variable X notée $V(x)$ est la somme des carrés des écarts à la moyenne divisée par le nombre d'observations (Effectif total N).

- **Le cas d'une variable discrète**

$$V(x) = \frac{1}{N}\sum_{i=1}^{i=r} n_i (x_i - \bar{x})^2 = \sum_{i=1}^{i=r} f_i (x_i - \bar{x})^2$$

- **Le cas d'une variable continue**

$$V(x) = \frac{1}{N}\sum_{i=1}^{i=r} n_i (c_i - \bar{x})^2 = \sum_{i=1}^{i=r} f_i (c_i - \bar{x})^2$$

avec $c_i = \frac{b_{i-1} + b_i}{2}$, le centre de la classe $[b_{i-1}, b_i[$.

Remarque 2.2.5 La variance peut être écrite sous une autre forme dite « formule développée »

- **La formule développée de la variance pour le cas d'une variable discrète**

$$V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=r} n_i x_i^{\,2}\right) - \bar{x}^2 = \left(\sum_{i=1}^{i=r} f_i x_i^{\,2}\right) - \bar{x}^2$$

- **La formule développée de la variance pour le cas d'une variable continue**

$$V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=r} n_i c_i^{\,2}\right) - \bar{x}^2 = \left(\sum_{i=1}^{i=r} f_i c_i^{\,2}\right) - \bar{x}^2$$

Preuve de la formule développée

On a :

$$V(x)=\frac{1}{N}\sum_{i=1}^{i=r}n_i\left(x_i-\bar{x}\right)^2=\frac{1}{N}\sum_{i=1}^{i=r}n_i\left({x_i}^2-2x_i\bar{x}+\bar{x}^2\right)$$

$$\Rightarrow V(x)=\frac{1}{N}\sum_{i=1}^{i=r}n_i{x_i}^2-2\frac{\bar{x}}{N}\sum_{i=1}^{i=r}n_ix_i+\frac{\bar{x}^2}{N}\sum_{i=1}^{i=r}n_i$$

$$\Rightarrow V(x)=\left(\frac{1}{N}\sum_{i=1}^{i=r}n_i{x_i}^2\right)-2\bar{x}.\bar{x}+\bar{x}^2$$

Remarque 2.2.6

1. Cette formule développée de la variance est plus facile à retenir et plus rapide à calculer.

2. La variance est exprimée dans le carré de l'unité de la variable. Par exemple, la variance de la variable âge est exprimée en « années au carré (année²)» car :

2.2.4 L'écart type

On appelle écart type que l'on le note par $\sigma(x)$, la racine carrée de la variance :

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

Remarque 2.2.7

1. L'écart type est exprimé dans la même unité de mesure que la variable.

2. Il est utilisé comme un indicateur de la dispersion de la série statistique, de façon que dans un rangement croissant la moyenne $\bar{x}$ partage la population en deux partie tel que les individus ayant la valeur de la variable inférieur à $\bar{x}$ auront approximativement $\bar{x}-\sigma(x)$, les autres $(X>\bar{x})$ auront $\bar{x}+\sigma(x)$.

3. Plus l'écart type est grand, plus la dispersion des observations autour de la moyenne de la variable est forte.

4. Une distribution aura un écart-type proche de 0 si ces valeurs seront ramassées autour de la moyenne.

Exemple 2.2.3 Considérons les notes suivantes en statistique d'un groupe de 20 étudiants :

x_i	n_i	$n_i.x_i$	$n_i.x_i^2$
2	2	4	8
3	2	6	18
7	4	28	196
8	2	16	128
12	3	36	432
17	2	34	578
18	5	90	1620
Total	20	214	2980

Donc $\bar{x} = \sum_{i=1}^{i=r} \frac{n_i x_i}{N} = \frac{214}{20} = 10{,}7$ et $V(x) = \left(\frac{1}{N} \sum_{i=1}^{i=r} n_i {x_i}^2 \right) - \bar{x}^2$

$$\Rightarrow V(x) = \frac{2980}{20} - (10{,}7)^2 \Rightarrow V(x) = 149 - 114{,}49 = 34{,}51$$

$$\Rightarrow \sigma(x) = 5{,}87 .$$

Donc, certains étudiants (les bons) auront approximativement la note moyenne $(10{,}7)$ plus $(+)$ $5{,}87$ $(=16{,}57)$ les autres (les mauvais) auront la note moyenne $(10{,}7)$ moins $(-)$ $5{,}87$ $(= 4{,}83)$.

2.2.5 L'écart absolu moyen

Pour une série statistique dont la variable peut prendre les valeurs $x_1, x_2, \ldots, x_r$, les effectifs correspondants étant $n_1, n_2, \ldots, n_r$, l'écart-absolu moyen de la variable X est la moyenne arithmétique des valeurs absolues des écarts à la moyenne arithmétique est défini par :

$$e_{\bar{x}} = \frac{n_1 |x_1 - \bar{x}| + n_2 |x_2 - \bar{x}| + \cdots + n_r |x_r - \bar{x}|}{N} = \frac{1}{N} \sum_{i=1}^{i=r} n_i |x_i - \bar{x}|$$

Remarque 2.2.8

1. L'écart absolu moyen permet de mesurer la dispersion d'une série. Par exemple, si un premier (1) étudiant a eu pour notes $5, 10, 15$ et un deuxième (2) étudiant $9, 10, 11$, ils ont même moyenne alors que clairement leurs cas sont très différents. L'écart absolu moyen du premier étudiant est $10/3 = 3{,}333$, tandis que celui du deuxième étudiant est $2/3 = 0{,}6666$. On lit donc bien cette différence sur l'écart absolu moyen.

2. Plus l'écart absolu moyen par rapport à la moyenne est élevé, plus il y a de valeurs éloignées de la moyenne : l'écart absolu moyen est donc bien un paramètre de dispersion, c'est-à-dire un indicateur de l'étalement des valeurs recueillies.

3. On peut, de manière similaire, définir un écart moyen par rapport à la médiane quand on juge que la médiane est un paramètre de position plus approprié à l'étude d'une série statistique que la moyenne arithmétique.

2.2.6 Le coefficient de variation

L'écart type, et la moyenne, s'exprime dans la même unité que la variable statistique, mais dans certain cas on peut être ramené à comparer les dispersions qui ne sont pas exprimées dans la même unité. Donc le coefficient de variation se calcule comme **le rapport de l'écart type à la moyenne**, et s'exprime en pourcentage.

$$CV(x)=\frac{\sigma(x)}{\bar{x}}$$

Remarque 2.2.9

1. Le coefficient de variation permet de comparer le degré de variation d'un échantillon à un autre, même si les moyennes sont différentes.

2. 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.

3. Plus le coefficient de variation est faible, plus les données statistiques sont regroupées autour de la moyenne et plus il est grand, plus les données sont dispersées.

Exemple 2.2.3 Dans une maternité on a relevé le poids (en *kilogramme*) à la naissance de 47 nouveaux nés. Les données collectées sont résumées dans le tableau suivant :

Classes	n_i	c_i	$n_i c_i$	$n_i c_i^2$
[2,5 ; 3,0 [	08	2,75	22,00	60,50
[3,0 ; 3,5 [	15	3,25	48,75	158,438
[3,5 ; 4,0 [	20	3,75	75,00	281,25
[4,0 ; 4,5 [	04	4,25	17,00	72,25
Total	47		162,75	572,438

Donc $\bar{x}=\sum_{i=1}^{i=4}\frac{n_i c_i}{N}=\frac{162,75}{47}=3,463$ et $V(x)=\left(\frac{1}{N}\sum_{i=1}^{i=r}n_i {x_i}^2\right)-\bar{x}^2$

$$\Rightarrow V(x)=\frac{572438}{47}-(3,463)^2 \Rightarrow V(x)=0,189$$

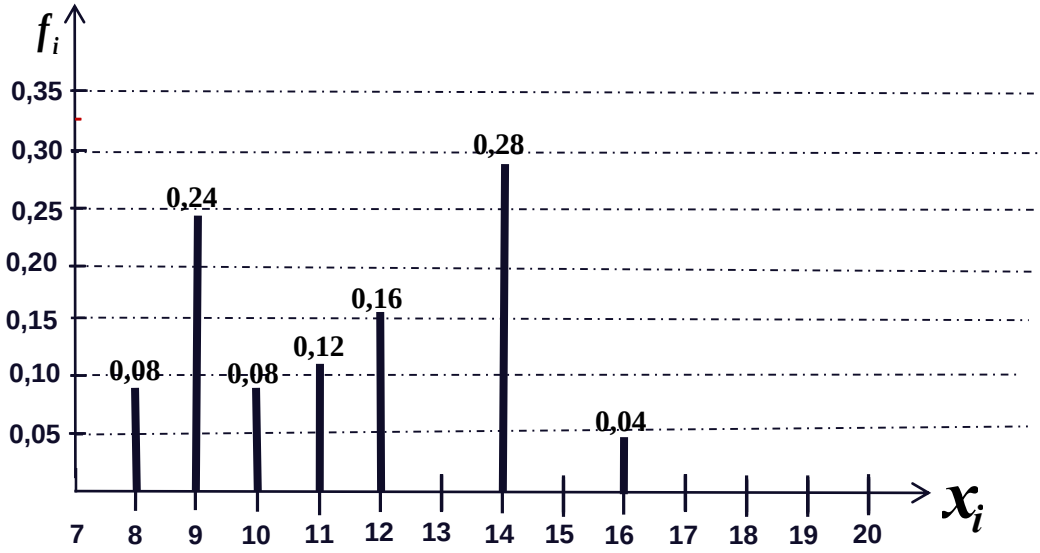
$$\Rightarrow \sigma(x)==\sqrt{V(x)}=0,434 \Rightarrow \quad CV(x)=\frac{\sigma(x)}{\bar{x}}=\frac{0,434}{3,463}=0,125.$$

Donc, Le coefficient de variation étant faible, le poids à la naissance est concentré autour de la moyenne.

2.3 Exercices du chapitre 2

Exercice 1

Le diagramme en bâtons des fréquences ci-dessous donne la répartition des notes obtenues à un contrôle de mathématiques par les élèves d'une classe de terminale.



1. A partir du diagramme en bâtons donnez la valeur du mode de cette série statistique. Expliquez ?
2. Construisez le tableau statistique des fréquences associé à cette série statistique.
3. Donnez la proportion en pourcentage (%) des élèves qui ont une note superieure à 11 $(X > 11)$.
4. Calculez la moyenne de cette série statstique.
5. Sachant que $\sum_{i=1}^{7} n_i x_i = 570$ calculez l'effectif total ainsi que l'effectif de chaque modalité.

Exercice 2

Le tableau ci-dessous donne la répartition des boulangeries d'une ville selon le prix auquel elles vendent la baguette :

Prix (Euro)	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90
Effectif	4	14	26	11	7	12	7	5

1. Calculez le prix moyen d'une baguette.
2. Déterminez le prix médian d'une baguette.
3. Déterminez les premier et troisième quartiles.

4. Calculez l'étendue de la série.
5. Déterminez le prix médian d'une baguette, les premier et troisième quartiles si on remplace le 1er tableau par le tableau suivant :

Prix (Euro)	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90
Effectif	4	14	4	11	7	8	11	5

Exercice 3

On a effectué avec les mêmes athlètes un sondage sur le temps d'entrainement par semaine, exprimé en heures, Les résultats sont donnés ci-dessous :

Temps d'entrainement	F_i
[0 , 10 [	0,04
[10 , 15 [	0,20
[15 , 20 [	F_3
[20 , 25 [	0,82
[25 , 30 [	1,00

1. Déterminer la classe du premier quartile Q_1 sachant que $0{,}30 \leq F_3$.
2. Calculer F_3 sachant que le premier quartile $Q_1 = 16$ heures.
   Dans toute la suite on prend $F_3 = 0{,}45$.
3. Calculer les fréquences de cette série statistique.
4. Déduire la moyenne arithmétique.
5. Calculer l'écart type $\sigma(x)$ de cette série.
6. Sachant que $\sum_{i=1}^{5} n_i c_i^2 = 42600$ calculer l'effectif total ainsi que l'effectif de chaque classe.

Exercice 4

Une enquete statistique chez 1000 commercants porte sur le nombre d'heures d'ouvertures hebdomadaire. On a obtenu les resultats suivants :

Nombre d'heures	*Nombre de commerçants*
[30 , 35 [	50
[35 , 37 [	100
[37, 39 [	200
[39 , 40 [	150
[40 , 41 [	120
[41 , 43 [	n_6
[43 , 45 [	130
[45 , 50 [	n_8

En prenant le nombre moyen d'heures d'ouverture hebdomadaires **40,38.**

1. Determiner les effectifs n_6 et n_8 .

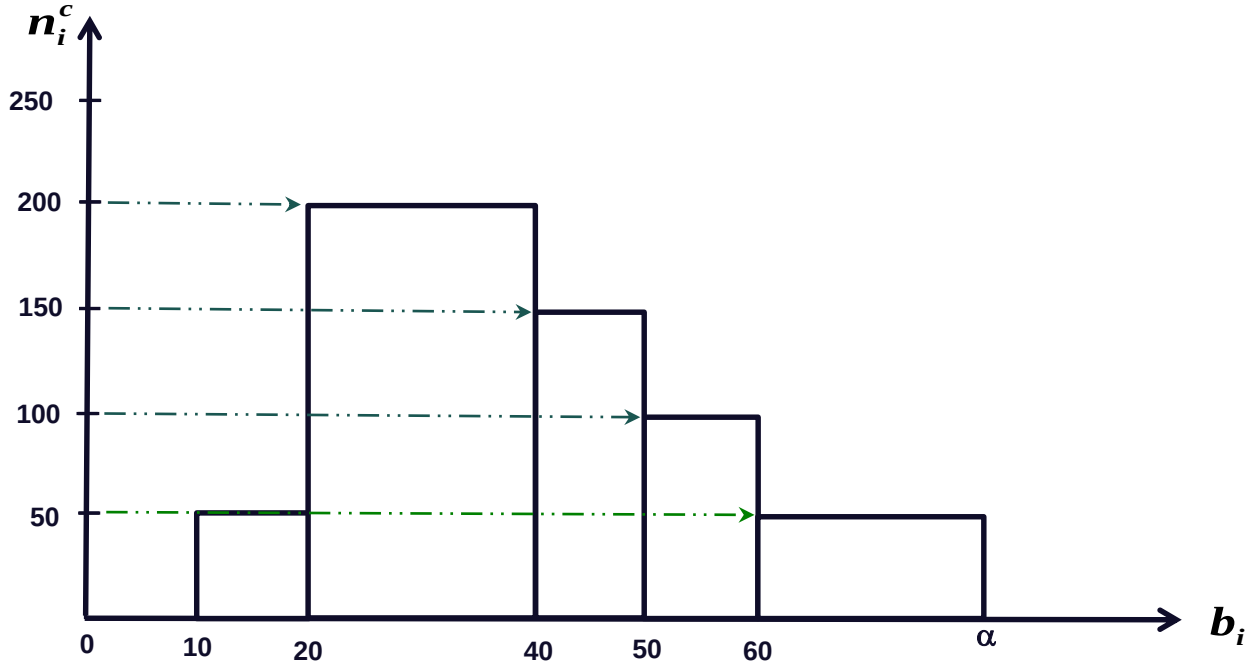
2. Calaculer mode et la mediane de cette distribution (on prend $a^* = 1$) ?

3. Calculer la variance puis l'ecart-type de cette distribution.

4. Calculer les premier et troisième quartiles.

Exercice 5

Dans une gare routière, on évalue le temps d'attente de 80 voyageurs en minutes.
Voici l'histogramme des effectifs de cette variable.



1. Représenter graphiquement le mode M_0 de cette série statistique.

2. Construisez le tableau statistique des effectifs associé à cette série (en fonction de α) avec $a^* = 100$.

3. Calculer α.

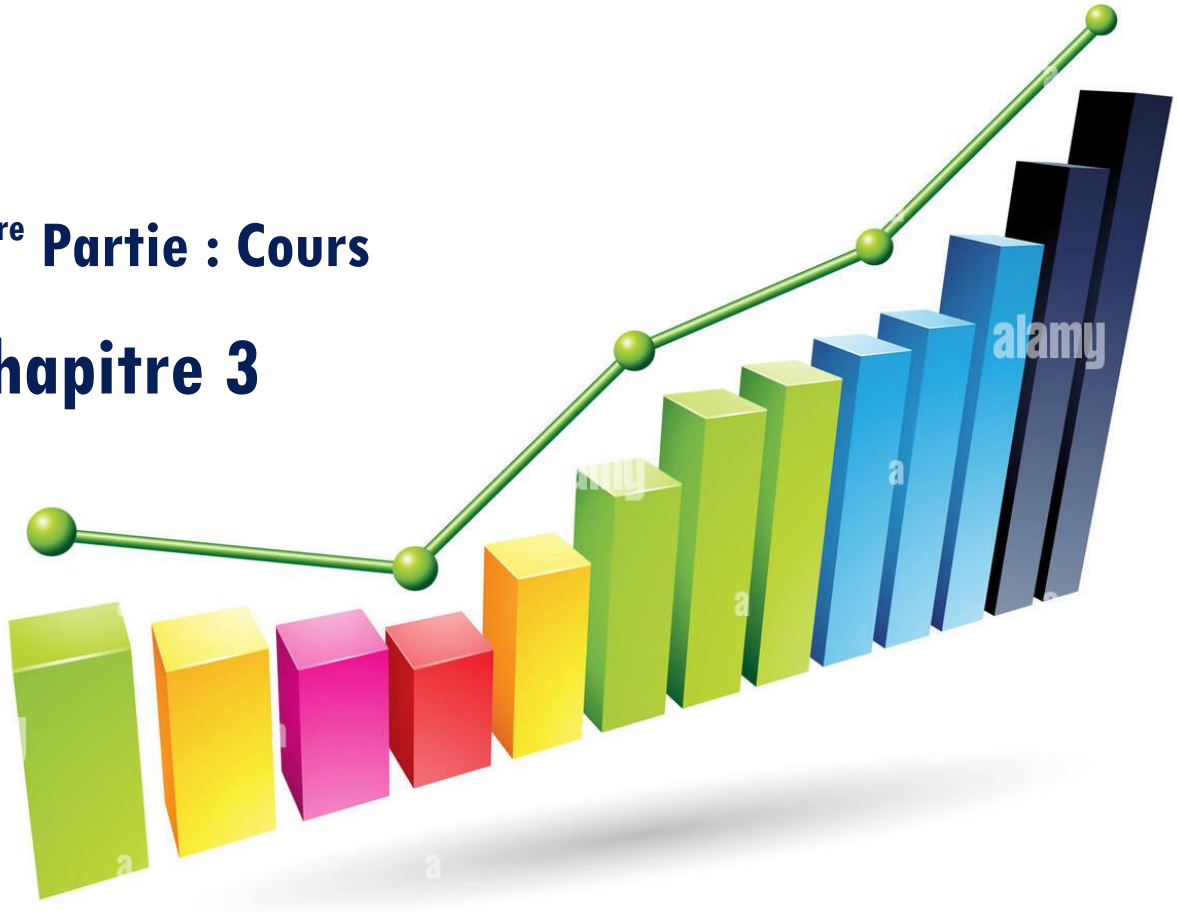
4. Dessiner le polygone des effectifs.

5. Déterminez le mode M_0 de cette série.

6. Calculer l'écart type $\sigma(x)$ de cette série.

1ère Partie : Cours

Chapitre 3



Statistique descriptive bivariée

On s'intéresse à deux variables X et Y. Ces deux variables sont mesurées simultanément sur les mêmes individus. On obtient donc deux mesures. La série statistique est alors une suite de couples des valeurs prises par les deux variables sur chaque individu.
Chacune des deux variables peut être, soit quantitative, soit qualitative. On examine juste le cas des deux variables sont quantitatives.
Pour bien voir ce genre de série statistique on va commencer par un exemple et par la suite on généralise.

Exemple 3.0.1

On considère le tableau suivant, relatif à une population de 100 ménages, tel que :
X = « le nombre d'enfants du ménage » et Y = « le nombre de pièces du logement »

X \ Y	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	Total
$x_1 = 2$	10	8	05	23
$x_2 = 3$	39	5	10	54
$x_3 = 4$	8	5	0	13
$x_4 = 5$	10	0	0	10
Total	67	18	15	100

- La valeur 39 indique que, parmi les 100 ménages observés, il y'a 39 ménages qui ont 3 enfants et qui habitent dans des logements de 3 pièces.

- La valeur 67 indique que, parmi les 100 ménages observés, il y a 67 ménages qui habitent dans des logements de 3 pièces.

- La valeur 54 indique que, parmi les 100 ménages observés, il y a 54 ménages qui ont 3 enfants.

3.1 Distributions et caractéristiques

Soient X et Y deux variables mesurées sur N individus d'une population, Avec les modalités : $M(X) = \{x_1, x_2, \ldots, x_r\}$, $M(Y) = \{y_1, y_2, \ldots, y_k\}$.

3.1.1 Distribution conjointe de X et Y

C'est la liste des $r \times k$ modalités conjointes (x_i, y_j) associées chacune à son effectif n_{ij} ou à sa fréquence f_{ij} . Ce qui donne le tableau des contingents suivant :

X \ Y	y_1	y_2	...	y_j	...	y_k	Total
x_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2.}$
⋮							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i.}$
⋮							
x_r	n_{r1}	n_{r2}		n_{rj}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.k}$	N

- Les effectifs qui sont notés par n_{ij} est le nombre de fois où la modalité x_i de la variable X et la modalité y_j de la variable Y ont été observées simultanément.
- L'effectif $n_{i.}$ appelé **effectif marginal** de la variable X est le nombre total d'observations de la modalité x_i de la variable X.

- L'effectif $n_{.j}$ appelé **effectif marginal de** la variable Y est le nombre total d'observations de la modalité y_j de la variable Y.

3.1.2 Distributions marginales

La distribution marginale de X (respectivement de Y) est la distribution de X (respectivement de Y) sur l'échantillon, calculée à partir de la distribution conjointe.
Ces deux distributions peuvent se présenter sous forme de tableaux statistiques suivants :

Distribution marginale de X

X	*Effectif marginal*
x_1	$n_{1.}$
x_2	$n_{2.}$
⋮	
x_i	$n_{i.}$
⋮	
x_r	$n_{r.}$
Total	N

Distribution marginale de Y

Y	*Effectif marginal*
y_1	$n_{.1}$
y_2	$n_{.2}$
⋮	
y_j	$n_{.j}$
⋮	
y_k	$n_{.k}$
Total	N

Remarque 3.1.1 Pour deux variables X et Y mesurées sur N individus d'une population, la distribution conjointe se donne sous forme de tableaux des contingent des effectifs ou des fréquences comme suit :

Distribution conjointe en effectif de X et Y

X \ Y	y_1	y_2	...	y_j	...	y_k	Total
x_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2.}$
⋮							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i.}$
⋮							
x_r	n_{r1}	n_{r2}		n_{rj}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.k}$	N

Distribution conjointe en fréquence de X et Y

X \ Y	y_1	y_2	...	y_j	...	y_k	Total
x_1	f_{11}	f_{12}		f_{1j}		f_{1k}	$f_{1.}$
x_2	f_{21}	f_{22}		f_{2j}		f_{2k}	$f_{2.}$
⋮							
x_i	f_{i1}	f_{i2}		f_{ij}		f_{ik}	$f_{i.}$
⋮							
x_r	f_{r1}	f_{r2}		f_{rj}		f_{rk}	$f_{r.}$
Total	$f_{.1}$	$f_{.2}$		$f_{.j}$		$f_{.k}$	1

- L'effectif $n_{i.}$ appelé effectif marginal de X est le nombre total d'observations de la modalité x_i de la variable X.

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

- L'effectif $n_{.j}$ appelé effectif marginal de Y est le nombre total d'observations de la modalité y_j de la variable Y.

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

- L'effectif total de la distribution conjointe noté N, peut être obtenu à partir de l'effectif marginal de X ou bien à partir de l'effectif marginal de Y :

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

- La fréquence conjointe, noté f_{ij} est $f_{ij} = \dfrac{n_{ij}}{N}$
- La fréquence $f_{i.}$ appelée **fréquence marginale** de X est le nombre

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

- La fréquence $f_{.j}$ appelée **fréquence marginale** de Y est le nombre

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

$$\text{et } \sum_{i=1}^{i=r} f_{i.} = \sum_{j=1}^{i=k} f_{.j} = \sum_{i=1}^{i=r}\sum_{j=1}^{j=k} f_{ij} = 1$$

- **Les moyennes marginales et les variances marginales**

Distribution marginale de X

X	Effectif marginal
x_1	$n_{1.}$
x_2	$n_{2.}$
⋮	
x_i	$n_{i.}$
⋮	
x_r	$n_{r.}$
Total	N

Distribution marginale de Y

Y	Effectif marginal
y_1	$n_{.1}$
y_2	$n_{.2}$
⋮	
y_j	$n_{.j}$
⋮	
y_k	$n_{.k}$
Total	N

Les moyennes marginales de X et de Y, ainsi que les variances marginales se calculent à partir des distributions marginales par les formules suivantes :

$$\bar{x} = \sum_{i=1}^{i=r} \frac{n_{i.}x_i}{N} = \sum_{i=1}^{i=r} f_{i.}x_i; \qquad \bar{y} = \sum_{j=1}^{j=k} \frac{n_{.j}y_j}{N} = \sum_{j=1}^{j=k} f_{.j}y_j$$

$$V(x) = \frac{1}{N}\sum_{i=1}^{i=r} n_{i.}(x_i - \bar{x})^2 = \sum_{i=1}^{i=r} f_{i.}(x_i - \bar{x})^2;$$

$$V(y) = \frac{1}{N}\sum_{j=1}^{j=k} n_{.j}(y_j - \bar{y})^2 = \sum_{j=1}^{j=k} f_{.j}(y_j - \bar{y})^2$$

Exemple 3.1.1

On considère le tableau des effectifs suivant, relatif à une population de 20 adolescents, tel que : X = « la taille » et Y = « le poids ».

X \ Y	[40,50[	[50,70[	[70,90[	*Total*
[120,140[	2	1	0	3
[140,160[	2	6	0	8
[160,180[	1	3	5	9
Total	5	10	5	20

Donc la distribution conjointe en fréquences est : $f_{ij} = \frac{n_{ij}}{N}$

X \ Y	[40,50[	[50,70[	[70,90[	*Total*
[120,140[	0,10	0,05	0,00	0,15
[140,160[	0,10	0,30	0,00	0,40
[160,180[	0,05	0,15	0,25	0,45
Total	0,25	0,50	0,25	1,00

$\Rightarrow$

Distribution marginale en fréquence de X

X	$f_{i.}$
[120,140[	0,15
[140,160[	0,40
[160,180[	0,45
Total	1,00

et

Distribution marginale en fréquence Y

Y	$f_{.j}$
[40,50[	0,25
[50,70[	0,50
[70,90[	0,25
Total	1,00

De même on obtient à partir du même tableau les distributions marginales des effectifs.

En introduisant les centres des classes pour calculer les moyennes et les variances marginales :

Distribution marginale en effectif de X

X	$n_{i.}$	c_i	$n_{i.}\, c_i$	$n_{i.}\, c_i^2$
[120,140[	3	130	390	50700
[140,160[	8	150	1200	180000
[160,180[	9	170	1530	260100
Total	20		3120	490800

et

Distribution marginale en effectif Y

Y	$n_{.j}$	c_j	$n_{.j}\, c_j$	$n_{j.}\, c_j^2$
[40,50[	5	45	225	10125
[50,70[	10	60	600	36000
[70,90[	5	80	400	32000
Total	20		1225	78125

Donc les moyennes marginales de X et de Y sont :

$$\bar{x} = \frac{3120}{20} = 156 \ \ cm \ \text{ et } \ \bar{y} = \frac{1225}{20} = 61{,}25 \ \ kg$$

les variances marginales de X et de Y sont :

$$V(x) = \left(\frac{1}{N} \sum_{i=1}^{i=3} n_{i.} c_i^{\,2} \right) - \bar{x}^2 = \frac{490800}{20} - (156)^2 = 24540 - 24336 \Rightarrow V(x) = 204.$$

$$\text{et } V(y) = \left(\frac{1}{N} \sum_{j=1}^{j=3} n_{.j} c_j^{\,2} \right) - \bar{y}^2 = \frac{78125}{20} - (61{,}25)^2 = 154{,}69$$

3.1.3 Distributions conditionnelles

Les distributions conditionnelles s'obtiennent en fixant la valeur d'une des deux variables (où la modalité d'une des deux variables).

Exemple 3.1.2

On considère le tableau suivant, relatif à une population de 100 ménages, tel que :
X = « le nombre d'enfants du ménage » et Y = « le nombre de pièces du logement »

X \ Y	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	*Total*
$x_1 = 2$	15	10	05	30
$x_2 = 3$	30	5	10	45
$x_3 = 4$	10	5	0	15
$x_4 = 5$	10	0	0	10
Total	65	20	15	100

1. La distribution conditionnelle de X sachant $Y = 3$ est donnée par la première colonne du tableau.
2. La distribution conditionnelle de X sachant $Y = 4$ est donnée par la deuxième colonne du tableau.
3. La distribution conditionnelle de Y sachant $X = 2$ est donnée par la première ligne du tableau.
4. De même, la distribution conditionnelle de Y sachant $X = 5$ est donnée par la quatrième ligne du tableau.

Ces quatre distributions se présentent dans les tableaux suivants :

Distribution conditionnelle de X sachant Y = 3

$X / Y = 3$	$n_{i/1}$
$x_1 = 2$	15
$x_2 = 3$	30
$x_3 = 4$	10
$x_4 = 5$	10
Total	65

Distribution conditionnelle de X sachant Y = 4

$X / Y = 4$	$n_{i/2}$
$x_1 = 2$	10
$x_2 = 3$	5
$x_3 = 4$	5
$x_4 = 5$	0
Total	20

Distribution conditionnelle
de Y *sachant* $X=2$

$Y/X=2$	$n_{j/1}$
$y_1=3$	15
$y_2=4$	10
$y_3=5$	05
Total	30

et

Distribution conditionnelle
de Y *sachant* $X=5$

$Y/X=5$	$n_{j/4}$
$y_1=3$	10
$y_2=4$	0
$y_3=5$	0
Total	10

Remarque 3.1.2 $n_{i/1}=n_{i1}$; $n_{i/2}=n_{i2}$; $n_{j/1}=n_{1j}$; $n_{j/4}=n_{4j}$.

En général si on prend le tableau des contingents des effectifs suivant :

Distribution conjointe en effectif de X et Y

X \ Y	y_1	y_2	...	y_j	...	y_k	Total
x_1	n_{11}	n_{12}		n_{1j}		n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}		n_{2j}		n_{2k}	$n_{2.}$
⋮							
x_i	n_{i1}	n_{i2}		n_{ij}		n_{ik}	$n_{i.}$
⋮							
x_r	n_{r1}	n_{r2}		n_{rj}		n_{rk}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$		$n_{.j}$		$n_{.k}$	N

$\Rightarrow$

Distribution conditionnelle
de X *sachant* $Y=y_j$

$X/Y=y_j$	$n_{i/j}$
x_1	n_{1j}
x_2	n_{2j}
⋮	
x_i	n_{ij}
⋮	
x_r	n_{rj}
Total	$n_{.j}$

et

Distribution conditionnelle
de Y *sachant* $X=x_i$

$Y/X=x_i$	$n_{j/i}$
y_1	n_{i1}
y_2	n_{i2}
⋮	
y_j	n_{ij}
⋮	
y_k	n_{ik}
Total	$n_{i.}$

Remarque 3.1.3 La distribution conditionnelle de chacune des variables X et Y peut être définie à partir des fréquences.

1. Dans le cas de la distribution conditionnelle de X sachant $Y=y_j$, on a :

$$f_{i/j}=\frac{n_{ij}}{n_{.j}}=\frac{n_{ij}\times N}{n_{.j}\times N}=\frac{f_{ij}}{f_{.j}};\quad \text{avec } \sum_{i=1}^{r} f_{i/j}=1$$

2. Dans le cas de la distribution conditionnelle de Y sachant $X = x_i$, on a :

$$f_{j/i} = \frac{n_{ij}}{n_{i.}} = \frac{n_{ij} \times N}{n_{i.} \times N} = \frac{f_{ij}}{f_{i.}}; \text{ avec } \sum_{j=1}^{k} f_{j/i} = 1$$

Exemple 3.1.3

On considère le tableau suivant, relatif à une population de 100 ménages, tel que :
X = « le nombre d'enfants du ménage » et Y = « le nombre de pièces du logement ».
Pour calculer les **distributions conditionnelles en fréquences** de X sachant $Y = 4$ et de Y sachant $X = 2$, on a deux façons pour le faire. En effet ;

X \ Y	$y_1 = 3$	$y_2 = 4$	$y_3 = 5$	*Total*
$x_1 = 2$	15	10	05	30
$x_2 = 3$	30	5	10	45
$x_3 = 4$	10	5	0	15
$x_4 = 5$	10	0	0	10
Total	65	20	15	100

1ere méthode : En passant par les distributions conditionnelles des effectifs.

Distribution conditionnelle de X sachant Y = 4

$X / Y = 4$	$n_{i/2}$
$x_1 = 2$	10
$x_2 = 3$	5
$x_3 = 4$	5
$x_4 = 5$	0
Total	20

Distribution conditionnelle de Y sachant X = 2

$Y / X = 2$	$n_{j/1}$
$y_1 = 3$	15
$y_2 = 4$	10
$y_3 = 5$	05
Total	30

⇒

Distribution conditionnelle de X sachant Y = 4

$X / Y = 4$	$f_{i/2}$
$x_1 = 2$	0,5
$x_2 = 3$	0,25
$x_3 = 4$	0,25
$x_4 = 5$	0
Total	1,00

Distribution conditionnelle de Y sachant X = 2

$Y / X = 2$	$f_{j/1}$
$y_1 = 3$	0,5
$y_2 = 4$	0,33
$y_3 = 5$	0,17
Total	1,00

2eme méthode : En passant par la distribution conjointe en fréquences.

Donc

Distribution conjointe en fréquences

X \ Y	$y_1=3$	$y_2=4$	$y_3=5$	*Total*
$x_1=2$	0,15	0,10	0,05	0,30
$x_2=3$	0,30	0,05	0,10	0,45
$x_3=4$	0,10	0,05	0	0,15
$x_4=5$	0,10	0	0	0,10
Total	0,65	0,20	0,15	1,00

En utilisant les formules $f_{i/j}=\frac{f_{ij}}{f_{.j}}$ et $f_{j/i}=\frac{f_{ij}}{f_{i.}}$ on obtient

Distribution conditionnelle de X sachant Y = 4

$X/Y=4$	$f_{i/2}$
$x_1=2$	0,5
$x_2=3$	0,25
$x_3=4$	0,25
$x_4=5$	0
Total	1,00

Distribution conditionnelle de Y sachant X = 2

$Y/X=2$	$f_{j/1}$
$y_1=3$	0,5
$y_2=4$	0,33
$y_3=5$	0,17
Total	1,00

Remarque 3.1.4 $f_{i/1}\neq f_{i1}$; $f_{i/2}\neq f_{i2}$; $f_{j/1}\neq f_{1j}$; $f_{j/4}\neq f_{4j}$.

3.1.4 Variables indépendantes

Les variables X et Y sont indépendantes si et seulement si

$$\forall i,j \Rightarrow f_{ij}=f_{i.}\times f_{.j}$$

Remarque 3.1.5 On a l'équivalence suivante :

$$\forall i,j : f_{ij}=f_{i.}\times f_{.j} \Leftrightarrow n_{ij}=\frac{n_{i.}\times n_{.j}}{N}$$

Exemple 3.1.4 On considère le tableau suivant, relatif à une population de 54 ménages, tel que :

X \ Y	$y_1=3$	$y_2=4$	$y_3=5$	*Total*
$x_1=2$	2	4	12	18
$x_2=3$	4	8	24	36
Total	6	12	36	54

Les deux variables X et Y sont indépendantes car :

1	$\frac{n_{1.} \times n_{.1}}{N} = \frac{18 \times 6}{54} = 2 = n_{11}$	4	$\frac{n_{2.} \times n_{.1}}{N} = \frac{36 \times 6}{54} = 4 = n_{21}$
2	$\frac{n_{1.} \times n_{.2}}{N} = \frac{18 \times 12}{54} = 4 = n_{12}$	5	$\frac{n_{2.} \times n_{.2}}{N} = \frac{36 \times 12}{54} = 8 = n_{22}$
3	$\frac{n_{1.} \times n_{.3}}{N} = \frac{18 \times 36}{54} = 12 = n_{13}$	6	$\frac{n_{2.} \times n_{.3}}{N} = \frac{36 \times 36}{54} = 24 = n_{23}$

3.2 Paramètres de liaison

Dans toute la suite on considère N observations sur les deux variables X et Y.

3.2.1 Covariance entre X et Y

La covariance est égale à la moyenne des écarts des couples les (x_i, y_i) de X et Y par rapport au point $(\bar{x}, \bar{y})$.

$$Cov(x,y) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i - \bar{x})(y_i - \bar{y})$$

- **Le rôle de la covariance**

La covariance indique le sens de la relation entre les variables X et Y. Ainsi, On peut distinguer les cas suivants :

1er Cas : Si $Cov(x,y) > 0$, alors on peut dire que **la relation entre les deux variables est positive.** Dans ce cas, **ces deux variables varient dans le même sens.**

2eme Cas : Si $Cov(x,y) < 0$, alors on peut dire que **la relation entre les deux variables est négative.** Dans ce cas, **ces deux variables varient en sens inverse.**

3eme Cas : Si $Cov(x,y) = 0$, alors on peut dire qu'**il n'y a pas de relation entre les deux variables.** Dans ce cas, **les variations de l'une n'entraînent pas la variation de l'autre.**

3.2.2 Les propriétés de la covariance

1ere *Propriété*

$$Cov(ax+b,cy+d)=ac.Cov(x,y)$$

Démonstration : En effet,

$$Cov(ax+b,cy+d)=\frac{1}{N}\sum_{i=1}^{i=N}[(ax_i+b)-(a\bar{x}+b)].[(cy_i+d)-(c\bar{y}+d)]$$

$$=\frac{1}{N}\sum_{i=1}^{i=N}(ax_i-a\bar{x})(cy_i-c\bar{y})$$

$$=\frac{1}{N}\sum_{i=1}^{i=N}(a\times c)(x_i-\bar{x})(y_i-\bar{y})$$

$$=(a\times c)\left[\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})\right]$$

$$\Rightarrow Cov(ax+b,cy+d)=ac.Cov(x,y)$$

2eme *Propriété*

$$Cov(y,x)=Cov(x,y)$$

Démonstration : En effet,

$$Cov(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{N}\sum_{i=1}^{i=N}(y_i-\bar{y})(x_i-\bar{x})$$

$$\Rightarrow Cov(y,x)=Cov(x,y)$$

3eme *Propriété*

$$Cov(x,x)=V(x)$$

Démonstration : En effet,

$$Cov(x,x)=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(x_i-\bar{x})=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})^2=V(x)$$

4eme *Propriété*

$$Cov(x,y)=\left(\frac{1}{N}\sum_{i=1}^{i=N}x_iy_i\right)-(\bar{x}.\bar{y})$$

Démonstration : En effet,

$$Cov(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})=\frac{1}{N}\sum_{i=1}^{i=N}(x_iy_i-\bar{y}x_i-\bar{x}y_i+\bar{x}.\bar{y})$$

$$=\frac{1}{N}\sum_{i=1}^{i=N}(x_iy_i)-\frac{1}{N}\sum_{i=1}^{i=N}(\bar{y}x_i)-\frac{1}{N}\sum_{i=1}^{i=N}(\bar{x}y_i)+\frac{1}{N}\sum_{i=1}^{i=N}(\bar{x}.\bar{y})$$

$$=\frac{1}{N}\left(\sum_{i=1}^{i=N}(x_iy_i)-\bar{y}\underbrace{\sum_{i=1}^{i=N}(x_i)}_{N.\bar{x}}-\bar{x}\underbrace{\sum_{i=1}^{i=N}(y_i)}_{N.\bar{y}}+\underbrace{\sum_{i=1}^{i=N}(\bar{x}.\bar{y})}_{N.\bar{x}.\bar{y}}\right)$$

$$=\frac{1}{N}\left(\sum_{i=1}^{i=N}(x_iy_i)-N.\bar{x}.\bar{y}-N.\bar{x}.\bar{y}+N.\bar{x}.\bar{y}\right)$$

$$\Rightarrow Cov(x,y)=\left(\frac{1}{N}\sum_{i=1}^{i=N}x_iy_i\right)-(\bar{x}.\bar{y}).$$

3.2.3 Le coefficient de corrélation linéaire entre X et Y

Le coefficient de corrélation linéaire entre X et Y est

$$r_{x,y}=\frac{Cov(x,y)}{\sigma(x)\sigma(y)}=\frac{Cov(x,y)}{\sqrt{V(x)V(y)}}$$

Remarque 3.2.1 Le coefficient de corrélation linéaire est un nombre sans dimension car :

$$r_{x,y}=\frac{Cov(x,y)}{\sigma(x)\sigma(y)} \text{ et } Cov(x,y)=\frac{1}{N}\sum_{i=1}^{i=N}(x_i-\bar{x})(y_i-\bar{y})$$

3.2.4 Les propriétés du coefficient de corrélation linéaire

1ere *Propriété*

$$r_{ax+b,cy+d}=(Signe\ de\ a)(Signe\ de\ c).r_{x,y}$$

Démonstration : En effet,

$$r_{ax+b,cy+d}=\frac{Cov(ax+b,cy+d)}{\sqrt{V(ax+b)}\sqrt{V(cy+d)}}=\frac{(a\times c).Cov(x,y)}{|a|\sqrt{V(x)}|c|\sqrt{V(y)}}$$

$$= \frac{(a \times c)}{|a| \times |c|} \frac{Cov(x,y)}{\sqrt{V(x)}\sqrt{V(y)}}$$

$$\Rightarrow r_{ax+b,cy+d} = (Signe\ de\ a)(Signe\ de\ c).r_{x,y}$$

2eme Propriété

$$r_{y,x} = r_{x,y}$$

Démonstration : En effet,

$$r_{y,x} = \frac{Cov(y,x)}{\sqrt{V(y)V(x)}} = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} = r_{x,y}$$

3eme Propriété

$$r_{x,x} = 1$$

Démonstration : En effet,

$$Cov(x,y) = \left(\frac{1}{N}\sum_{i=1}^{i=N} x_i y_i\right) - (\bar{x}.\bar{y}) \Rightarrow Cov(x,x) = \left(\frac{1}{N}\sum_{i=1}^{i=N} x_i^{\,2}\right) - \bar{x}^2$$

$$\Rightarrow Cov(x,x) = V(x)$$

$$\Rightarrow r_{x,x} = \frac{Cov(x,x)}{\sqrt{V(x)V(x)}} = \frac{V(x)}{\sqrt{V(x)V(x)}} \Rightarrow r_{x,x} = 1$$

4eme Propriété

Le coefficient de corrélation linéaire est compris entre -1 et 1, c'est-à-dire :

$$-1 \le r_{x,y} \le +1$$

Démonstration : Admis.

- **Le rôle du coefficient de corrélation linéaire**

Le coefficient de corrélation linéaire permet de mesurer le degré ou l'intensité de la liaison linéaire entre deux variables statistiques. C'est-à-dire :

1. Si $r_{x,y} = 1$, on dit qu'il y a une parfaite corrélation linéaire positive entre les deux

variables.

2. Si $r_{x,y} = -1$, on dit qu'il y a une parfaite corrélation linéaire négative entre les deux variables.
3. Si $r_{x,y} = 0$, on dit qu'il y a absence de corrélation linéaire entre les deux variables.
4. On dit qu'il y a une forte corrélation linéaire entre les deux variables (ou forte dépendance linéaire) si $r_{x,y}$ est proche de ± 1.
5. En revanche, si $r_{x,y}$ est proche de zéro (0), on dit qu'il y a une faible corrélation linéaire entre les deux variables.

3.3 Ajustement d'un nuage de points

Dans toute la suite on considère N observations sur les deux variables X et Y.

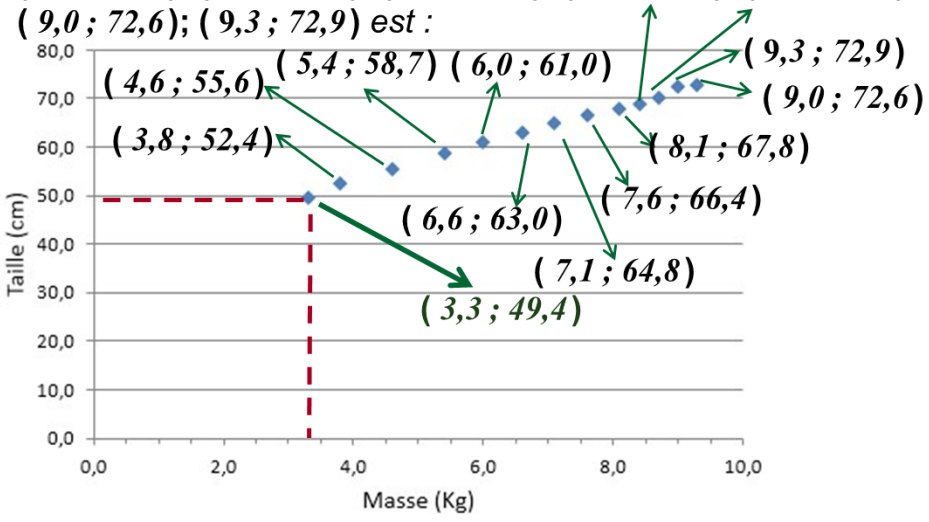
3.3.1 Nuage de points

Ensemble de points isolés représentés dans un graphique cartésien ; c'est-à-dire des points $M_1, M_2, \ldots, M_n$ de coordonnées (x_1, y_1) ; (x_2, y_2); … ; (x_n, y_n)

Exemple 3.3.1 On considère le tableau suivant, relatif à une population associé à deux variables mesurées sur 13 bébés tels que, X = « le poids du bébé » et Y = « la taille du bébé »

Masse (kg)	3,3	3,8	4,6	5,4	6,0	6,6	7,1	7,6	8,1	8,4	8,7	9,0	9,3
Taille (cm)	49,4	52,4	55,6	58,7	61,0	63,0	64,8	66,4	67,8	69,0	70,3	72,6	72,9

Le nuage des points de coordonnées $(3,3 ; 49,4)$; $(3,8 ; 52,4)$; $(4,6 ; 55,6)$; $(5,4 ; 58,7)$; $(6,0 ; 61,0)$; $(6,6 ; 63,0)$; $(7,1 ; 64,8)$; $(7,6 ; 66,4)$; $(8,1 ; 67,8)$; $(8,4 ; 69,0)$; $(8,7 ; 70,3)$; $(9,0 ; 72,6)$; $(9,3 ; 72,9)$ *est :*



3.3.2 Ajustement linéaire d'un nuage de points

On considère N observations sur les deux variables X et Y, donc ;

1. Ces observations peuvent être représentées par un nuage de points.
2. Notre but est d'exprimer Y en fonction de X.
3. La représentation du nuage de points peut nous renseigner sur l'allure de la courbe de régression.

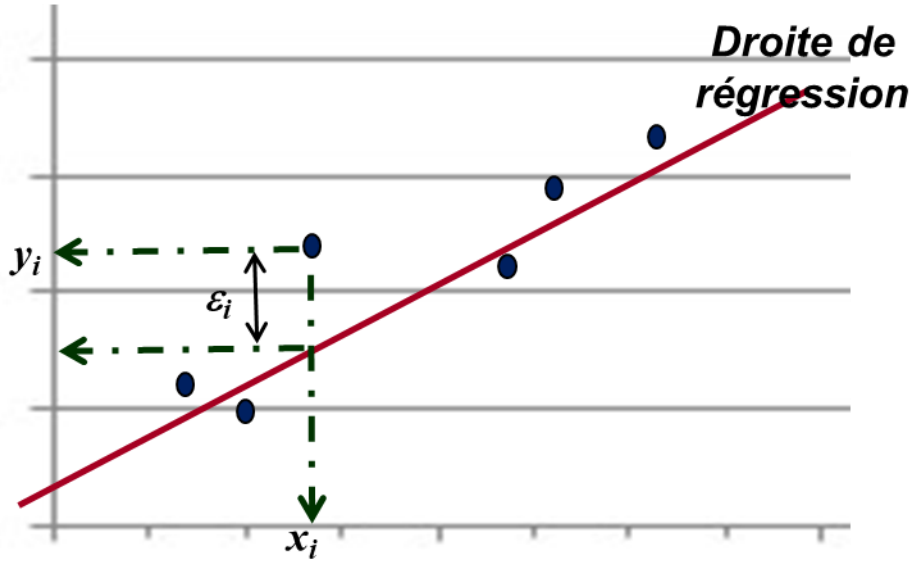
Remarque 3.3.1

1. L'ajustement linéaire consiste à trouver l'équation d'une droite du type $y = ax + b$, appelée droite de régression. Cette droite donne l'évolution de la variable Y (variable expliquée) en fonction de la variable explicative X.

2. La méthode d'ajustement que nous allons exposer est appelée « méthode des Moindres Carrés Ordinaires » ou simplement **« MCO ».**

- **La méthode des Moindres Carrés Ordinaires**

Considérons N couples d'observations (x_i, y_i), leurs nuage est :



Donc les couples (x_i, y_i) vérifient :

$$y_i = (ax_i + b) + \varepsilon_i \quad \forall i \in \{1, \cdots, N\}$$

où ε_i représente le résidu du couple (x_i, y_i). On peut alors écrire :

$$\varepsilon_i = y_i - (ax_i + b)$$

Remarque 3.3.2 La méthode des **MCO** consiste à minimiser $\sum_{i=1}^{i=N} \varepsilon_i^{\,2}$ tels que :

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

Les deux conditions de premier ordre de la minimisation de cette fonction f par rapport à a et à b sont :

$$\frac{\partial\left(\sum_{i=1}^{i=N} \varepsilon_i^{\,2}\right)}{\partial a} = 0 \quad \text{et} \quad \frac{\partial\left(\sum_{i=1}^{i=N} \varepsilon_i^{\,2}\right)}{\partial b} = 0$$

$$\Rightarrow \frac{\partial\left(\sum_{i=1}^{i=N} \varepsilon_i^{\,2}\right)}{\partial a} = 2\sum_{i=1}^{i=N} (y_i - ax_i - b)(-x_i) = 0 \Rightarrow \sum_{i=1}^{i=N} (y_i - ax_i - b)(x_i) = 0 \qquad (1)$$

et $\frac{\partial\left(\sum_{i=1}^{i=N} \varepsilon_i^{\,2}\right)}{\partial b} = 2\sum_{i=1}^{i=N} (y_i - ax_i - b)(-1) = 0 \Rightarrow \sum_{i=1}^{i=N} (y_i - ax_i - b) = 0 \qquad (2)$

$$(1) \Rightarrow \sum_{i=1}^{i=N} (y_i x_i - ax_i^{\,2} - bx_i) = \sum_{i=1}^{i=N} y_i x_i - a\sum_{i=1}^{i=N} x_i^{\,2} - b\sum_{i=1}^{i=N} x_i = 0 \quad (3)$$

$$(2) \Rightarrow \sum_{i=1}^{i=N} (y_i - ax_i - b) = \sum_{i=1}^{i=N} y_i - a\sum_{i=1}^{i=N} x_i - Nb = 0 \qquad (4)$$

En divisant les deux membres de l'équation (4) par N, on obtient :

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

Sachant que $\bar{x} = \frac{1}{N}\sum_{i=1}^{i=N} x_i$ et $\bar{y} = \frac{1}{N}\sum_{i=1}^{i=N} y_i$ donc l'équation devient :

$$\bar{y} - a\bar{x} - b = 0 \quad (5) \Leftrightarrow b = \bar{y} - a\bar{x}$$

En remplaçant, dans l'équation (3), b par $\bar{y} - a\bar{x}$, d'après l'équation (5) on obtient

$$\sum_{i=1}^{i=N} y_i x_i - a\sum_{i=1}^{i=N} x_i^{\,2} - (\bar{y} - a\bar{x})\sum_{i=1}^{i=N} x_i = 0$$

$$\Leftrightarrow \sum_{i=1}^{i=N} y_i x_i - a\sum_{i=1}^{i=N} x_i^{\,2} - \bar{y}\underbrace{\sum_{i=1}^{i=N} x_i}_{N.\bar{x}} + a\bar{x}\underbrace{\sum_{i=1}^{i=N} x_i}_{N.\bar{x}} = 0$$

$$\Leftrightarrow \sum_{i=1}^{i=N} y_i x_i - a\sum_{i=1}^{i=N} x_i^{\,2} - N\bar{x}.\bar{y} + aN\bar{x}^2 = 0$$

$$\Leftrightarrow \sum_{i=1}^{i=N} y_i x_i - N\bar{x}.\bar{y} = a\left(\sum_{i=1}^{i=N} {x_i}^2 - N\bar{x}^2\right)$$

Ainsi, on obtient la valeur estimée de la pente de la droite de régression :

$$\hat{a} = \frac{\sum_{i=1}^{i=N} x_i y_i - N\bar{x}.\bar{y}}{\sum_{i=1}^{i=N} {x_i}^2 - N\bar{x}^2} \Rightarrow \quad \hat{b} = \bar{y} - \hat{a}\bar{x}$$

Donc l'équation de la droite de régression est :

$$y = \hat{a}x + \hat{b}$$

Remarque 3.3.3 On peut aussi calculer la valeur estimée de la pente de la droite de régression en utilisant l'une de ces deux expressions suivantes :

$$\hat{a} = \frac{Cov(x,y)}{V(x)} \quad \text{et} \quad \hat{a} = \frac{\sum_{i=1}^{i=N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{i=N}(x_i - \bar{x})^2} \text{ car}$$

$$\hat{a} = \frac{\sum_{i=1}^{i=N} y_i x_i - N\bar{x}.\bar{y}}{\sum_{i=1}^{i=N} {x_i}^2 - N\bar{x}^2} = \frac{\frac{1}{N}\left(\sum_{i=1}^{i=N} y_i x_i - N\bar{x}.\bar{y}\right)}{\frac{1}{N}\left(\sum_{i=1}^{i=N} {x_i}^2 - N\bar{x}^2\right)} = \frac{\frac{1}{N}\sum_{i=1}^{i=N} y_i x_i - \bar{x}.\bar{y}}{\frac{1}{N}\sum_{i=1}^{i=N} {x_i}^2 - \bar{x}^2}$$

$$\hat{a} = \frac{Cov(x,y)}{V(x)} \quad \text{et} \quad \hat{a} = \frac{\sum_{i=1}^{i=N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{i=N}(x_i - \bar{x})^2}$$

Remarque 3.3.4 La droite de régression passe par le point moyen de coordonnées $(\bar{x}, \bar{y})$.

En effet

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \Rightarrow \bar{y} = \hat{a}\bar{x} + \hat{b}.$$

Exercice d'application

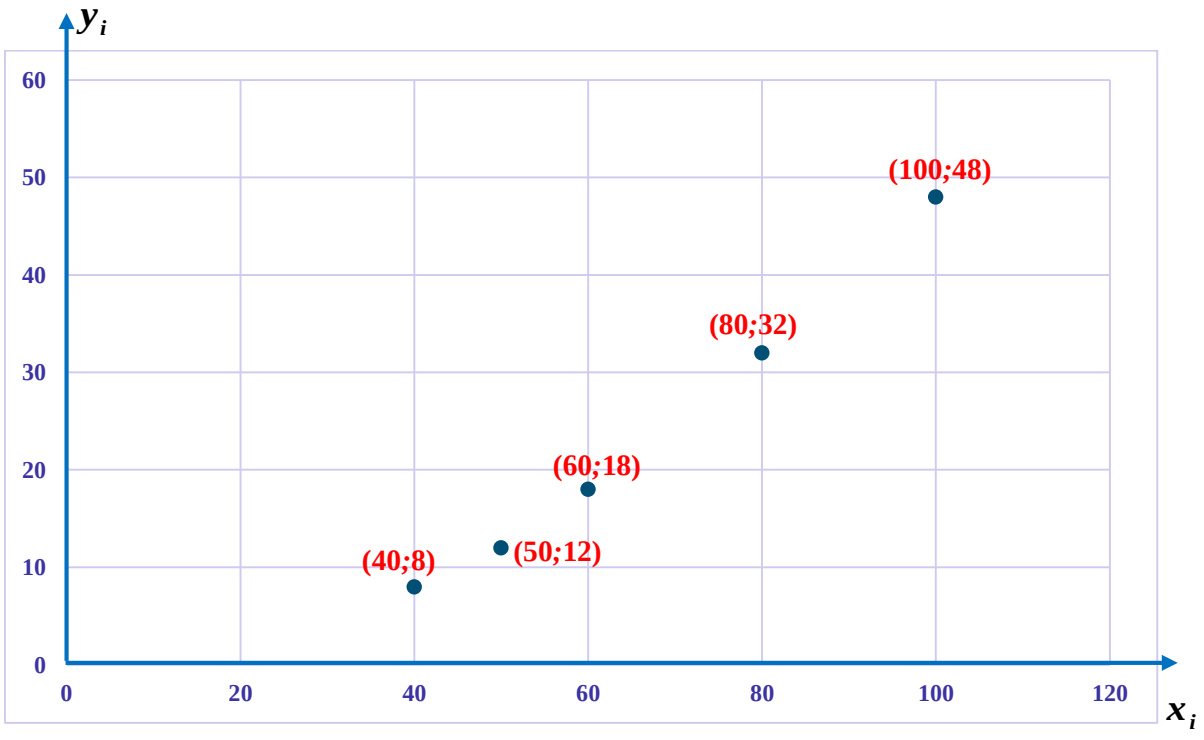
Le tableau suivant donne la distance de freinage d'un véhicule automobile sur une route sèche, en fonction de sa vitesse.

Vitesse en Km/h (x_i)	*Distance en m* (y_i)
40	8
50	12
60	18
80	32
100	48

1. Construire le nuage des points.
2. Calculer la covariance entre la vitesse X et la distance Y.
   Que peut on déduire sur la relation entre X et Y.
3. Calculer le coefficient de corrélation linéaire.
   Conclure sur l'intensité de la liaison entre X et Y.
4. Déterminer, en utilisant la méthode des moindres carrés, l'équation de la droite de régression permettant d'estimer la distance de freinage en fonction de la vitesse du véhicule.
5. Interpréter la pente et la constante de l'équation de la droite obtenue.
6. A combien peut-on estimer la distance de freinage d'un véhicule roulant à 120 km/h.
7. Déterminer cette même droite sachant qu'une sixième mesure a donné pour : $x_i = 0$; $y_i = 0$.

Solution de l'exercice

1. Le nuage des points (x_i, y_i)



2. La covariance entre la vitesse X et la distance Y.

$$Cov(x,y)=\left(\frac{1}{5}\sum_{i=1}^{i=5}x_i y_i\right)-(\bar{x}.\bar{y}) \text{ avec } \bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i \text{ et } \bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i$$

i	x_i	y_i	$x_i y_i$
1	40	8	320
2	50	12	600
3	60	18	1080
4	80	32	2560
5	100	48	4800
Total	330	118	9360

$$\Rightarrow \bar{x}=\frac{330}{5}=66;\ \bar{y}=\frac{118}{5}=23{,}6$$

$$\Rightarrow Cov(x,y)=\frac{9360}{5}-66\times 23{,}6=1872-1557{,}6$$

$$\Rightarrow Cov(x,y)=314{,}4.$$

Conclusion

Comme $Cov(x, y) > 0$ alors la relation entre la vitesse et la distance de freinage est positive et les deux variables varient dans le même sens.

3. Le coefficient de corrélation linéaire.

Conclure sur l'intensité de la liaison entre X et Y.

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	40	8	320	1600	64
2	50	12	600	2500	144
3	60	18	1080	3600	324
4	80	32	2560	6400	1024
5	100	48	4800	10000	2304
Total	330	118	9360	24100	3860

$$r_{x,y}=\frac{Cov(x,y)}{\sqrt{V(x)V(y)}} \text{ avec } V(x)=\left(\frac{1}{N}\sum_{i=1}^{i=r}{x_i}^2\right)-\bar{x}^2 \text{ et } V(y)=\left(\frac{1}{N}\sum_{i=1}^{i=r}{y_i}^2\right)-\bar{y}^2$$

$$\Rightarrow V(x)=\frac{24100}{5}-(66)^2=4820-4356\Rightarrow V(x)=464.$$

Et $V(y)=\frac{3860}{5}-(23{,}6)^2=772-556{,}96\Rightarrow V(y)=215{,}04$

$$\Rightarrow r_{x,y} = \frac{314{,}4}{\sqrt{464 \times 215{,}04}} \Rightarrow r_{x,y} = 0{,}99$$

Conclusion

Les variables varient dans le même sens. La valeur de $r_{x,y}$, proche de 1, cela traduit une forte corrélation linéaire entre les deux variables.

4. L'équation de la droite de régression permettant d'estimer la distance de freinage en fonction de la vitesse en utilisant la méthode des moindres carrés ordinaire.

$$\hat{a} = \frac{\sum_{i=1}^{i=N} x_i y_i - N\bar{x}.\bar{y}}{\sum_{i=1}^{i=N} x_i{}^2 - N\bar{x}^2} \Rightarrow \hat{a} = \frac{9360 - 5 \times 66 \times 23{,}6}{24100 - 5 \times 66^2} \Rightarrow \hat{a} = 0{,}67$$

De plus l'ordonnée à l'origine est égale à :

$$\hat{b} = \bar{y} - \hat{a}\bar{x} = 23{,}6 - 0{,}67 \times 66 \Rightarrow \hat{b} = -20{,}62$$

donc L'équation s'écrit : $y = 0{,}67x - 20{,}62$

5. Interprétation de la pente et la constante de l'équation de la droite obtenue :

- Lorsque la vitesse augmente de 1 km/h la distance de freinage augmente de $\hat{a} = 0{,}67m$.
- La constante $\hat{b} = -20{,}62$ Indique qu'à l'arrêt le véhicule est en retard d'une distance de $20{,}62\text{m}$.

6. L'estimation de la distance de freinage d'un véhicule roulant à $120\ km/h$.

L'équation étant : $y = 0{,}67x - 20{,}62$.

En remplaçant x par 120 on obtient :

$y = 0{,}67 \times 120 - 20{,}62 = 59{,}78$.

Donc la distance de freinage d'un véhicule roulant à $120\ km/h$ est $y = 59{,}78\ m$.

7. Détermination de la même droite sachant qu'une sixième mesure a donné pour :

$x_i = 0$; $y_i = 0$.

C'est-à-dire l'équation des moindres carrés avec $(x_i = 0\ ;\ y_i = 0)$

Il suffit de refaire les calculs avec les mêmes sommes mais en divisant par le nouveau nombre d'observations = effectif total qui est égal à $N = 6$.

$$\Rightarrow \bar{x} = \frac{330}{6} = 55;\ \ \bar{y} = \frac{118}{6} = 19{,}67$$

$$\Rightarrow \hat{a} = \frac{\sum_{i=1}^{i=N} y_i x_i - N\bar{x}.\bar{y}}{\sum_{i=1}^{i=N} {x_i}^2 - N\bar{x}^2} = \frac{9360 - 6\times 55\times 19{,}67}{24100 - 6\times 55^2} \Rightarrow \hat{a} = 0{,}48$$

$$\Rightarrow \hat{b} = \bar{y} - \hat{a}\bar{x} = 19{,}67 - 0{,}48\times 55 \Rightarrow \hat{b} = -6{,}74$$

donc L'équation s'écrit : $y = 0{,}48x - 6{,}74$

3.3.3 Ajustement non linéaire d'un nuage de points

On considère N observations sur les deux variables X et Y.
Dans le cas général, la relation entre X et Y semble être plutôt non linéaire, c'est-à-dire n'est pas de la forme $y = ax + b$.
En fait lorsque le nuage de points manifeste en tendance courbe et que le coefficient de corrélation linéaire n'est pas proche de 1 en valeur absolue, l'ajustement de ce nuage par une droite est hasardeux et aboutira à des estimations de mauvaise qualité. Dans ce cas, on peut tenter d'utiliser un des modèles proposés dans ce paragraphe. En fait, chacun de ces modèles utilise le principe d'ajustement par la méthode des moindres carrés (donc ils utilisent tous une droite) mais en "transformant" au préalable les données pour obtenir un modèle linéaire à partir du modèle non-linéaire considéré.
On va voir les deux modèles d'ajustements non linéaires suivants :

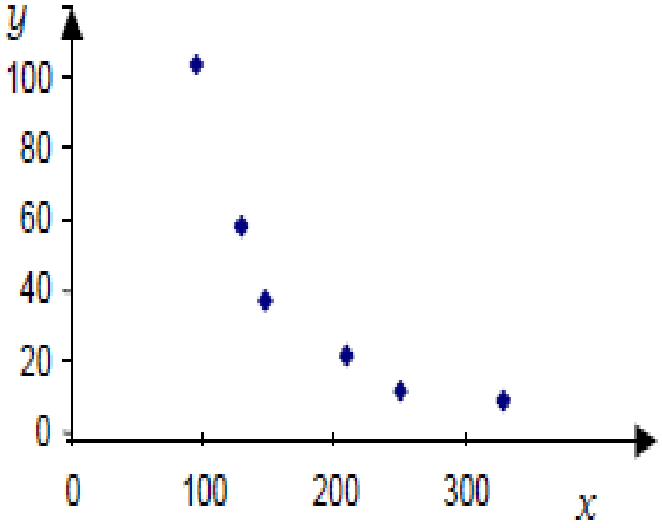
- Ajustement hyperbolique.
- Ajustement par une fonction puissance.

1er modèle : L'ajustement hyperbolique

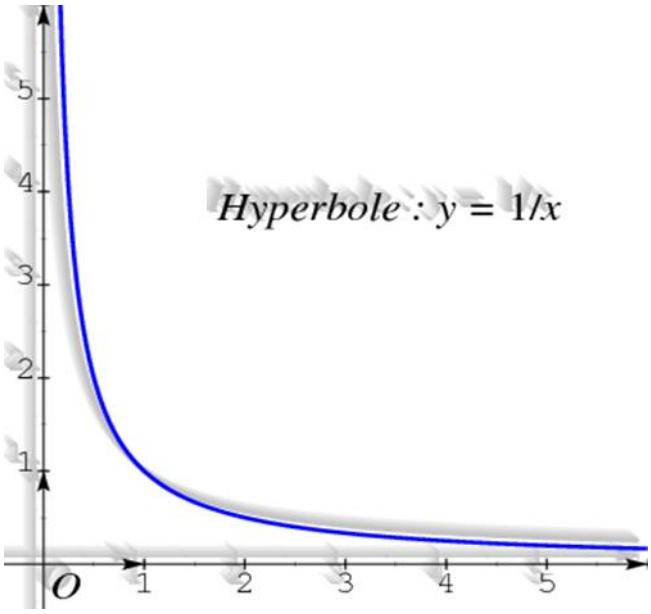
Les N points (x_i, y_i) ne sont pas alignés, mais plutôt proches d'une courbe représentant une fonction hyperbolique de la forme :

$$y = \frac{b}{x^a} \text{ avec } a > 0, \quad b > 0$$

Dans ce cas le nuage aura l'allure suivante



Qui ressemble à la courbe de la fonction $f(x)=\frac{1}{x}$ qui apparait comme suit :



Comment peut-on estimer b et a ?

Nous sommes en présence d'une relation non linéaire entre y et x.
Afin d'utiliser la méthode des **MCO**, il faut d'abord retrouver, moyennant une transformation, dans ce cas logarithmique, une forme linéaire :
On cherche a et b tels que :

$$y=\frac{b}{x^a}=bx^{-a}$$

En utilisant le logarithme népérien dans cette équation on trouve :

$$\Rightarrow \ln y=\ln bx^{-a}=\ln b-a\ln x$$

Et si on suppose que : $\beta=\ln b$ et $\alpha=-a$.

Le modèle linéaire est alors de la forme :

$$\ln y = \alpha \ln x + \beta$$

Donc en utilisant la méthode des **MCO**, on peut retrouver α et β :

$$\hat{\alpha} = \frac{Cov(\ln x, \ln y)}{V(\ln x)} = \frac{\sum_{i=1}^{i=N} (\ln x_i)(\ln y_i) - N\overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N} (\ln x_i)^2 - N\overline{\ln x}^2}$$

$$\Rightarrow \hat{\beta} = \overline{\ln y} - \hat{\alpha}\overline{\ln x}$$

On peut maintenant retrouver la valeur de b et la valeur de a :

$\beta = \ln b \Rightarrow \hat{b} = e^{\hat{\beta}}$ et $\alpha = -a \Rightarrow \hat{a} = -\hat{\alpha}$.

Exercice d'application

Une entreprise fabrique un équipement. Le prix unitaire Y (en Dollar) de ce produit est en fonction du nombre X d'unités produites. On a relevé les résultats suivants.

X	22	23	24	30	60	174
Y	120	60	25	10	4	1

1. Calculer la covariance entre le nombre d'unités produites X et le prix unitaire Y. Que peut-on déduire sur la relation entre X et Y.
2. Calculer le coefficient de corrélation linéaire $r_{x,y}$. Conclure sur l'intensité de la liaison entre les deux variables X et Y.
3. Représenter le nuage de points (x_i, y_i).
4. Compte tenue de cette représentation, donner la forme de l'ajustement de ce nuage de points et retrouver la relation entre les deux variables X et Y.
5. Quelle est Le prix unitaire du produit avec cette approximation pour produire 15 unités.

Solution de l'exercice

1. La covariance entre le nombre d'unités produites X et le prix unitaire Y.

$$Cov(x,y) = \left(\frac{1}{6}\sum_{i=1}^{i=6} x_i y_i\right) - (\bar{x}.\bar{y}) \text{ avec } \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \ \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

	x_i	y_i	x_i^2	$x_i \cdot y_i$
	22	120	484	2640
	23	60	529	1380
	24	25	576	600
	30	10	900	300
	60	4	3600	240
	174	1	30276	174
Total	333	220	36365	5334

$$\Rightarrow \bar{x} = \frac{333}{6} = 55{,}5; \quad \bar{y} = \frac{220}{6} = 36{,}667 \Rightarrow Cov(x,y) = \frac{5334}{6} - 55{,}5 \times 36{,}667$$

$$\Rightarrow Cov(x,y) = 889 - 2035{,}019 = -1146{,}019.$$

Conclusion

Comme $Cov(x,y) < 0$ alors alors la relation entre les deux variables est négative. Dans ce cas, ces deux variables varient en sens inverse.

2. Le coefficient de corrélation linéaire $r_{x,y}$.

$$r_{x,y} = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}} \text{ avec } V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=r} x_i^{\,2}\right) - \bar{x}^2 \text{ et } V(y) = \left(\frac{1}{N}\sum_{i=1}^{i=r} y_i^{\,2}\right) - \bar{y}^2$$

$$V(x) = \left(\frac{1}{N}\sum_{i=1}^{i=r} x_i^{\,2}\right) - \bar{x}^2 = \frac{36365}{6} - (55{,}5)^2 \Rightarrow V(x) = 2980{,}583.$$

Et $V(y) = \frac{18742}{6} - (36{,}667)^2 = 3123{,}667 - 1344{,}469 = 1779{,}198$

$$r_{x,y} = \frac{Cov(x,y)}{\sigma(x)\sigma(y)} = \frac{-1146{,}019}{\sqrt{2980{,}583 \times 1779{,}198}}$$

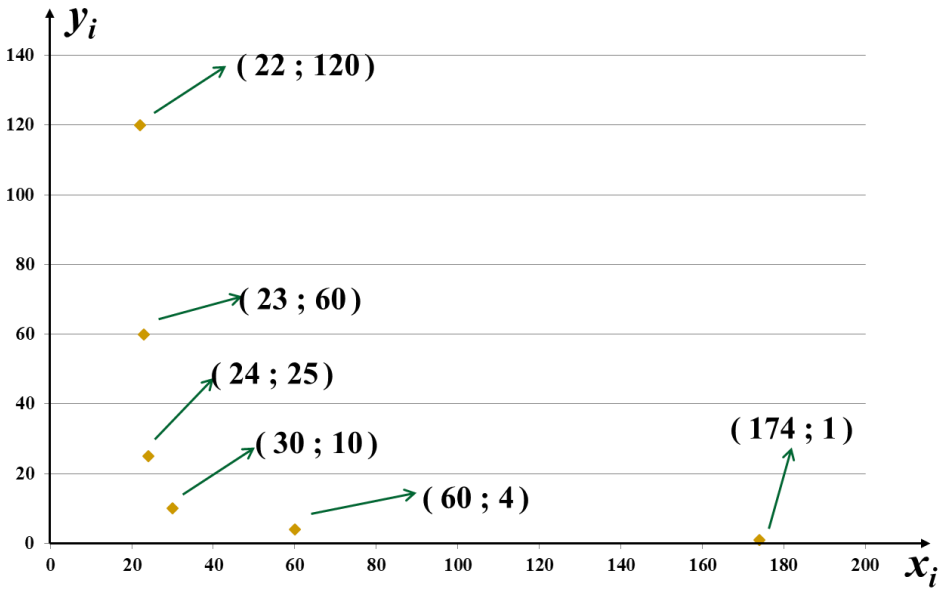
$$\Rightarrow r_{x,y} = \frac{-1146{,}019}{\sqrt{5303047{,}312}} = \frac{-1146{,}019}{2302{,}835} = -\,0{,}498$$

Conclusion

La valeur de $r_{x,y}$ n'est pas proche ni de ± 1 ni de 0 cela traduit qu'il n'ya pas ni forte corrélation linéaire entre les deux variables ni faible corrélation linéaire.

Il y'a une juste moyenne corrélation linéaire entre les deux variables (La valeur de $r_{x,y}$ est proche de $0{,}5$).

3. Le nuage de points (x_i, y_i).



4. Compte tenue de cette représentation, donner la forme de l'ajustement de ce nuage de points et retrouver la relation entre les deux variables X et Y.

L'allure du nuage ressemble à une hyperbole, donc la forme théorique de l'ajustement de ce nuage de points est une forme hyperbolique de la forme :

$$y = \frac{b}{x^a} = bx^{-a} \text{ avec } a > 0, \quad b > 0$$

$$\Rightarrow \ln y = \ln bx^{-a} = \ln b - a \ln x$$

Posons : $\beta = \ln b$ et $\alpha = -a$.

$$\Rightarrow \ln y = \alpha \ln x + \beta$$

En considérant les deux (02) nouvelles variables et en utilisant la méthode des **MCO**, on peut retrouver α et β, tels que :

$$\Rightarrow \hat{\alpha} = \frac{\sum_{i=1}^{i=N} (\ln x_i)(\ln y_i) - N\overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N} (\ln x_i)^2 - N\overline{\ln x}^2} \text{ et } \hat{\beta} = \overline{\ln y} - \hat{\alpha}\overline{\ln x}$$

	x_i	y_i	lnx_i	lny_i	$(lnx_i)(lny_i)$	$(lnx_i)^2$
	22	120	3,091	4,787	14,797	9,554
	23	60	3,135	4,094	12,835	9,828
	24	25	3,178	3,219	10,230	10,100
	30	10	3,401	2,303	7,833	11,567
	60	4	4,094	1,386	5,674	16,761
	174	1	5,159	0	0	26,615
Total	333	220	22,058	15,789	51,369	84,425

Donc $\overline{\ln x} = \frac{1}{6}\sum_{i=1}^{6} \ln x_i = \frac{22{,}058}{6} = 3{,}676$

$\overline{\ln y} = \frac{1}{6}\sum_{i=1}^{6} \ln y_i = \frac{15{,}789}{6} = 2{,}632.$

$\Rightarrow \hat{\alpha} = \frac{51{,}369 - 6 \times 3{,}676 \times 2{,}632}{84{,}425 - 6 \times (3{,}676)^2} = -1{,}9964459 \cong -2$

$\Rightarrow \hat{\beta} = \overline{\ln y} - \alpha \overline{\ln x} = 2{,}632 + 2 \times 3{,}676 = 2{,}632 + 7{,}352 = 9{,}984$

$\Rightarrow a = 2, \ \beta = \ln b$ et $b = e^{9{,}984} = 21676{,}847$

Donc l'ajustement est : $y = \frac{b}{x^a} = \frac{21676{,}847}{x^2}$

5. Le prix unitaire du produit avec cette approximation pour produire **15** unités.

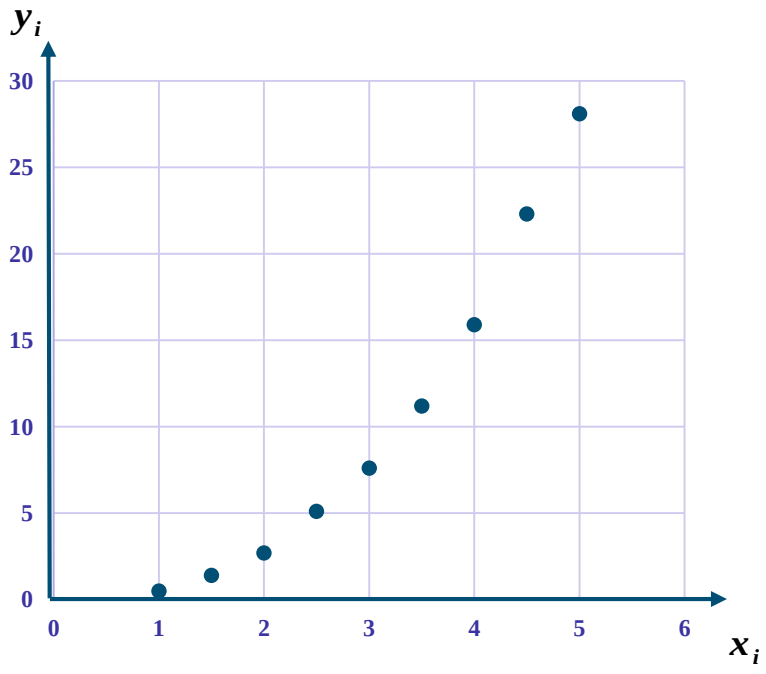
$y = \frac{b}{x^a} = \frac{21676{,}847}{x^2} = \frac{21676{,}847}{15^2} = 96{,}342.$

2eme modèle : L'ajustement par une fonction puissance

Les N points (x_i, y_i) ne sont pas alignés, mais plutôt proches d'une courbe représentant une fonction puissance de la forme :

$$y = b \cdot x^a \text{ avec } a > 0, \ b > 0$$

Dans ce cas le nuage aura l'allure suivante



Qui ressemble à la courbe de la fonction $f(x) = x^2$.

Comment peut-on estimer b et a ?

Nous raisonnons de la même façon que dans le cas hyperbolique et afin d'utiliser la méthode des **MCO**, moyennant une transformation, dans ce cas logarithmique, une forme linéaire :

On cherche a et b tels que :

$$y = b \cdot x^a \text{ avec } a > 0, \quad b > 0$$

En utilisant le logarithme népérien dans cette équation on trouve :

$$\Rightarrow \ln y = \ln b x^a = \ln b + a \ln x$$

Et si on suppose que : $\beta = \ln b$ et $\alpha = a$.

Le modèle linéaire est alors de la forme :

$$\ln y = \alpha \ln x + \beta$$

Donc en utilisant la méthode des **MCO**, on peut retrouver α et β :

$$\hat{\alpha} = \frac{Cov(\ln x, \ln y)}{V(\ln x)} = \frac{\sum_{i=1}^{i=N} (\ln x_i)(\ln y_i) - N\overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N} (\ln x_i)^2 - N\overline{\ln x}^2}$$

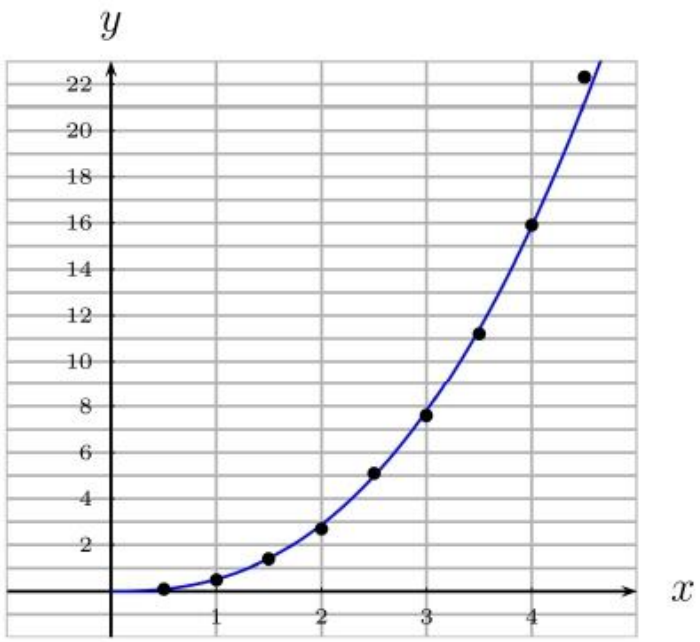
$$\Rightarrow \hat{\beta} = \overline{\ln y} - \hat{\alpha}\overline{\ln x}$$

On peut maintenant retrouver la valeur de b et la valeur de a :

$\beta = \ln b \Rightarrow \hat{b} = e^{\hat{\beta}}$ et $\alpha = a \Rightarrow \hat{a} = \hat{\alpha}$.

Par exemple, on obtient l'ajustement ci-dessous si on applique cette méthode aux données suivantes.

X	0,5	1,0	1,5	2,0	2,5	3,0	3,5	4,0	4,5	5,0
Y	0,1	0,5	1,4	2,7	5,1	7,6	11,2	15,9	22,3	28,1



3.4 Exercices du chapitre 3

Exercice 1

On donne le tableau de répartition suivant

X : nombre de fréquentations hebdomadaires d'un magasin,

Y : montant des achats

X \ Y	[0,50[	[50,100[	[100,200[
1	40	60	150
2	60	90	140
3	80	70	60
4	220	20	10

1. Calculer les distributions jointes et marginales en fréquences.
2. Calculer les moyennes et variances de ces distributions marginales.
   Conclure sur l'indépendance de ces distributions.
3. Calculer les distributions conditionnelles de $X / Y = 25$ et $Y / X = 3$.
   Calculer les moyennes et variances de ces distributions.

Exercice 2

On considère le tableau des fréquences suivant, relatif à un groupe de personnes, tel que :

X : L'âge de chaque individu,

Y : Le poids de chaque individu.

X \ Y	[40,50[	[50,60[	[60,80[
[20,30[	0,09	0,30	0,21
[30,50[	0,06	0,20	0,14

1. Calculer les distributions marginales en fréquences.
2. Est-ce que les variables X et Y sont indépendantes. Et pourquoi ?
3. Calculer les distributions conditionnelles de $X / Y = 45$ et $Y / X = 40$.
4. En comparant la distribution conditionnelle de $X / Y = 45$ avec la distribution marginale de X; et la distribution conditionnelle de $Y / X = 40$ avec la distribution marginale de Y, que peut on déduire et pour quoi ?

Exercice 3

Le tableau suivant donne le nombre fabriqué d'un produit Y dans une entreprise en fonction du nombre d'employés X.

X (*Nombre d'employés*)	1	2	3	4	5	6	7	8
Y (*Nombre fabriqué du produit en unités*)	3	10	20	32	40	46	52	60

1. Représenter le nuage de points (x_i, y_i).
2. Calculer la covariance entre le nombre d'employés X et le nombre fabriqué du produit Y. Que peut-on déduire sur la relation entre X et Y.
3. Calculer le coefficient de corrélation linéaire $r_{x,y}$.
4. Conclure sur l'intensité de la liaison entre les deux variables X et Y.
5. Déterminer l'équation de la droite des moindres carrés ordinaires du nombre fabriqué du produit Y en fonction du nombre d'employés X.
6. Interpréter la pente de l'équation de la droite obtenue.
7. Déduire le nombre du produit fabriqué si l'entreprise a 16 employés.
8. Déterminer l'équation de la droite des moindres carrés ordinaires du nombre fabriqué du produit Y en fonction du nombre d'employés X si on est certain que si $X = 0$ alors $Y = 0$.

***Remarque** : Pour l'exercice* **3** *Les calculs se feront 3 chiffres après la virgule avec arrondissement.*

Exercice 4

Considérons les données suivantes sur le prix et les quantités vendues d'un certain bien.

Quantités Y	104	58	37	22	12	9
Prix X	95	130	148	210	250	330

1. Représenter le nuage de points (x_i, y_i).
2. Compte tenu de cette représentation, donner la forme de l'ajustement de ce nuage de points et retrouver la relation entre les deux variables.
3. Donner une estimation de la demande lorsque le prix du bien est égal à 50 puis lorsque le prix est égal à 300.

2ème Partie : Travaux dirigés

Solutions des exercices

4.1 Solutions des exercices du chapitre 1

Exercice 1

1.

1.1 La population étudiée est le groupe de personnes.

1.2 La variable étudiée est X = «nombre d'enfants que chacun d'entre eux avait au 31 décembre 2016 ».

2.

2.1 La nature de la variable étudiée est quantitative discrète.

2.2 L'ensemble **M** des modalités est

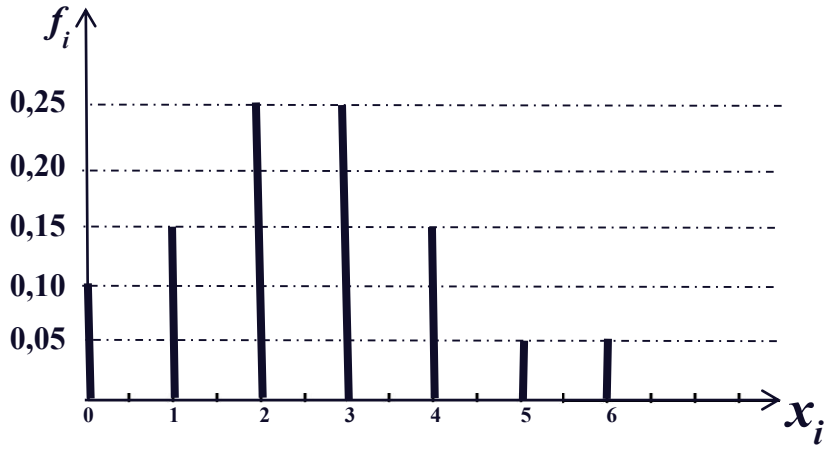
M = {0 ; 1 ; 2 ; 3 ; 4 ; 5 ; 6},

ie $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$, $x_6 = 5$, $x_i = 6$.

3. Représentez la distribution des fréquences par un diagramme en bâtons.
Le tableau statistique associé à X est le suivant.

x_i	n_i	$N_i = n_{ic}$	f_i	$F_i = f_{ic}$
0	**2**	**2**	**0,10**	**0,10**
1	**3**	**5**	**0,15**	**0,25**
2	**5**	**10**	**0,25**	**0,50**
3	**5**	**15**	**0,25**	**0,75**
4	**3**	**18**	**0,15**	**0,90**
5	**1**	**19**	**0,05**	**0,95**
6	**1**	**20**	**0,05**	**1**
Total	**20**		**1,00**	

Et le diagramme en bâtons des fréquences est :



Pour les questions **4** et **5** voir tableau suivant :

x_i	n_i	$N_i = n_{ic}$	f_i	$F_i = f_{ic}$	$N'_i = n_{id}$	$F'_i = f_{id}$
0	**2**	**2**	**0,10**	**0,10**	**20**	**1,00**
1	**3**	**5**	**0,15**	**0,25**	**18**	**0,90**
2	**5**	**10**	**0,25**	**0,50**	**15**	**0,75**
3	**5**	**15**	**0,25**	**0,75**	**10**	**0,50**
4	**3**	**18**	**0,15**	**0,90**	**5**	**0,25**
5	**1**	**19**	**0,05**	**0,95**	**2**	**0,10**
6	**1**	**20**	**0,05**	**1**	**1**	**0,05**
Total	**20**		**1,00**			

Exercice 2

1.

1.1 La population étudiée est le groupe de 50 personnes.

1.2 La variable étudiée est X = « le dernier diplôme obtenu ».

2.

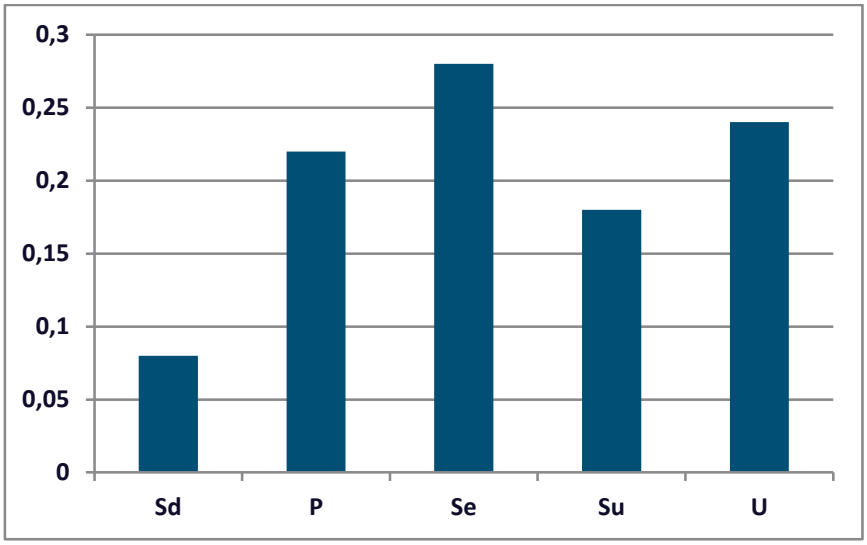
2.1 La nature de la variable étudiée est qualitative ordinale.

2.2 L'ensemble **M** des modalités est **M** = { ***Sd*** ; ***P***; ***Se*** ; ***Su***; ***U*** }.

3. Le tableau statistique complet de cette distribution :

x_i	n_i	$N_i = n_{ic}$	f_i	$F_i = f_{ic}$
Sd	**04**	**04**	**0,08**	**0,08**
P	**11**	**15**	**0,22**	**0,30**
Se	**14**	**29**	**0,28**	**0,58**
Su	**09**	**38**	**0,18**	**0,76**
U	**12**	**50**	**0,24**	**1,00**
Total	**50**		**1,00**	

4. Le diagramme à bandes de la distribution des fréquences est :



Exercice 3

1.

1.1 La population étudiée est composée des familles de la ville.

1.2 La variable étudiée est

X = «le nombre de pièces des appartements des familles de la ville».

1.3 La nature de la variable étudiée est quantitative discrète.

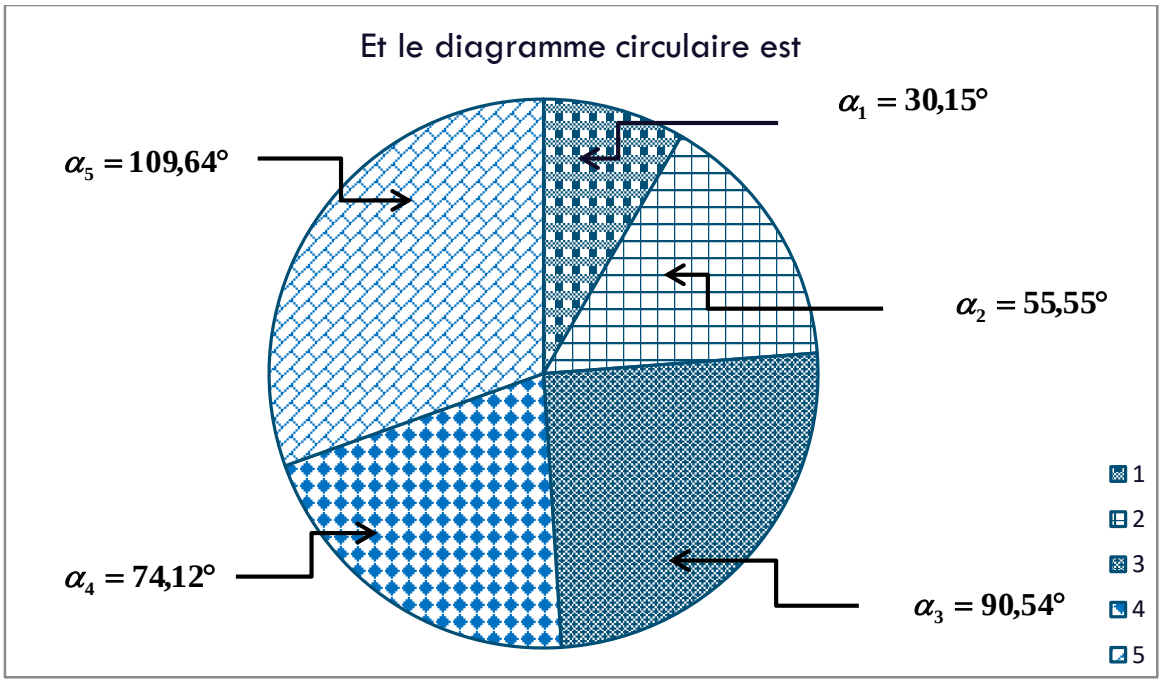
1.4 L'ensemble **M** des modalités est $\mathbf{M} = \{1 ; 2 ; 3 ; 4 ; 5\}$.

2. Pour représenter la distribution par diagramme circulaire on a :

$$\alpha_i = f_i \times 360° = \frac{n_i}{N} \times 360°$$

Donc

x_i	n_i	$N_i = n_{ic}$	f_i	α_i
1	25125	25125	0,08375	30,15
2	46290	71415	0,1543	55,548 ≈ 55,55
3	75453	146868	0,25151	90,5436 ≈ 90,54
4	61767	208635	0,20589	74,1204 ≈ 74,12
5	91365	300000	0,30455	109,638 ≈ 109,64
Total	300000		1,00	360



3.

3.1 Calculons la fonction de répartition :

Par définition la fonction de répartition est :

$$F(x) = \begin{cases} 0 & si \quad x < x_1 \\ F_i & si \quad x_i \le x < x_{i+1} \\ 1 & si \quad x_r \le x \end{cases}$$

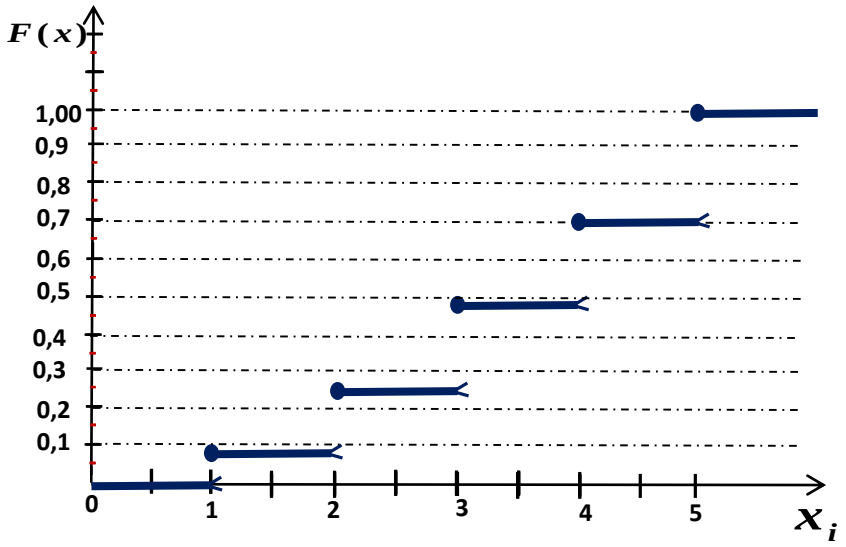
Donc on abesoin de calculer les fréquences cumulées croissantes $F_i = f_{ic}$ avec $i \in \{1,2,3,4,5\}$ ($r = 5$).

$$F_1 = f_1, F_2 = f_1 + f_2 \quad et \quad F_i = f_{ic} = f_1 + f_2 + \cdots + f_i = \sum_{p=1}^{i} f_p$$

x_i	n_i	$N_i = n_{ic}$	f_i	$F_i = f_{ic}$
1	25125	25125	0,08375	0,08375
2	46290	71415	0,1543	0,23805
3	75453	146868	0,25151	0,48956
4	61767	208635	0,20589	0,69545
5	91365	300000	0,30455	1,00
Total	300000		1,00	

$$\Rightarrow \quad F(x) = \begin{cases} 0 \quad si \quad x < 1 \\ 0{,}08375 \quad si \quad 1 \le x < 2 \\ 0{,}23805 \quad si \quad 2 \le x < 3 \\ 0{,}48956 \quad si \quad 3 \le x < 4 \\ 0{,}69545 \quad si \quad 4 \le x < 5 \\ 1 \quad si \quad x \ge 5 \end{cases} \approx \begin{cases} 0 \quad si \quad x < 1 \\ 0{,}08 \quad si \quad 1 \le x < 2 \\ 0{,}24 \quad si \quad 2 \le x < 3 \\ 0{,}49 \quad si \quad 3 \le x < 4 \\ 0{,}70 \quad si \quad 4 \le x < 5 \\ 1 \quad si \quad x \ge 5 \end{cases}$$

Donc la courbe cumulative est



3.2 Nombre d'appartement de cette ville qui sont composées d'au moins 3 pièces ? au plus 4 pièces.

a. « Au moins 3 pièces » corréspond aux appartements qui ont 3, 4, 5 pièces ; donc au nombre des familles habittant dans ces appartement : c'est à dire toutes les familles de la villes sauf celles qui présentent les modalités x_1 et x_2 , autrement dit sauf celles qui ont 1, 2 pièces :

D'où le nombre des appartement est l'effectif cumulé décroissant

$N'_3 = n_{3d} = N - (n_1 + n_2) = 228585$

$N'_3 = n_{3d} = N - (n_1 + n_2) = 300000 - (25125 + 46290) = 228585$

b. « Au plus 4 pièces » corréspond aux appartements qui ont 1, 2, 3, 4 ; donc au nombre des familles qui présentent les modalités x_1, x_2, x_3 et x_4 :
D'où le nombre des appartement est l'effectif cumulé croissant
$N_4 = n_1 + n_2 + n_3 + n_4 = 25125 + 46290 + 75453 + 61767 = 208635.$

Exercice 4

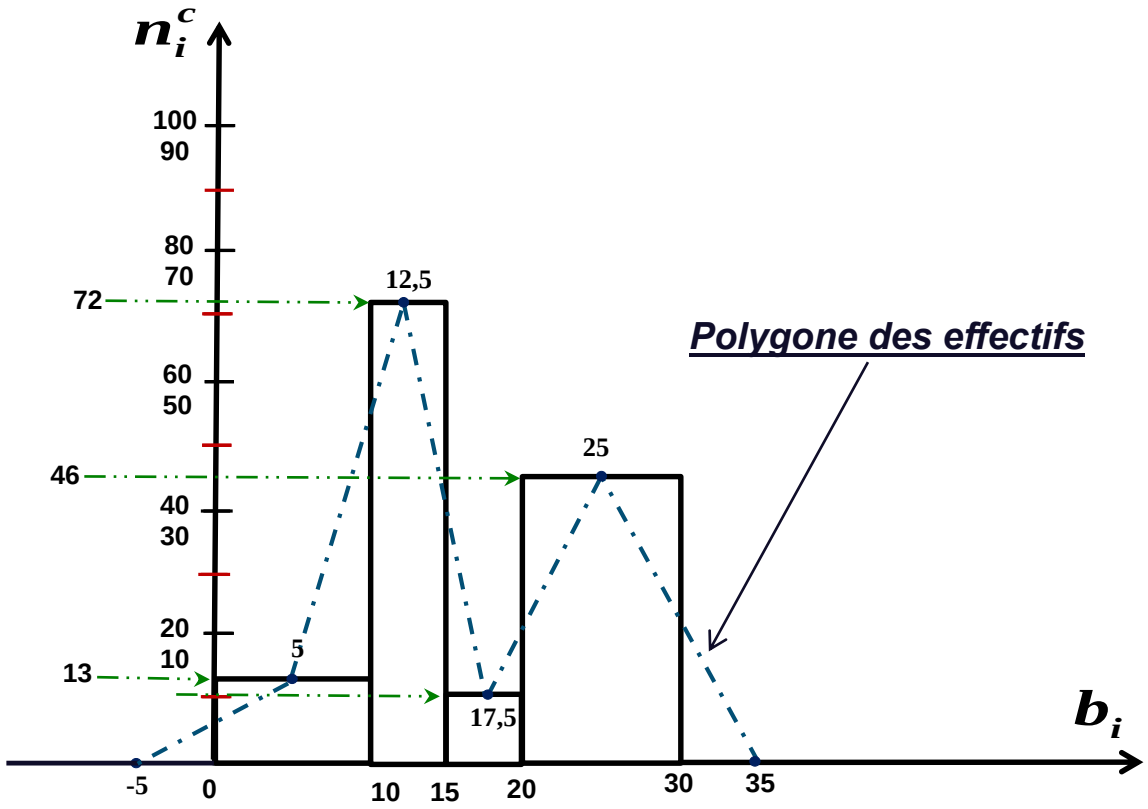
1.1 La population étudiée est composée des 100 athlètes.
1.2 La variable étudiée est X = «sur le temps d'entrainement par semaine».

1.3 La nature de la variable étudiée est quantitative continue.
1.4 Les modalités se sont les valeurs situées dans l'intervalle **[0 , 30[**.

2. Le tableau statistique des effectifs associé à cette série statistique est

Temps de trajet	n_i	f_i	F_i	c_i	N_i	N_i'	F_i'	a_i	d_i	n_i^c
[0 ,10 [	13	0,13	0,13	5	13	100	1,00	10	1,3	13
[10 , 15 [	36	0,36	0,49	12,5	49	87	0,87	5	7,2	72
[15 , 20 [	5	0,05	0,54	17,5	N_i = 54	51	0,51	5	1	10
[20 , 30 [	46	0,46	1,00	25	100	46	0,46	10	4,6	46
Total	100									

3. l'histogramme des effectifs de la distribution (On prend a^*=10).



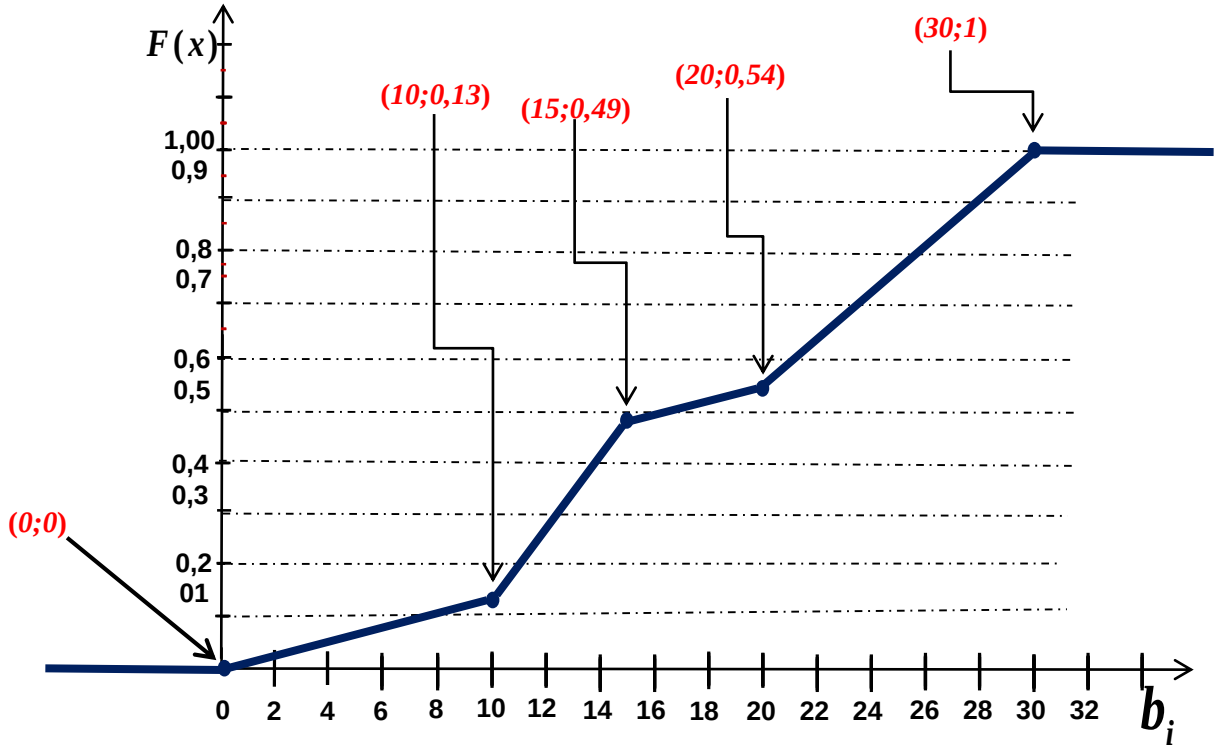
4. Le polygone des effectifs (voir l'histogramme).

5. La courbe cumulative de la série statistique

Donc les points (b_i , F_i) sont :

$(b_0 ; F_0) = (0 ; 0)$; $(b_1 ; F_1) = (10 ; 0{,}13)$; $(b_2 ; F_2) = (15 ; 0{,}49)$;

$(b_3 , F_3) = (20 ; 0{,}54)$; $(b_4 ; F_4) = (30 ; 1{,}00)$ .Donc **la courbe cumulative** est :



4.2 Solutions des exercices du chapitre 2

Exercice 1

1. La valeur du mode de cette série statistique est $Mo = 14$, car c'est la valeur de la variable ayant la fréquence la plus élevée.

2. Le tableau statistique des fréquences associé à cette série statistique est

x_i	f_i	$f_i\, x_i$	$f_i\, x_i^2$
8	0,08	0,64	5,12
9	0,24	2,16	19,44
10	0,08	0,80	8,00
11	0,12	1,32	14,52
12	0,16	1,92	23,04
14	0,28	3,92	54,88
16	0,04	0,64	10,24
Total	1,00	11,4	135,24

3. La proportion en pourcentage (%) des élèves qui ont une note superieure à 11 $(X > 11)$ corréspond à la fréquence cumulée décroissante d'ordre 5 $(x_5 = 12)$, c'est-à-dire :

$F'_5 = f_{5d} = 1 - (f_1 + f_2 + f_3 + f_4) = f_5 + f_6 + f_7 = 0{,}16 + 0{,}28 + 0{,}04 = 0{,}48$ donc 48%.

4. La moyenne de cette série statstique est

$$\bar{x} = \sum_{i=1}^{i=r} \frac{n_i x_i}{N} = \sum_{i=1}^{i=r} f_i x_i = 11{,}4$$

5. calcul de l'effectif total ainsi que l'effectif de chaque modalité si $\sum_{i=1}^{7} n_i x_i = 570$

On a

$$\bar{x} = \sum_{i=1}^{i=r} \frac{n_i x_i}{N} = \sum_{i=1}^{i=r} f_i x_i = 11{,}4 \Rightarrow N = \frac{\sum_{i=1}^{7} n_i x_i}{\bar{x}} = \frac{570}{11{,}4} = 50$$

$$\Rightarrow n_i = N \times f_i \Rightarrow n_i = N \times f_i$$

$$\Rightarrow n_1 = N \times f_1 = 50 \times 0{,}08 = 4;\quad n_2 = N \times f_2 = 50 \times 0{,}24 = 12;\quad n_3 = N \times f_3 = 50 \times 0{,}08 = 4;$$

$$n_4 = N \times f_4 = 50 \times 0{,}12 = 6;\quad n_5 = N \times f_5 = 50 \times 0{,}16 = 8;$$

$$n_6 = N \times f_6 = 50 \times 0{,}28 = 14;\quad n_7 = N \times f_7 = 50 \times 0{,}04 = 2$$

Exercice 2

1. Calculez le prix moyen d'une baguette.

On a : $\bar{x} = \frac{1}{N} \sum_{i=1}^{8} n_i x_i$

x_i	n_i	N_i	$n_i . x_i$
0,55	4	4	2,2
0,60	14	18	8,4
0,65	26	44	16,9
0,7	11	55	7,7
0,75	7	62	5,25
0,8	12	74	9,6
0,85	7	81	5,95
0,9	5	86	4,5
Total	86		60,5

$$\Rightarrow \bar{x} = \frac{1}{N} \sum_{i=1}^{8} n_i x_i = \frac{60{,}5}{86} = 0{,}70$$

2. Calcul du prix médian d'une baguette.

Dans ce cas l'effectif total est $N = 86$ qui est un nombre pair avec $\frac{N}{2} = 43$,

$N_2 = 18$ et $N_3 = 44$; donc :

$$Mé = \frac{v_{43} + v_{44}}{2} = \frac{0{,}65 + 0{,}65}{2} = 0{,}65$$

3.1 Pour Q_1 $N \times p = 86 \times \frac{1}{4} = 21{,}5$ qui n'est pas un nombre entier, donc

$$Q_1 = v_{\lceil 21{,}5 \rceil} = v_{22} = 0{,}65$$

3.2 Pour Q_3 $N \times p = 86 \times \frac{3}{4} = 64{,}5$ qui n'est un nombre entier, donc

$$Q_3 = v_{\lceil 64{,}5 \rceil} = v_{65} = 0{,}80$$

4. l'étendue de la série est : $E = x_{max} - x_{min} = 0{,}9 - 0{,}55 = 0{,}35.$

5.

x_i	n_i	N_i
0,55	4	4
0,60	14	18
0,65	4	22
0,7	11	33
0,75	7	40
0,80	8	48
0,85	11	59
0,9	5	64
Total	64	

5.1 Dans ce cas l'effectif total est $N = 64$ qui est un nombre pair, donc

$$Mé = \frac{v_{32} + v_{33}}{2} = \frac{0,70 + 0,70}{2} = 0,70$$

5.2 Pour Q_1 $N \times p = 64 \times \frac{1}{4} = 16$ qui est un nombre entier, donc

$$Q_1 = \frac{v_{16} + v_{17}}{2} = \frac{0,60 + 0,60}{2} = 0,60$$

5.3 Pour Q_3 $N \times p = 64 \times \frac{3}{4} = 48$ qui est un nombre entier, donc

$$Q_3 = \frac{v_{48} + v_{49}}{2} = \frac{0,80 + 0,85}{2} = 0,825.$$

Exercice 3

1. la classe du premier quartile Q_1 est $[15 , 20[$ car $0,25 < 0,30 \leq F_3$, l'ordre de Q_1 est $p = 0,25$ et $F_2 = 0,20$.

2. Calculer F_3 sachant que le premier quartile $Q_1 = 16$. On a :

$$Q_1 = b_{i-1} + a_i \left(\frac{0,25 - F_{i-1}}{F_i - F_{i-1}} \right) = 15 + 5 \left(\frac{0,25 - 0,20}{F_3 - 0,20} \right)$$

Temps de trajet	*Nombre d'élèves* n_i	f_i	F_i	c_i	$f_i c_i$	$f_i c_i^2$	a_i
[0 ,10 [	4	0,04	0,04	5	0,20	1,00	10
[10 , 15 [	16	0,16	0,20	12,5	2,00	25,00	5
[15 , *20* [	25	0,25	0,45	17,5	4,375	76,5625	5
[*20* , 25 [	37	0,37	0,82	22,5	8,325	187,3125	5
[25 , 30 [	18	0,18	1,00	27,5	4,950	136,125	5
Total	100				19,85	426	

$$\Rightarrow 16 = 15 + 5 \frac{0,05}{F_3 - 0,20} \Rightarrow F_3 = 0,20 + 0,25 \Rightarrow F_3 = 0,45$$

3. Pour le calcul des fréquences si $F_3 = 0,45$ voir le tableau.

4. La moyenne arithmétique est $\bar{x} = \sum_{i=1}^{5} f_i c_i = 19{,}85$

5. Le calcul de l'écart-type

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\sum_{i=1}^{5} f_i c_i^2 - (\bar{x})^2}$$

$$\Rightarrow \sigma(x) = \sqrt{426 - (19{,}85)^2} \Rightarrow \sigma(x) = 5{,}6549$$

6. Le calcul de l'effectif total ainsi que l'effectif de chaque classe si $\sum_{i=1}^{5} n_i c_i^2 = 42600$.

$$\left(\sum_{i=1}^{5} f_i c_i^2\right) = \frac{1}{N}\left(\sum_{i=1}^{5} n_i c_i^2\right) \Rightarrow N = \frac{\left(\sum_{i=1}^{5} n_i c_i^2\right)}{\left(\sum_{i=1}^{5} f_i c_i^2\right)} = \frac{42600}{426} = 100 \Rightarrow N = 100$$

$\Rightarrow n_i = N \times f_i$

$\Rightarrow n_1 = N \times f_1 = 100 \times 0{,}04 = 4,\ n_2 = N \times f_2 = 100 \times 0{,}16 = 16,\ n_3 = N \times f_3 = 100 \times 0{,}25 = 25,$
$n_4 = N \times f_4 = 100 \times 0{,}37 = 37,\ et\ n_5 = N \times f_5 = 100 \times 0{,}18 = 18,$

Exercice 4

1. Determination des effectifs n_6 et n_8.

L'enquete porte sur **1000** commerçant (N = 1000).

Le nombre moyen d'heures d'ouverture hebdomadaires est égal à **40,38** ($\bar{x} = 40{,}38$).

Donc

$$\begin{cases} N = 1000 \\ \bar{x} = \frac{1}{N}\sum_{i=1}^{5} n_i c_i = 19{,}85 \end{cases} \Rightarrow \begin{cases} n_6 + n_8 = 250 \\ 42n_6 + 47{,}5n_8 = 11050 \end{cases} \Rightarrow n_6 = 150,\ \ n_8 = 100.$$

Nombre d'heures	n_i	a_i	$n_i^c = d_i$	c_i	$n_i . c_i$	f_i	$(c_i)^2$	$f_i(c_i)^2$	N_i
[30 , 35[	50	5	10	32,5	1625	0,05	1056,25	52,8125	50
[35 , 37[	100	2	50	36	3600	0,10	1296,00	129,6000	150
[37, 39[	200	2	100	38	7600	0,20	1444,00	288,8000	350
[39, 40[	150	1	150	39,5	5925	0,15	1560,25	234,0375	500
[40 , 41[	120	1	120	40,5	4860	0,12	1640,25	196,8300	620
[41 , 43[	n_6 = 150	2	75	42	$n_6 \times 42$	0,15	1764,00	264,6000	770
[43, 45[	130	2	65	44	5720	0,13	1936,00	251,6800	900
[45, 50[	n_8 = 100	5	20	47,5	$n_8 \times 47{,}5$	0,10	2256,25	225,6250	1000
Total	1000				29330 + $n_6 \times 42$ + $n_8 \times 47{,}5$	1,00		1643,9850	

2.1 Calacul du mode avec $a^* = 1$

Comme $n_4^c = 150$ est l'effectif corrigé le plus élevé, alors la calsse modale est **[39 , 40[** et on a :

$$Mo = b_{i-1} + a_i \frac{m_1}{m_1 + m_2}, \quad m_1 = n_i^c - n_{i-1}^c \quad m_2 = n_i^c - n_{i+1}^c$$

$$\Rightarrow Mo = 39 + \left(\frac{50}{50+30}\right) = 39{,}625h$$

2.2 Calacul de la mediane de cette distribution

On a $N = 1000 = 2\times500$, $N_3 = 350$ et $N_4 = 500$ donc la classe médiane est **[39 , 40[** et on :

$$Mé = b_{i-1} + a_i\left(\frac{\frac{N}{2} - N_{i-1}}{N_i - N_{i-1}}\right) = 39 + 1\times\left(\frac{500-350}{500-350}\right) = 40$$

3. En prenant $\bar{x} = 40{,}38h$, on aura :

$$V(X) = \sum_{i=1}^{8} f_i(c_i)^2 - \bar{x} = 13{,}4406 \Rightarrow \sigma(X) = 3{,}6661$$

4.1 Calcul Q_1 :

L'ordre de Q_1 est $p = 0{,}25$ donc $\lceil N\times p\rceil = 250$, $N_2 = 150$ et $N_3 = 350$ donc la classe du premier quartile Q_1 est **[37, 39[** et on a :

$$Q_1 = b_{i-1} + a_i\left(\frac{N/4 - N_{i-1}}{N_i - N_{i-1}}\right) = 37 + 2\left(\frac{250-150}{350-150}\right) = 38\ h.$$

4.2 Calcul Q_3 :

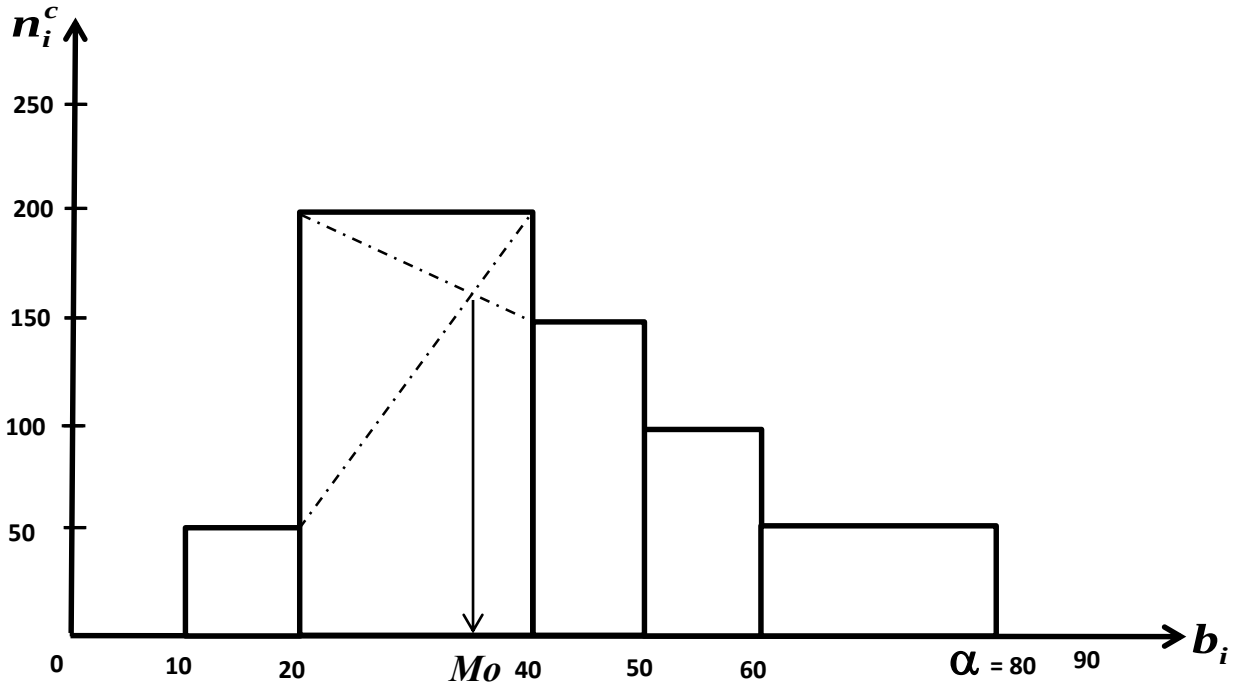
L'ordre de Q_3 est $p = 0{,}75$ donc $\lceil N\times p\rceil = 750$, $N_5 = 620$ et $N_6 = 770$ donc la classe du Troisième quartile Q_3 est **[41 , 43** et on a :

$$Q_3 = b_{i-1} + a_i\left(\frac{N(3/4) - N_{i-1}}{N_i - N_{i-1}}\right) = 41 + 2\left(\frac{750-620}{770-620}\right) = 42{,}75\ h.$$

Exercice 5

1. Représentation graphique du mode M_0 de cette série statistique.

D'après l'histogramme $n_2^c = 200$ est l'effectif corrigé le plus grand alors la calsse modale est **[20 , 40[** donc



2. Le tableau statistique des effectifs associé à cette série (en fonction de α) avec a^*=100 est :

Classes	N_i	$n_i^c = a^*(n_i/a_i)$	a_i	c_i	n_i
[10 , 20[	5	50	10	15	5
[20 , 40[	45	200	20	30	40
[40 , 50[	60	150	10	45	15
[50 , 60[	70	100	10	55	10
[60 , α [	80	50	α − 60 = 20	70	0,5(α−60)=10
Total					80

3. Calcule de α :

On a

$$n_1 + n_2 + n_3 + n_4 + n_5 = 5 + 40 + 15 + 10 + 0{,}5(\alpha - 60) = 80$$

$$0{,}5(\alpha - 60) = 80 - 70 \Rightarrow (\alpha - 60) = \frac{80 - 70}{0{,}5} \Rightarrow \alpha = 20 + 60 = 80$$

4. Calcul mode M_0 de cette série : Comme $n_2^c = 200$ est le plus effectif corrigé alors la calsse modale est [20 , 40[et on a :

$$Mo = b_{i-1} + a_i \frac{m_1}{m_1 + m_2}, \quad m_1 = n_i^c - n_{i-1}^c \quad m_2 = n_i^c - n_{i+1}^c$$

$$\Rightarrow Mo = 20 + 20\left(\frac{200-50}{200-50+200-150}\right) = 35.$$

4.3 Solutions des exercices du chapitre 3

Exercice 1

1. Le calcul des distributions jointes et marginales en fréquences.
La distribution jointe des effectifs est donné par le tableau suivant :

X \ Y	[0,50[	[50,100[	[100,200[	Total
1	40	60	150	250
2	60	90	140	290
3	80	70	60	210
4	220	20	10	250
Total	400	240	360	1000

Calculons d'abord l'effectif total : $N = \sum_{i=1}^{4}\sum_{J+1}^{3} n_{ij} = 1000$; $f_{ij} = \frac{n_{ij}}{N}$

Donc la distribution jointe en fréquences est

X \ Y	[0,50[	[50,100[	[100,200[	*Total*
1	0,040	0,060	0,150	0,25
2	0,060	0,090	0,140	0,29
3	0,080	0,070	0,060	0,21
4	0,220	0,020	0,010	0,25
Total	0,4	0,24	0,36	1

La distribution marginale en fréquence de X est

x_i	$f_{i.}$	$f_{i.}x_i$	$f_{i.}x_i^2$
1	0,25	0,25	0,25
2	0,29	0,58	1,16
3	0,21	0,63	1,89
4	0,25	1	4
Total	1	2,46	7,3

La distribution marginale en fréquence de Y est

y_j	$f_{.j}$	c_j	$f_{.j}\,y_j$	$f_{.j}y_j^2$
[0,50[	0,4	25	10	250
[50,100[	0,24	75	18	1350
[100,200[	0,36	150	54	8100
Total	1		82	9700

2. Le calcul des moyennes et variances de ces distributions marginales.
Pour calculer les moyennes et variances de ces distributions marginales, on peut compléter les deux tableaux relatives à chaque distribution marginale (voir les deux tableaux en haut).

2.1 La moyenne marginale de X :

$$\bar{x} = \sum_{i=1}^{4} f_{i.} x_i = 2{,}46$$

2.2 La variance marginale de X

$$V(x) = (\sum_{i=1}^{4} f_{i.} x_i^2) - \bar{x}^2 = 7{,}3 - (2{,}46)^2 = 1{,}24$$

2.3 La moyenne marginale de Y :

$$\bar{y} = \sum_{j=1}^{3} f_{.j} y_j = 82$$

2.2 La variance marginale de Y

$$V(y) = (\sum_{j=1}^{3} f_{.j} y_j^2) - \bar{y}^2 = 9700 - (82)^2 = 2976$$

2.2 Conclure sur l'indépendance de ces distributions
Les deux variables ne sont pas indépendantes. En effet, on peut remarquer que :

$$f_{11} \neq f_{1.} \times f_{.1} \Leftrightarrow 0{,}040 \neq 0{,}4 \times 0{,}25$$

3. Calcul des distributions conditionnelles de $X / Y = 25$ et $Y / X = 3$.

3.1 Distributions conditionnelles de $X / Y = 25$

$X / Y = 25$	$n_{i/1} = n_{i1}$
1	40
2	60
3	80
4	220
Total	400

3.2 Distribution conditionnelle de $Y / X = 3$

$Y / X = 3$	$n_{j/3} = n_{3j}$
[0,50[	80
[50,100[	70
[100,200[	60
Total	210

3.3 Pour calculer les moyennes et variances de ces distributions conditionnelles, on peut compléter les deux tableaux relatifs à chaque distribution conditionnelle de la même manière que celle de la première question. On peut ainsi vérifier que

$X / Y = 25$	n_{i1}	$n_{i1}\, x_i$	$n_{i1}\, x_i^2$
1	40	40	40
2	60	120	240
3	80	240	720
4	220	880	3520
Total	$N_1 = 400$	1280	4520

$Y / X = 3$	n_{3j}	c_j	$n_{3j}\, c_j$	$n_{3j}\, c_j{}^2$
[0,50[	80	25	2000	50000
[50,100[	70	75	5250	393750
[100,200[	60	150	9000	1350000
Total	$N_2 = 210$		16250	1793750

Ce qui donne

$$\overline{Y / X = 3} = \sum_{j=1}^{j=3} \frac{n_{3j} c_j}{N_2} = \frac{16250}{210} = 77{,}38$$

$$\overline{X / Y = 25} = \sum_{i=1}^{i=4} \frac{n_{i1} x_i}{N_1} = \frac{1280}{400} = 3{,}2$$

$$V(X / Y = 25) = \frac{1}{N_1}\left(\sum_{i=1}^{4} n_{i1} x_i^2\right) - \left(\overline{X / Y = 25}\right)^2 = \frac{4520}{400} - (3{,}2)^2 = 1{,}06$$

$$V(Y / X = 3) = \frac{1}{N_2}\left(\sum_{j=1}^{3} n_{3j} c_j^2\right) - \left(\overline{Y / X = 3}\right)^2 = \frac{1793750}{210} - (77{,}38)^2 = 2554$$

Exercice 2

1. Calculer les distributions marginales en fréquences. On a :

X \ Y	[40,50[	[50,60[	[60,80[	*Total*
[20,30[	0,09	0,30	0,21	0,60
[30,50[	0,06	0,20	0,14	0,40
Total	0 ,15	0,50	0,35	1,00

Donc la distribution marginale de X est :

X	$f_{i.}$
[20,30[	0,60
[30,50[	0,40
Total	1,00

et la distribution marginale de Y est

Y	$f_{.j}$
[40,50[	0,15
[50,60[	0,50
[60,80[	0,35
Total	1,00

2. Est-ce que les variables X et Y sont indépendantes. pourquoi ?

Les variables X et Y sont indépendantes car $\forall i,j \Rightarrow f_{ij} = f_{i.} \times f_{.j}$

en effet ;

$$f_{1.} \times f_{.1} = 0{,}60 \times 0{,}15 = 0{,}09 = f_{11}; \quad f_{1.} \times f_{.2} = 0{,}60 \times 0{,}50 = 0{,}30 = f_{12};$$
$$f_{1.} \times f_{.3} = 0{,}60 \times 0{,}35 = 0{,}21 = f_{13};$$
$$f_{2.} \times f_{.1} = 0{,}40 \times 0{,}15 = 0{,}06 = f_{21}; \quad f_{2.} \times f_{.2} = 0{,}40 \times 0{,}50 = 0{,}20 = f_{22};$$
$$f_{2.} \times f_{.3} = 0{,}40 \times 0{,}35 = 0{,}14 = f_{23};$$

3.1 La distribution conditionnelles de $X/Y = 45$ est :

$X/Y = 45$	$f_{i/1}$	
[20,30[	0,60	car $f_{i/1} = \dfrac{f_{i1}}{f_{.1}}$
[30,50[	0,40	
Total	1,00	

3.2 La distribution conditionnelles de $Y/X = 40$ est :

$Y/X = 40$	$f_{j/2}$	
[40,50[	0,15	car $f_{j/2} = \dfrac{f_{2j}}{f_{2.}}$
[50,60[	0,50	
[60,80[	0,35	
Total	1,00	

4. On déduit que :

la distribution conditionnelle de $X/Y = 45$ et la distribution marginale de X sont identiques.

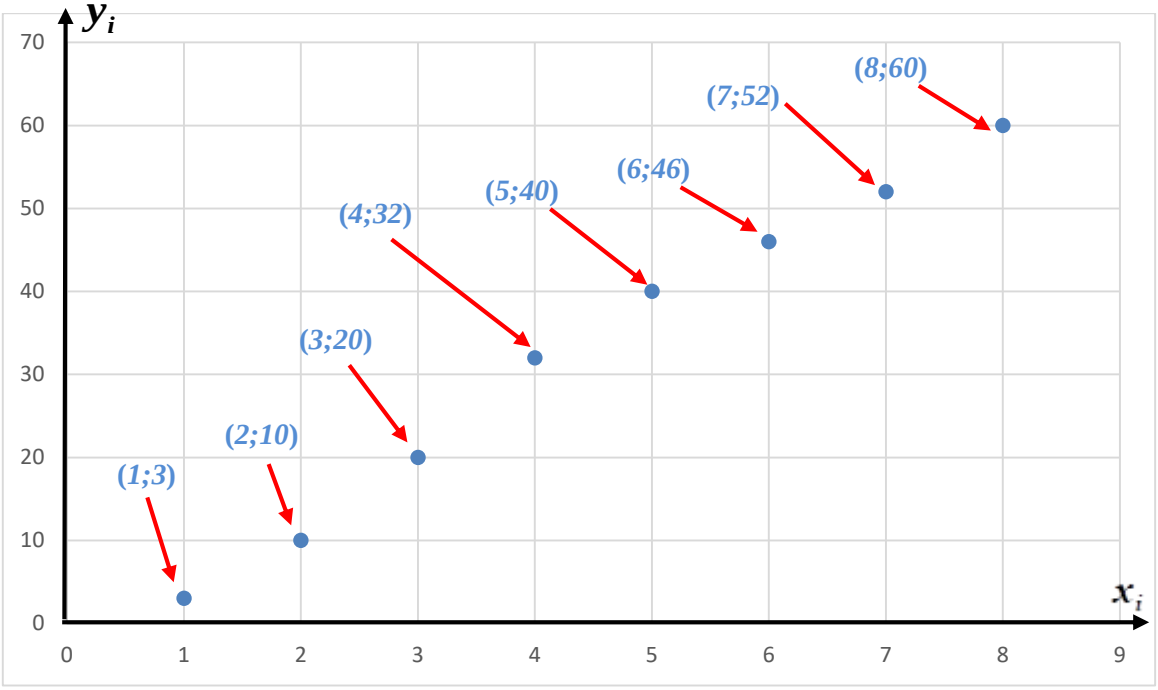
la distribution conditionnelle de $Y/X = 40$ et la distribution marginale de Y sont identiques.

Et cela est dû à l'indépendance des deux variables X et Y, en effet :

$$\forall i,j \Rightarrow f_{ij} = f_{i.} \times f_{.j} \quad \forall i \Rightarrow f_{i/1} = \frac{f_{i1}}{f_{.1}} = \frac{f_{i.} \times f_{.1}}{f_{.1}} = f_{i.} \text{ et } \forall j \Rightarrow f_{j/2} = \frac{f_{2j}}{f_{2.}} = \frac{f_{2.} \times f_{.j}}{f_{2.}} = f_{.j}$$

Exercice 3

1. Le nuage des points de coordonnées (x_i , y_i) est :



2. La covariance entre le nombre d'employés X et le nombre fabriqué du produit Y.

	x_i	y_i	$x_i . y_i$	$(x_i)^2$	$(y_i)^2$
	1	3	3	1	9
	2	10	20	4	100
	3	20	60	9	400
	4	32	128	16	1024
	5	40	200	25	1600
	6	46	276	36	2116
	7	52	364	49	2704
	8	60	480	64	3600
Total	36	263	1531	204	11553

$$\bar{x} = \frac{1}{8}\sum_{1}^{8} x_i = \frac{36}{8} = 4{,}5 \quad \text{et} \quad \bar{y} = \frac{1}{8}\sum_{1}^{8} y_i = \frac{263}{8} = 32{,}875$$

$$Cov(x,y) = \left(\frac{1}{N}\sum_{i=1}^{i=N} x_i y_i\right) - (\bar{x}.\bar{y}) = \frac{1531}{8} - 4{,}5 \times 32{,}875 = 191{,}375 - 147{,}938 = 43{,}437$$

Conclusion : $Cov(x,y) > 0$, donc la relation entre les deux variables est positive et les deux variables dans le même sens.

3. Le coefficient de corrélation linéaire $r_{x,y}$.

$$r_{x,y} = \frac{Cov(x,y)}{\sqrt{V(x)V(y)}}$$

	x_i	y_i	$x_i . y_i$	$(x_i)^2$	$(y_i)^2$
	1	3	3	1	9
	2	10	20	4	100
	3	20	60	9	400
	4	32	128	16	1024
	5	40	200	25	1600
	6	46	276	36	2116
	7	52	364	49	2704
	8	60	480	64	3600
Total	36	263	1531	204	11553

$$V(x)=\left(\frac{1}{N}\sum_{i=1}^{i=r}{x_i}^2\right)-\bar{x}^2=\frac{204}{8}-(4{,}5)^2=25{,}5-20{,}25=5{,}25$$

et $V(y)=\left(\frac{1}{N}\sum_{i=1}^{i=r}{y_i}^2\right)-\bar{y}^2=\frac{11553}{8}-(32{,}875)^2=1444{,}125-1080{,}766=363{,}359$

$$r_{x,y}=\frac{Cov(x,y)}{\sigma(x)\sigma(y)}=\frac{43{,}437}{\sqrt{5{,}25\times 363{,}359}}=\frac{43{,}437}{\sqrt{1907{,}635}}=\frac{43{,}437}{43{,}676}=0{,}995$$

4. Conclure sur l'intensité de la liaison entre les deux variables X et Y.

Conclusion : $r_{x,y}$ est proche de 1 cela traduit qu'il y'a une forte corrélation linéaire positive.

5. L'équation de la droite des moindres carrés ordinaires du nombre fabriqué du produit Y en fonction du nombre d'employés X.

L'équation de la droite recherchée est $y=\hat{a}x+\hat{b}$, telle que la pente de cette droite de régression est obtenue par :

$$\hat{a}=\frac{\sum_{i=1}^{i=8}x_i y_i-8\bar{x}.\bar{y}}{\sum_{i=1}^{i=8}{x_i}^2-8\bar{x}^2}=\frac{Cov(x,y)}{V(x)}$$ et la constante de la droite $\hat{b}=\bar{y}-\hat{a}\bar{x}$

	x_i	y_i	$x_i . y_i$	$(x_i)^2$	$(y_i)^2$
	1	3	3	1	9
	2	10	20	4	100
	3	20	60	9	400
	4	32	128	16	1024
	5	40	200	25	1600
	6	46	276	36	2116
	7	52	364	49	2704
	8	60	480	64	3600
Total	36	263	1531	204	11553

$$\Rightarrow \hat{a}=\frac{1531-8\times 4{,}5\times 32{,}875}{204-8\times(4{,}5)^2}=\frac{1531-1183{,}5}{204-162}=\frac{347{,}5}{42}=8{,}274$$

$\Rightarrow \hat{b} = 32{,}875 - 8{,}274 \times 4{,}5 = -4{,}358$

Donc L'équation s'écrit : $y = 8{,}274x - 4{,}358$

6. Interpréter la pente de l'équation de la droite obtenue.

Interprétation de la pente $\hat{a} = 8{,}274$:

La valeur de la pente de la droite signifie que le nombre d'unités fabriqués du produit augmente de $\hat{a} = 8{,}274$ par employé.

7. Déduire le nombre du produit fabriqué si l'entreprise a 16 employés.

C'est-à-dire $x = 16 \Rightarrow y = 8{,}274 \times 16 - 4{,}358 = 128{,}026 \approx 128$ unités produites.

8. L'équation de la droite des moindres carrés ordinaires du nombre fabriqué du produit Y en fonction du nombre d'employés X si on est certain que si $X = 0$ alors $Y = 0$.

C'est-à-dire la constante de la droite est $b = 0$

Donc les couples (x_i , y_i) vérifient :

$$y_i = ax_i + \varepsilon_i \quad \forall i \in \{1, \cdots, 8\} \Rightarrow \varepsilon_i = y_i - ax_i$$

$$\Rightarrow \sum_{i=1}^{i=8} {\varepsilon_i}^2 = \sum_{i=1}^{i=8} (y_i - ax_i)^2 = f(a)$$

La méthode **MCO** consiste à minimiser $\sum_{i=1}^{i=8} {\varepsilon_i}^2$; et en utilisant la condition de minimisation de cette fonction f par rapport à a on obtient :

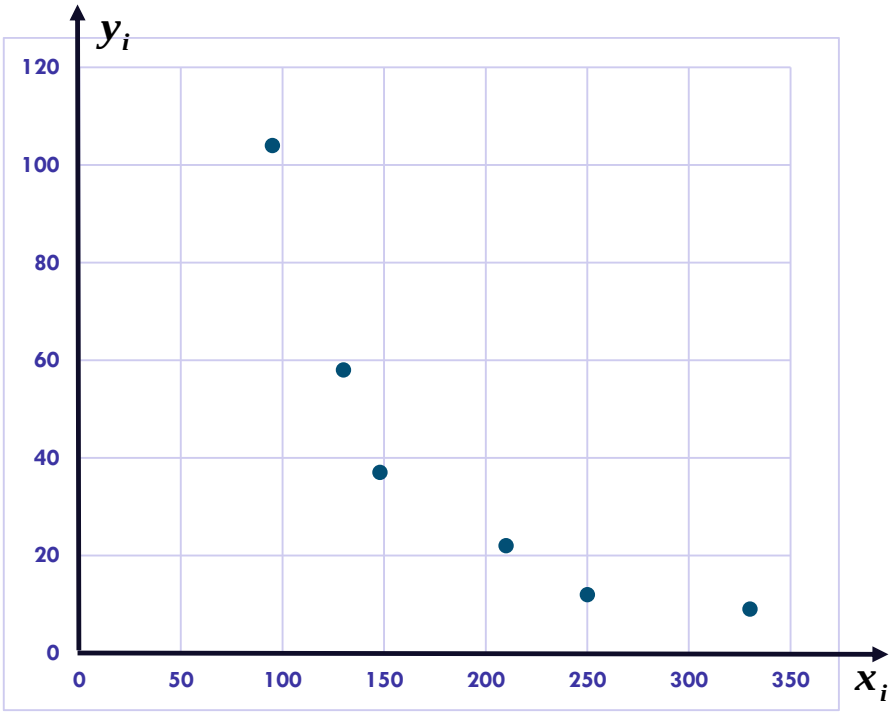
$$f'(a) = \frac{\partial\left(\sum_{i=1}^{i=8} {\varepsilon_i}^2\right)}{\partial a} = 0 \Rightarrow 2\sum_{i=1}^{i=8} (y_i - ax_i)(-x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{i=8} (y_i x_i - a{x_i}^2) = \sum_{i=1}^{i=8} x_i y_i - a\sum_{i=1}^{i=8} {x_i}^2 = 0 \Rightarrow a = \frac{\sum_{i=1}^{i=N} x_i y_i}{\sum_{i=1}^{i=N} {x_i}^2} = \frac{1531}{204} = 7{,}505$$

Donc L'équation s'écrit : $y = 7{,}505x$

Exercice 4

1. Représenter le nuage de points (x_i, y_i).



2. La forme de l'ajustement de ce nuage :

Il est clair que la forme de ce nuage ne suggère pas un ajustement du type $y = ax + b$.

La fonction permettant de représenter ce nuage de points est une fonction hyperbolique du type : $y = \frac{b}{x^a} = bx^{-a}$

Afin d'estimer les coefficients de cette fonction par la méthode des **MCO**, on peu passer d'abord par le logarithme népérien qui nous donne une relation linéaire.

$$y = \frac{b}{x^a} = bx^{-a} \Rightarrow \ln y = \alpha \ln x + \beta$$

Avec $\beta = \ln b$ et $\alpha = -a$

En considérant les deux (02) nouvelles variables et en utilisant la méthode des **MCO**, on peut retrouver α et β, tels que :

$$\hat{\alpha} = \frac{\sum_{i=1}^{i=N} (\ln x_i)(\ln y_i) - N\overline{\ln x}.\overline{\ln y}}{\sum_{i=1}^{i=N} (\ln x_i)^2 - N\overline{\ln x}^2} \quad \text{et} \quad \hat{\beta} = \overline{\ln y} - \hat{\alpha}\overline{\ln x}$$

	x_i	y_i	lnx_i	lny_i	$(lnx_i)(lny_i)$	$(lnx_i)^2$
	95	104	4,554	4,644	21,150	20,738
	130	58	4,868	4,060	19,764	23,693
	148	37	4,997	3,611	18,045	24,972
	210	22	5,347	3,091	16,528	28,592
	250	12	5,521	2,485	13,720	30,487
	330	9	5,799	2,197	12,742	33,629
Total			31,086	20,089	101,949	162,110

Donc : $\overline{\ln x} = \frac{31,086}{6} = 5,181$ et $\overline{\ln y} = \frac{20,089}{6} = 3,348$

$$\hat{\alpha} = \frac{101,949 - 6 \times 5,181 \times 3,348}{162,110 - 6 \times (5,181)^2} \approx -2 \quad \text{et} \quad \hat{\beta} = 3,348 - [(-2) \times 5,181] = 13,71$$

L'équation de la droite de régression de $\ln(y)$ sur $\ln(x)$ est de la forme

$$\Rightarrow \ln(y) = -2\ln(x) + 13,71$$

On peut maintenant retrouver la valeur de a et de b

$$\beta = \ln b \Rightarrow \ln b = 13,71 \Rightarrow b = e^{13,71} = 899864$$

$$\alpha = -a \Rightarrow a = -\alpha = 2$$

Enfin, l'équation de la courbe donnant la relation entre les deux variables est :

$$y = \frac{899864}{x^2} = 899864x^{-2}$$

3. Lorsque le prix du bien est égal à $x = 50$ alors la demande sera :

$$y = \frac{899864}{x^2} = \frac{899864}{(50)^2} \approx 360$$

Donc la demande d'élève à 360 unités.

Lorsque le prix du bien est égal à $x = 300$ alors la demande sera :

$$y = \frac{899864}{x^2} = \frac{899864}{(300)^2} \approx 10$$

Donc la demande sera 10 unités.

Bibliographie

1. B. Verlant, G. Saint-Pierre, Statistiques et probabilités, Manuel de cours Exercices corrigés – Sujets d'examens, Edition BERTI.
2. B. Oukacha, M. Benmessaoud Statistique Descriptive et Calcul des Probabilités, Les Manuels de l'Etudiant, Pages Bleues.
3. Bernard PY, 1996, Statistique descriptive, nouvelle méthode pour bien comprendre et réussir, 4e édition, Economica.
4. Bernard PY, 1994, Exercices corrigés de statistique descriptive, 2e édition, Economica.
5. Alain PILLER, 2004, Statistique Descriptive, éditions Premium.
6. Maurice LETHIELLEUX, 2003, Statistique Descriptive, éditions Dunod, collection « Express ».
7. Fabrice MAZEROLLE, 2003, Statistique Descriptive, Gualino éditeur, EJA-Paris-2006.

Statistique descriptive
Cours et exercices

Février 2023

Ce manuel est conçu spécialement aux étudiants de première année universitaire, ces objectifs pédagogiques sont de favoriser l'apprentissage et développer l'autonomie de tous les étudiants, il est très utile pour l'apprentissage et pour l'entrainement à l'examen, il prend en compte tous les besoins en statistiques des étudiants de première année universitaire, notamment ceux de première année CPST des écoles nationales polytechniques d'Algérie.

Dr Rafik Medjati
Enpo Oran, Algérie