République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 8 Mai 1945 Guelma

Faculté des Mathématiques et de l'Informatique et des Sciences de la Matière

Département de Mathématiques

Cours de Statistique déscriptive et théorie de Probabiltés avec des applicationns

Réaliser par :

Dr . REBIAI Ghania

Niveau : Deuxième année Sciences techniques LMD

Année universitaire : 2022-2023

Statistique déscriptive & théorie de probabilités

Table des matières

Avant propos

Ce document est un support de cours pour les enseignements des probabilités et de la statistique. Il couvre l'analyse combinatoire, le calcul des probabilités. Ce support correspond approximativement aux enseignements en L2 de Sciences Techniques

Un document ne vient jamais du néant. Pour élaborer ce support, je me suis appuyé sur différentes références, des ouvrages reconnus dans la discipline, mais aussi des ressources en ligne qui sont de plus en plus présents aujourd'hui dans la diffusion de la connaissance.

Ce polycopié est constitué de trois chapitres :

► Le premier chapitre est un rappel sur la statistique descriptive. Dans ce chapitre, nous avons introduit les différents terminologies concernant cette partie ainsi on présente les outils pour décrire une série statistique en utilisant les tableaux, les graphes et les paramètres.

► Le deuxième chapitre est consacré à la théorie d'analyse combinatoire qui étudié comment compter les objets et elle fournit des méthodes de dénombrements particulièrement utiles en théorie des probabilités.

► Enfin, le troisième et dernier chapitre est un rappel sur le calcul des probabilités. Dans ce chapitre, nous avons introduit la notion de probabilité et probabilité conditionnelle ainsi que la notion d'indépendance pour les événements et nous avons aussi donné la définition de la probabilité totale et on se termine par la formule de bayes.

Partie 1
Statistique descriptive

1. Statistique déscriptive

1.1. Introduction

L'histoire de la "statistique" remonte à une époque très ancienne. Les activités statistiques (dénombrements) ont commencé bien avant la création du mot, l'application de la méthode et de l'analyse statistique.

Depuis l'antiquité, les Empereurs, les Rois et les Hommes d'Eglise réalisaient des dénombrements de populations humaines et de terres pour les besoins de la guerre et de l'impôt.

- **Il y a plus de 4 ou 5000 ans, il existait déjà en Chine des descriptions chiffrées de la population et de l'agriculture.**
- **Les Égyptiens de l'époque des Pharaons procédaient au dénombrement de la population ou du bétail.**
- **A Rome, l'empereur Auguste fit procéder à une vaste enquête en dénombrant les soldats, les navires et les revenus publics.**

Jusqu'au moyen âge, les seules "statistiques" existante étaient les dénombrements faits dans des buts divers : assiettes de l'impôt, répartition des terres, recrutement dans l'armée est effectués avec des méthodes diverses (recensements des personnes, enregistrements de certains actes d'état civil ...). C'est à partir du XVIII siècle, qu'apparait le mot "statistique" crée par ACHENWAL en 1749 à partir du mot "STATISTA" (politique). Du simple dénombrement de populations humaines et de terres, la statistique est devenue une science qui a retenu et continue de retenir l'attention, non seulement des empereurs et de rois, mais surtout des personnes

de sciences. L'extension et l'utilisation du calcul des probabilités développé par J. BERNOULLI au 18ème siècle et l'application des études démographiques et sociales

ont permis à cette science de connaitre un essor considérable. Ainsi au 19 siècle, de la simple statistique descriptive, elle passe au stade de "Statistique Mathématique".

Depuis le 20ème siècle, les travaux de KARL PEARSON (1857-1936), de STUDENT (WILLIAM SEALY GOSSET, 1876-1937) et de RONAL FISHER (1890-1963) ont permis à cette science de connaitre un développement considérable et une application vaste et variée.

La statistique utilise les techniques et des méthodes de collecte, de présentation, d'étude et d'analyse des données quantitatives. La statistique n'est pas uniquement utilisée pour décrire, pour mieux connaitre un événement survenu dans le passé mais elle intervient de plus en plus dans les travaux de planification, dans le choix de prises de décisions et d'actions.

1.2- Les notions de base de la statistique

1- Elément: C'est une unité qui peut être:

2- Un individu: (êtres vivants : humain, animal, végétal ... ;

3- Un sujet: modules enseignés, les nationalités, les métiers ou professions

4- Population: On appelle population l'ensemble des éléments (individus) sur lequel porte notre étude statistique. Cet ensemble est noté *Ω*

Exemple 1

√ Une population humaine;

√ Une population de plantes;

√ Une population de poissons:

√ Des livres.

5- Echantillon: Pour des raisons techniques ou économiques, il n'est généralement pas possible de collecter des données sur tous les éléments de la population. En outre, si cette opération est possible il est rarement utile de la faire, car l'analyse d'un groupe restreint d'éléments extraits de la population fournit généralement des résultats de précision satisfaisante. Cette petite partie de la population qu'on va examiner

s'appelle « échantillon ».

Exemple 2 :

√ Etude de 20 étudiants pris à partir d'une population de 57.

√ Etude de 5 régions prises à partir d'une population de 25.

√ Etude de 5 modules pris à partir d'une population de 13.

√ Etude de 200 patients pris à partir d'une population de 660.

6- Caractère (variable statistique): C'est une propriété possédée par les unités statistiques (undividus) permettant de les décrire et de les distinguer les unes des autres. En Mathématique: On appel un caractère ou variable statistique (v.s) toute application

$$X: \Omega \rightarrow C$$

L'ensemble C est dit : ensemble des valeurs du caractère X (c'est ce qui est mesuré ou observé sur les individus.

REMARQUE 2.:Toute unité statistique peut être étudiée selon un ou plusieurs caractères.

Exemple 3 :

- Couleur des yeux ;

- Poids des souris ;

- Superficie d'une pièce ;

- La température de l'air.

7- Modalités: Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci..

Exemple 4:

√ Couleur des yeux : vert, bleu, noir ;

√ Poids des souris (en grammes) : 15, 18, 20, 39 ;

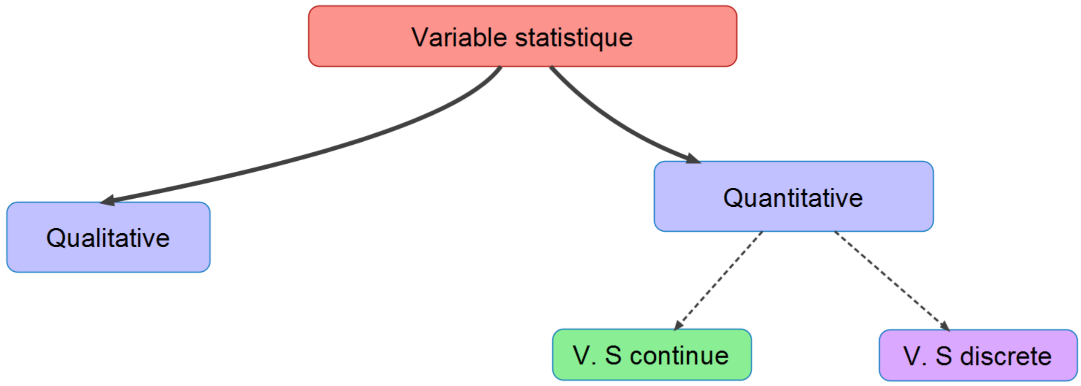
√ Superficie d'une pièce (en mètres) : 3, 5, 6 ;

√ La température de l'air (en °C) : 8, 16, 27, 30, 38.

REMARQUE.3: Le caractère est une variable statistique on la notée par X et ses valeurs appelés les modalités et les notées par x_i

1.3- Types des caractères

Nous distinguons deux catégories de caractères : les caractères qualitatifs et les caractères quantitatifs .



1.3.1-.Caractère qualitatif:

Un caractère est dit qualitatif lorsque ses modalités ne sont pas mesurables. Le nombre de valeurs que peut prendre la variable est limité. Il existe au sein de ce type deux échelles: nominale et ordinale.

1. Echelle nominale : Chaque modalité est exprimée par un nom ou un code. Les différentes modalités ne sont pas ordonnables.

Exemple 5: cas des noms

✓ Etat matrimoniale : marié, célibataire, veuf, divorcé ;

✓ Sexe : féminin, masculin ;

✓ Profession : enseignant, médecin ;

✓ Nationalité : Algérienne, Tunisienne.

Exemple 6: cas des codes

✓ Etat matrimoniale : marié (1), célibataire (2), veuf (3), divorcé (4) ;

✓ Sexe : féminin (1), masculin (2) ;

✓ Profession : enseignant (1), médecin (2) ;

✓ Nationalité : Algérienne (1), Tunisienne (2) ;

2. Echelle ordinale

Chaque modalité est explicitement significative du rang pris par chaque individu pour le caractère considéré.

Exemple 7:

✓ Degré d'intelligence : pas intelligent (0), peu intelligent (1), moyennement intelligent (2), très intelligent (3);

✓ Forme des fruits : petite (1), moyenne (2), grosse (3) ;

1.3.2- Caractère quantitatif

Un caractère est quantitatif si ses modalités s'expriment par des nombres. Le nombre de valeurs peut prendre la variable est illimité. De même, il est partagé en deux sortes de caractères, discret (discontinu) et continu

Exemple 8:

a) salaire d'employés d'une usine.

Modalités : 10000 da , 20000 da...

Type : Discret.

b) La taille d'une personne.

Modalités : [1,20, 1,75] m

Type: continu

1) La variable quantitative discrète: est une variable ne prenant que des valeurs entières (plus rarement décimales).
2) Une variable quantitative est dite continue lorsque les observations qui lui sont associées ne sont pas des valeurs précises, mais des intervalles. C'est le cas lorsque nous avons un grand nombre d'observations distinctes.

1.4- Effectif partiel - effectif cumulé

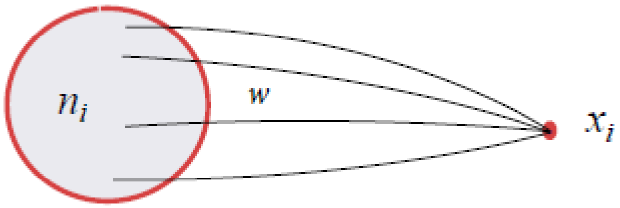
On étudie ici un caractère statistique numérique représenté par une suite x_i décrivant la valeur du caractère avec varie de 1 à k

1.4.1- Effectif partiel (fréquence absolue)

Pour chaque modalité x_i on pose un nombre entier positive n_i par définition

$$n_i = card\{\, \omega \in \Omega : X(\omega) = x_i\}$$

On dit que n_i est le nombre d'individu qui ont la même modalité x_i .



1.4.2- Effectif cumulé

Pour chaque valeur x_i, on pose par définition

$$n_i{}^c \nearrow = n_1 + n_2 + \dots + n_i.$$

L'effectif cumulé $n_i{}^c \nearrow$ d'une valeur est la somme de l'effectif de cette valeur et de tous les effectifs des valeurs qui précèdent.

1.5-.Fréquence partielle - Fréquence cumulée

Typiquement les effectifs n_i. sont grands et il est intéressant de calculer des grandeurs permettant de résumer la série.

1.5.1-. Fréquence partielle (fréquence relative)

Pour chaque valeur x_i, on pose par définition :

$$f_i := \frac{n_i}{N}.$$

f_i s'appelle la fréquence partielle de x_i. La fréquence d'une valeur est le rapport de l'effectif de cette valeur par l'effectif total.

REMARQUE 4

1. **On peut remplacer f_i par $p = f_i \times 100$ qui représente alors un pourcentage.**
2. $\sum_{i=1}^{k} f_i = 1$

1.5.2-. Fréquence cumulée

Pour chaque valeur x_i, on pose par définition

$$f_i^c \nearrow = f_1 + f_2 + \dots + f_i$$

La quantité $f_i^c \nearrow =$ s'appelle la fréquence cumulée de x_i.

REMARQUE 5

1- $f_i^c \nearrow$ est le pourcentage des individus tel que la valeur X est inférieure ou égale à x_i.

2- L'effectif cumulé décroissant est donné par :

$$n_i^c \searrow = n_i + n_{i+1} + \dots + n_k$$

3- La fréquence cumulée décroissante est donnée par

$$f_i^c \searrow = f_i + f_{i+1} + \dots + f_k$$

4- $f_i^c \searrow$ est le pourcentage des individus tel que la valeur X est supérieur ou égale à x_i.

1.6- Représentation graphique des séries statistiques

On distingue les méthodes de représentation d'une variable statistique en fonction de la nature de cette variable (qualitative ou quantitative).

Les représentations recommandées et les plus fréquentes sont les tableaux et les diagrammes (graphe).

Le graphique est un support visuel qui permet:

► La synthèse : visualiser d'un seul coup d'œil les principales caractéristiques (mais on perd une quantité d'informations), (voir Figure 1.1).

► La découverte : met en évidence les tendances.

► Le contrôle : on aperçoit mieux les anomalies sur un graphique que dans un tableau.

► La recherche des régularités: régularité dans le mouvement, répétition du phénomène.

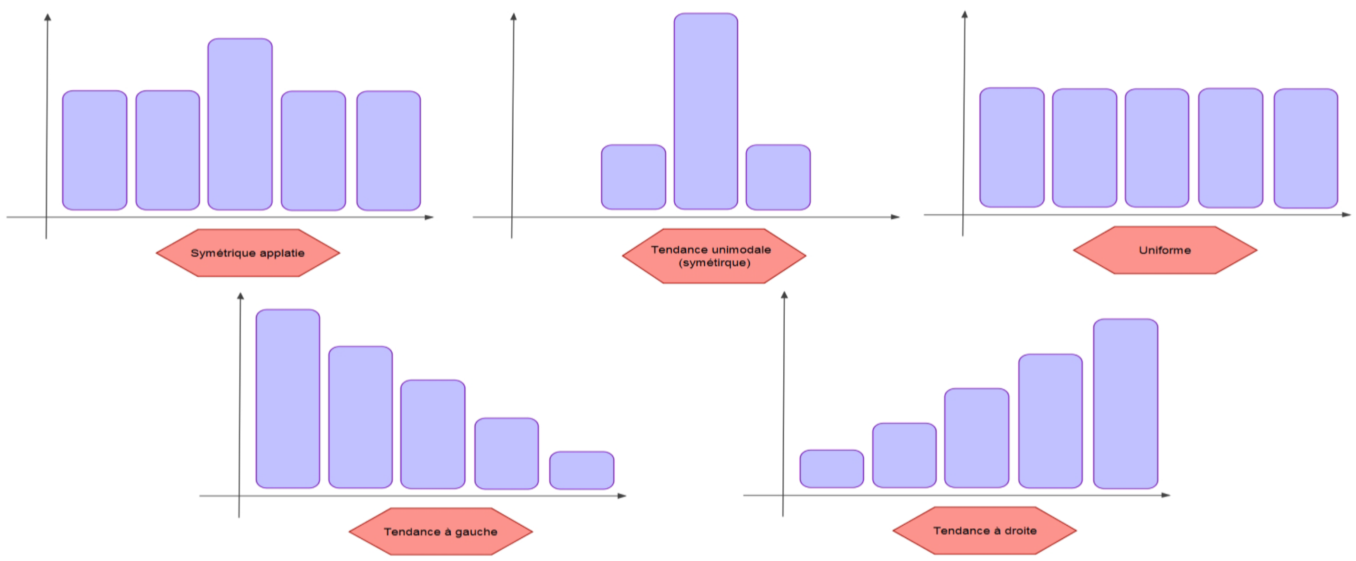


Figure 1.1: Quelques caractéristiques du graphique

1.6.1-Distribution à caractère qualitatif

A partir de l'observation d'une variable qualitative, deux diagrammes permettent de

représenter cette variable : le diagramme en bandes (dit tuyaux d'orgue) et le diagramme circulaire (dit camembert).

a)- Tuyaux d'orgues :

(en barres) est un graphique qui à chaque modalité d'une variable qualitative associe un rectangle de base constante dont la hauteur est proportionnelle à l'effectif ou à la fréquence. Les rectangles sont en général disjoints , (voir Figure 1.2).

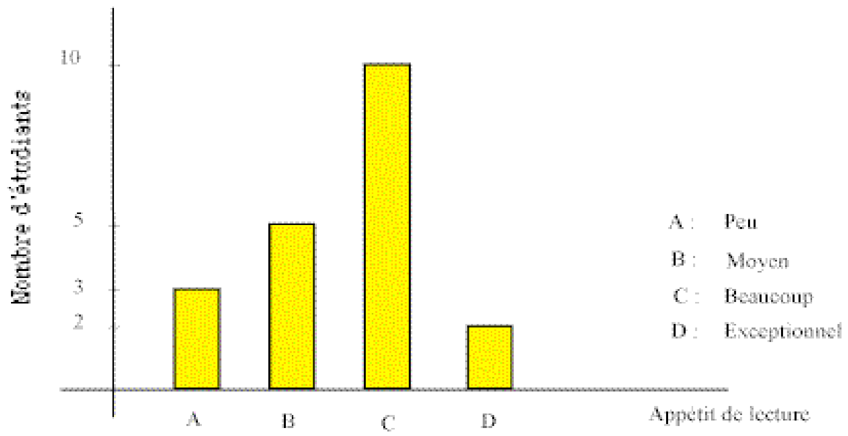


Figure 1 .2 : - Tuyaux d'orgues

b)- Diagramme par secteur (diagramme circulaire)

Les diagrammes circulaires, ou semi-circulaires, consistent à partager un disque ou un demi-disque, en tranches, ou secteurs, correspondant aux modalités observées et dont la surface est proportionnelle à l'effectif, ou à la fréquence, de la modalité, (voir Figure 1.3).

● Le degré d'un secteur est déterminé à l'aide de la règle de trois de la manière suivante :

$$N \rightarrow 360^{\circ}$$

$$n_i \rightarrow \alpha_i$$

Donc :

$$\mathring{\alpha}_i = \frac{n_i}{N} \times 360^{\circ} = f_i \times 360^{\circ}$$

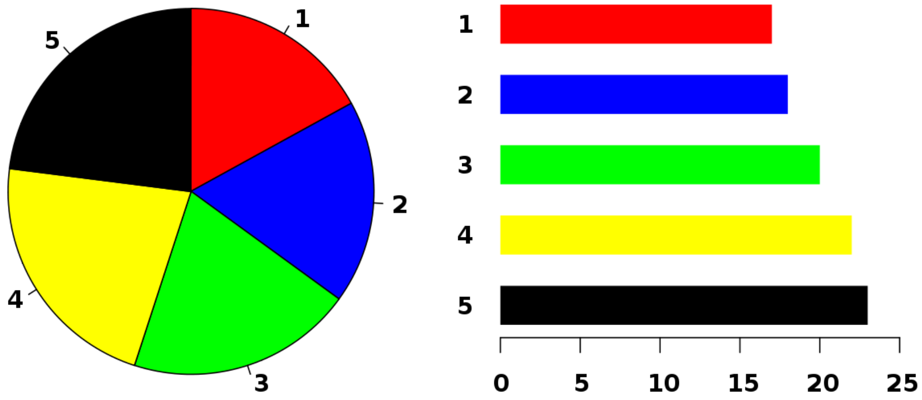


Figure 1.3 : Diagramme circulaire

1.6.2 Distribution à caractère quantitatif discret

A partir de l'observation d'une variable quantitative discrète, deux diagrammes permettent de représenter cette variable : le diagramme en bâtons et le diagramme cumulatif.

1) Diagramme à bâtons

On veut représenter cette répartition sous la forme d'un diagramme en bâtons. À chaque marque correspond un bâton. Les hauteurs des bâtons sont proportionnelles aux effectifs (ou fréquences) représentés. (Voir Figure 1.4).

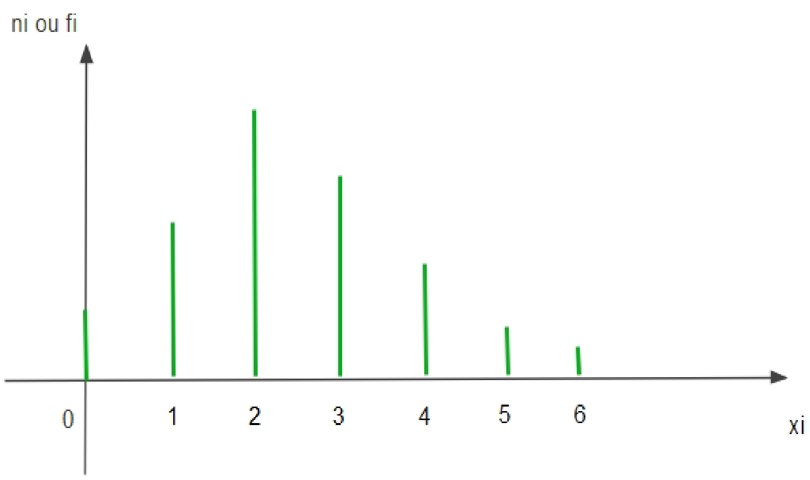


Figure 1.4 : Diagramme en bâtons

2- Polygone des fréquences ou des effectifs. Il s'agit de la ligne brisée joignant les sommets des bâtons du diagramme précédent. Par un segment (Voir Figure 1.5).

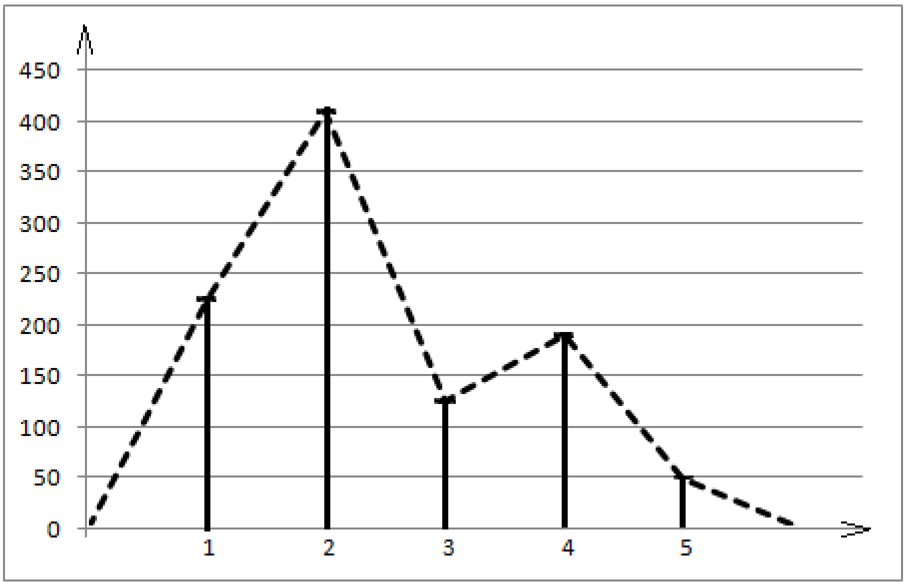


Figure 1.5 : Diagramme en bâtons et le polygone des effectifs

3- Représentation sous forme de courbe et fonction de répartition

Nous avons déjà abordé les distributions cumulées d'une variable statistique. Nous allons dans cette partie exploiter ses valeurs cumulées pour introduire la notion de la fonction de répartition. Cette notion ne concerne que les variables quantitatives.

Soit la fonction $F_X : \mathbb{R} \to [0,1]$ définie par :

$F_x(x) :=$ pourcentage des individus dont la valeur du caractère est $\leq x$.

Cette fonction s'appelle la fonction de répartition du caractère X.

REMARQUE .6

▶ Pour tout $i \in \{1,\ldots, n\}$, on :

$$F_x(x_i) = F_i.$$

▶ La courbe de F_x passe par les points (x_1, F_1), (x_2, F_2), ... et (x_n, F_n).

Cette courbe s'appelle "la courbe cumulative des fréquences".

► **La courbe cumulative est une courbe en escalier représentant les fréquences cumulées relatives.**

►**On peut tracer la courbe cumulative par les effectifs cumulés ou les fréquences cumulées par contre la fonction de répartition il faut la tracer seulement par les fréquences cumulées. . (Voir Figure 1.6).**

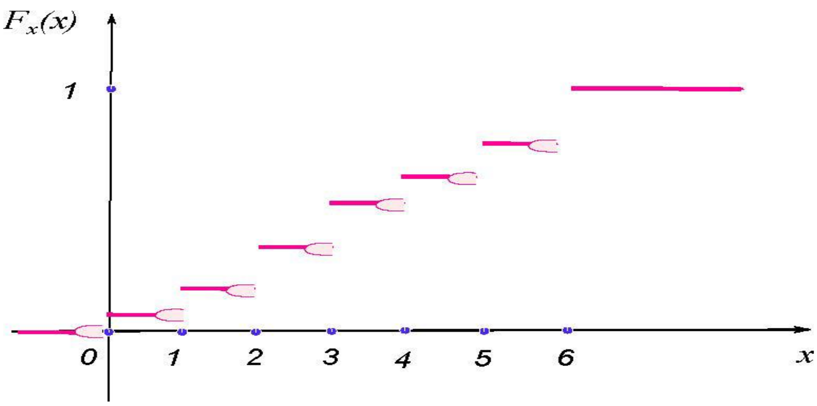


Figure 1.6: Fonction de répartition

Proposition

La fonction de répartition satisfait pour $i \in \{1, 2, \ldots n\}$;

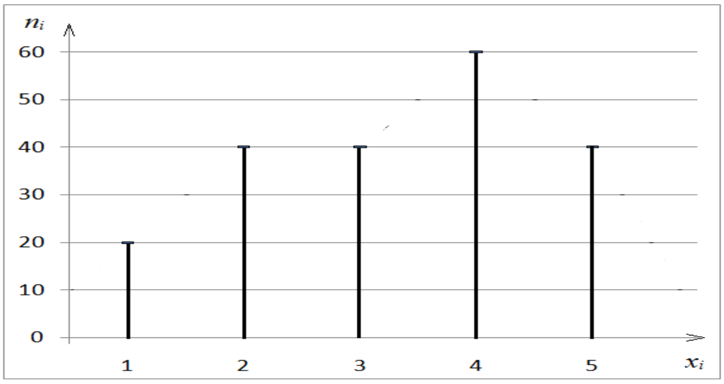
- **L'égalité ;** $F_x(x_i) = F_i$

L'expression ; $F_x(x_i) = \begin{cases} 0 & si \quad s < x_1 \\ F_i & si \quad x_i \leq x < x_{i+1} \\ 1 & si \quad x \geq x_n. \end{cases}$

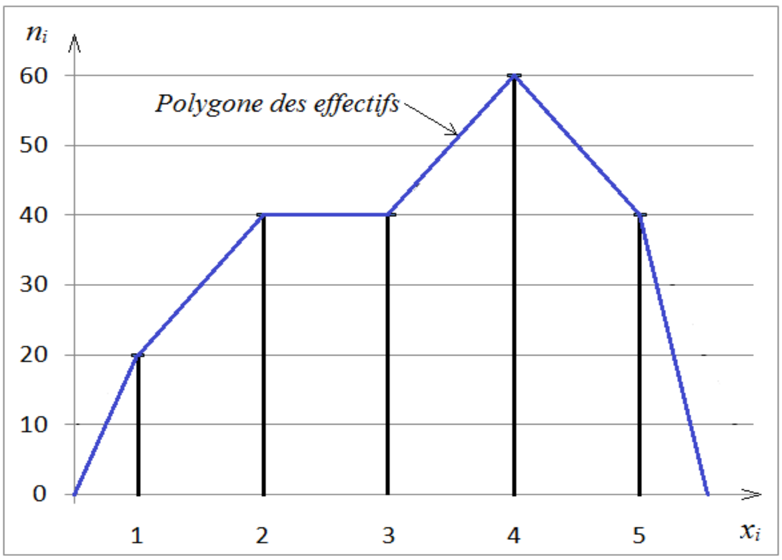
Exemple 9: Une enquête portant sur le nombre d'enfants à charge a été réalisée auprès des habitants d'une citée. Cette enquête a donné les résultats suivants

Nombre d'enfants (x_i)	Nombre de familles (n_i)
1	20
2	40
3	40
4	60
5	40

1) **Représenter graphiquement la distribution.**
2) **La distribution est un diagramme en bâtons**



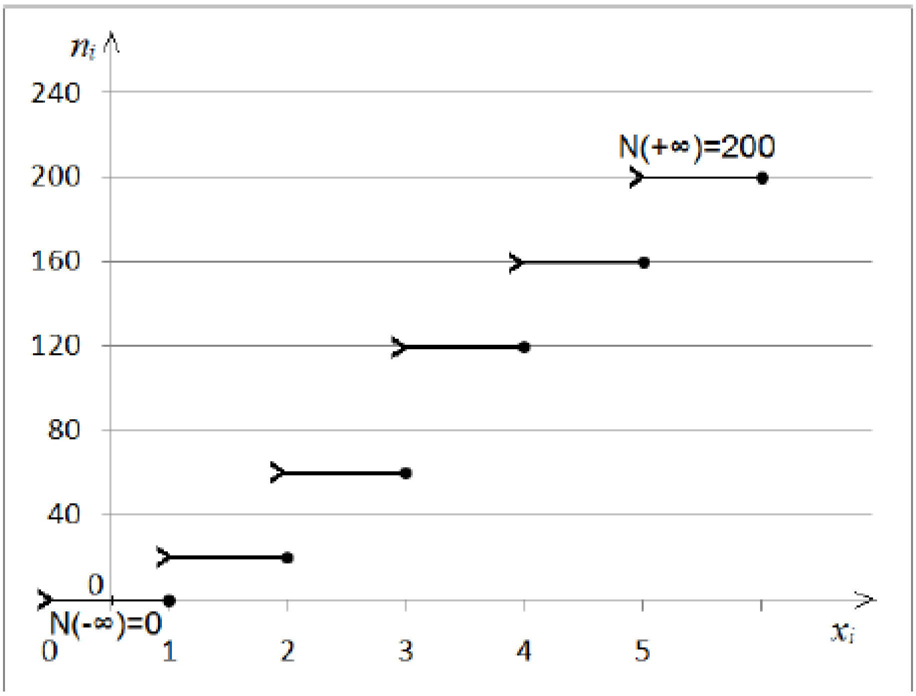
3) **Tracer le polygone des effectifs**



2)- Complété le tableau par effectifs cumulés et les fréquence cumulées croissantes et décroissante

x_i	n_i	N_i	f_i	$F_i\uparrow$	$F_i\downarrow$
1	20	0	0,1	0	1
2	40	20	0,2	0,1	0,9
3	40	60	0,2	0,3	0,7
4	60	120	0,3	0,5	0,5
5	40	160	0,2	0,8	0,2
Total	200	200	1,0	1,0	0,0

3)-Tracer la courbe cumulative croissante



1.7 Étude d'une variable statistique continue

Dans le cas continu, il est nécessaire de regrouper les résultats en classes à cause de leur grande masse. Un bon découpage correspond à des classes homogènes et séparées.

Dès qu'un caractère est identifié en tant que continu, ces modalités $x_i \in [L_i, L_{i+1}[$ sont des intervalles $[L_i, L_{i+1}[$ avec

– L_i : borne inférieure.

– L_{i+1} : borne supérieure.

– $a_i = L_{i+1} - L_i$: son amplitude, son pas ou sa longueur.

– $C_i = \frac{L_{i+1} + L_i}{2}$: son centre.

1 .7.1 Nombre de classes

Nous pouvons considérer dans ce cours trois formules .avec N désigne la taille de la population

1) $K \cong \sqrt{N}$,
2) **a formule de Sturge : $K = 1 + 3.3log_{10}(N)$,**
3) **la formule de Yule : $K = 2.5\sqrt[4]{N}$**

► **On va suivit les étapes suivantes pour déterminer ces classes :**

1 **Calculons l'étendu (E) tel que: $E = x_{max} - x_{min}$,**

2 **Calculons le nombre de classes K (K : un nombre entier positif)**

3 **D'où le pas ou l'amplitude de chaque classe est donné par : $a_i = \frac{E}{K}$**

REMARQUE.7

Nous mentionnons que les deux formules Yule et Sturge sont presque pareils si $N \ll 200$.par contre la formule $K \cong \sqrt{N}$

Exemple 10

On s'intéresse à la taille (cm) de 20 étudiants, les résultats sont obtenus comme suit :

140	144	150	156
142	146	152	157
143	147	153	158
143	148	154	159
144	150	155	163

Dans ce cas, on doit regrouper cette série en classes

- **Calculons l'étendu (E) tel que: $E = x_{max} - x_{min}$=163-140=23**
- **Le nombre de classes K est donné par $K \cong \sqrt{N} = \sqrt{20} \simeq 4.5$**
- **L'amplitude de chaque classe est donné par : $a_i = \frac{E}{K} = \frac{23}{\sqrt{20}} = 5.1 \simeq 5$**

Les classes sont : $[140.145[$., $[145.150[$, $[150.155[$, $[155.160[$, $[160.165[$ au

On obtient le tableau statistique suivant :

Les classes	$c_i = \frac{L_i + L_{i+1}}{2}$	n_i	f_i
$[140.145[$	142.5	6	0.30
$[145.150[$	147.5	3	0.15
$[150.155[$	152.5	5	0.25
$[155.160[$	157.5	5	0.25
$[160.165[$	162.5	1	0.05
Total		**20**	**1**

1.7.2 Représentation graphique d'un caractère continu

a) Histogramme des fréquences (ou effectifs)

Nous pouvons représenter le tableau statistique par un histogramme. Nous reportons les classes sur l'axe des abscisses et, au-dessus de chacune d'elles, nous traçons un rectangle dont l'aire est proportionnelle à la fréquence f_i (ou l'effectif

n_i) associée. Ce graphique est appelé l'histogramme des fréquences (ou l'histogramme des effectifs).

REMARQUE 8 :

► L'histogramme est constitué par l'ensemble des rectangles adjacents.

► Si en joignant les milieux des bases supérieures de chaque rectangle par des segments on obtient le polygone. Voir la figure (1.9).

► Pour tracer un histogramme, il est indispensable de distinguer deux cas :

1-Si les amplitudes de classes sont égales on trace directement l'histogramme

2- Si les amplitudes sont différentes, afin de constituer un histogramme, il est nécessaire de

► Calculer pour chaque classe l'amplitude a_i

► Calculer les densités qui sont définies par $d_i = \frac{n_i}{a_i}$ ou $d_i = \frac{f_i}{a_i}$

► Voir les deux figures suivantes 1.7 et 1.8

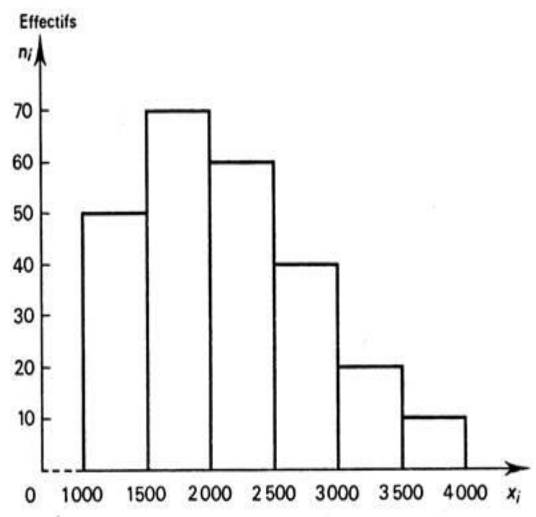


Figure 1.7 : Histogramme avec même amplitudes

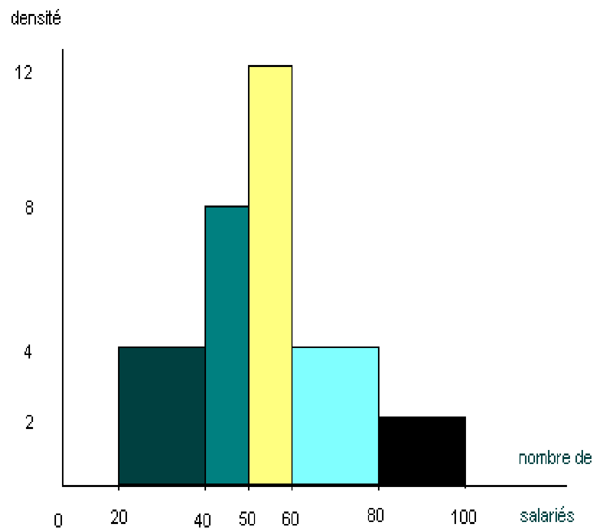


Figure 1.8: Histogramme avec des amplitudes différentes

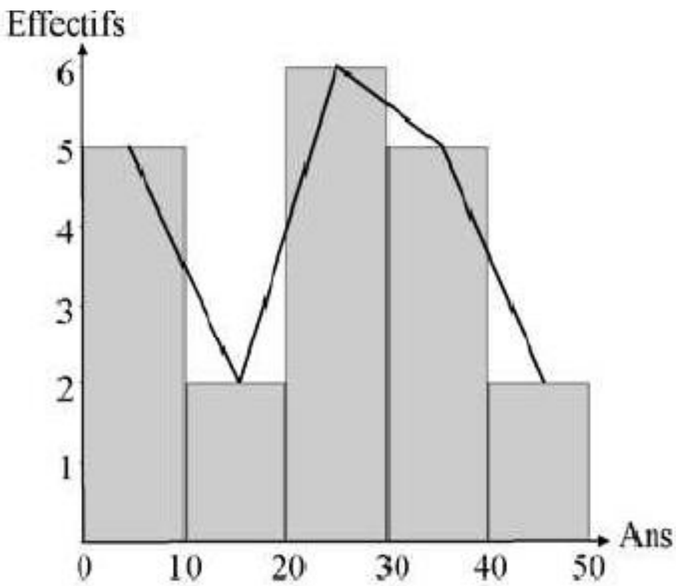


Figure 1.9: Histogramme et polygone des effectifs

C) La courbe cumulative croissante des effectifs ou fréquence

est présentée par les coordonnées : (L_1 ; 0) et (L_i ;$n_i^c \nearrow / f_i^c \nearrow$) et a_1 désigne la borne inférieure de la première classe.

b) La courbe cumulative décroissante des effectifs ou fréquence

est présentée par les coordonnées : (L_i ; $n_i^c \nearrow / f_i^c$) et (b_r ; 0) et b_r désigne la borne supérieure de la classe .(voir la figure 1.10)

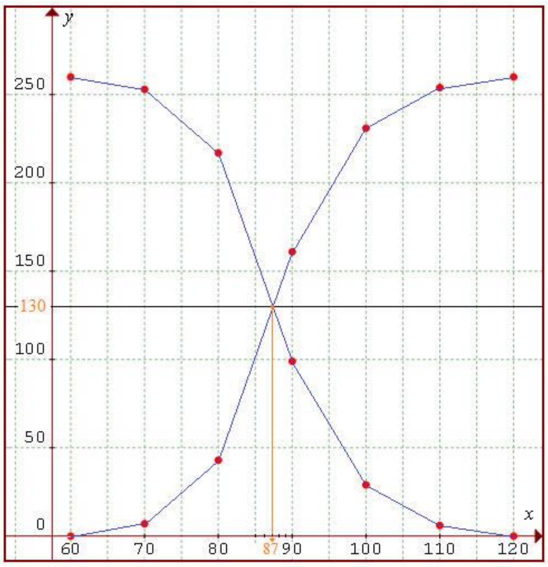


Figure 1.10: La courbe cumulative croissante et décroissante

b- La courbe de répartition

est présentée par les coordonnées : (L_i ; $f_i^c \nearrow$) et (b_r ; 0) et b_r désigne la borne supérieures de la dernière classe.

1.8. Paramètres caractéristiques des distributions statistiques

Trois aspects sont essentiels à l'interprétation d'une distribution :

- **Paramètre de position : le centre de la distribution et la répartition autour d'une valeur centrale (moyenne, mode, médiane, quantiles, ..)**

- **Paramètre de dispersion ou d'étendue : les valeurs sont-elles dispersées ou concentrées?**

- **Paramètre de forme : la forme de la distribution : la symétrie, l'aplatissement**

1.8.1. Paramètre de position et valeurs centrales

Le but des valeurs centrales est de résumer en une seule valeur l'ensemble des valeurs d'une distribution statistique. Il existe quatre valeurs de positions :

- Le mode (Mo),

- La moyenne ($\overline{X}$)

- La médiane ou le médian (Me)

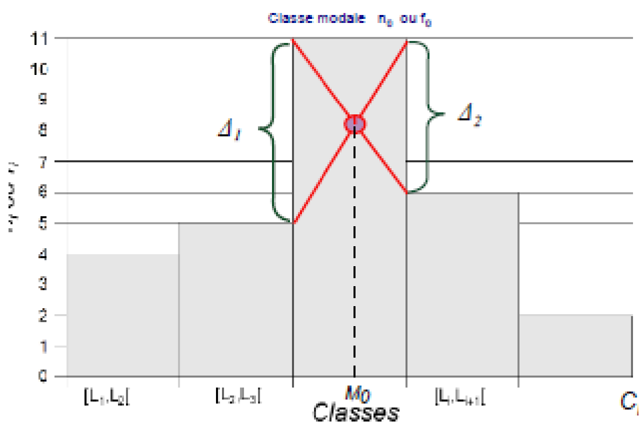
- Les fractiles (Quantiles) (Q_i)

Parmi ces valeurs les trois premières sont des valeurs de position centrales :

1. Le mode (valeur dominante)

* Caractère qualitatif et caractère discret : **le mode est la modalité ou la valeur qui a la fréquence simple la plus élevée (ou l'effectif le plus élevé, ce qui revient au même).**

* Caractère quantitatif continu : **Le mode est alors le centre de la classe modale, c'est à dire de la classe qui a la fréquence moyenne la plus élevée.**



2. La moyenne

Formalisation mathématique de la moyenne arithmétique

La moyenne arithmétique, noté $\bar{X}$ peut être résumée par la somme des observations divisée par l'effectif de l'échantillon étudié:

$$\bar{X} = \frac{\sum_{i=1}^{r} x_i n_i}{N} = \sum_{i=1}^{r} x_i f_i$$

REMRQUE 9

- **Elle est calculée pour les caractères quantitatifs.**
- **Si $n_i=1$ pour tout $i \in \{1, \dots, r\}$. D'où $\bar{X} = \frac{\sum_{i=1}^{r} x_i}{N}$**

Exemple. 11

Soit la série statistique suivante :

Valeurs	0	1	2	3	4
Effectifs	1	2	1	4	2

La moyenne est : $\bar{X} = \frac{\sum_{i=1}^{r} x_i n_i}{N} = \frac{0\times1+\times1\times2+2\times1+3\times4+4\times2}{10} = \frac{24}{10} = 2{,}4$

REMARQUE 10:

Si les données ont été regroupées en classes on ne peut calculer la valeur exacte de la moyenne. On peut toutefois en déterminer une bonne approximation en remplaçant chaque classe par son milieu. (centre de classe $c_i = \frac{x_i + x_{i+1}}{2}$) :

$$\bar{X} = \frac{\sum_{i=1}^{r} c_i n_i}{N}$$

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

4 . La médiane et la classe médiane

Définition générale :

On appelle médiane la valeur "du milieu". On dit qu'elle partage la série statistique en deux moitiés (C'est la donnée qui permet de diviser une série ordonnée d'une façon croissante en 2 parties égales (50%, 50%). La médiane ne peut être calculée que pour les caractères quantitatifs.

1) **Médiane, pour les données rangées : Les valeurs du caractère X étant classées par ordre croissant, la médiane est la valeur du caractère qui partage l'ensemble décrit par X en deux sous ensembles d'effectifs égaux : 50 % des éléments ont des valeurs de X supérieures à $X_{méd}$ méd et 50% prennent des valeurs inférieures.**

● Méthode de calcul

Soit une série statistique d'effectif total N, rangée par ordre croissant.

On a deux cas :

a) **N est impair**

La médiane (M_e) dans ce cas est la modalité qui a l'ordre $p+1$ c'est –à-dire : $M_e = x_{p+1}$

.

b) **N est pair**

La médiane (M_e) dans ce cas est le centre de classe $[x_p \ , x_{p+1}[$ c'est –à-dire

$$M_e = \frac{x_p \ + x_{p+1}}{2} .$$

Exemple 12:

Valeurs x_i	1	2	3	4	5	6
Effectif n_i	6	11	25	19	15	5
Effectif cumulé $n_i \nearrow$	6	17	42	61	76	81

► $N = 81$ est impair, $N = 81 = 2p + 1$ que ce implique $p = 81 - 1/2 = 40$

Donc $M_e = x_{p+1} = x_{40+1} = x_{41} = 3$

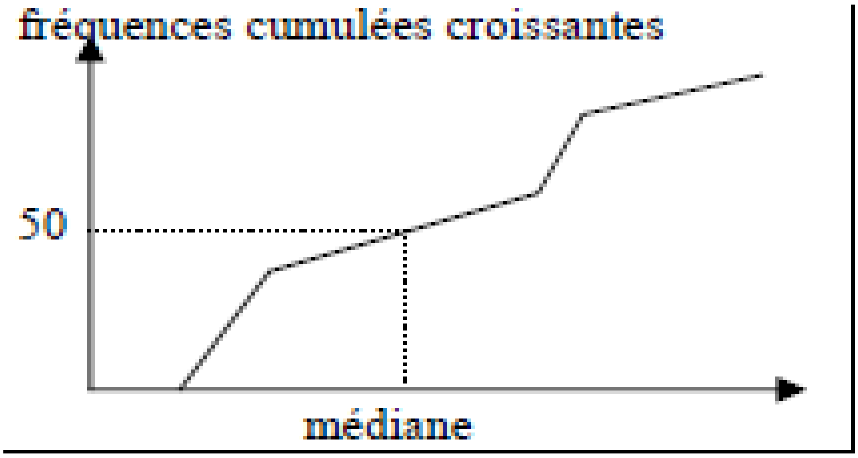
► Si l'effectif de la dernière modalité égale à 4, le N devient 80 dans ce cas est pair

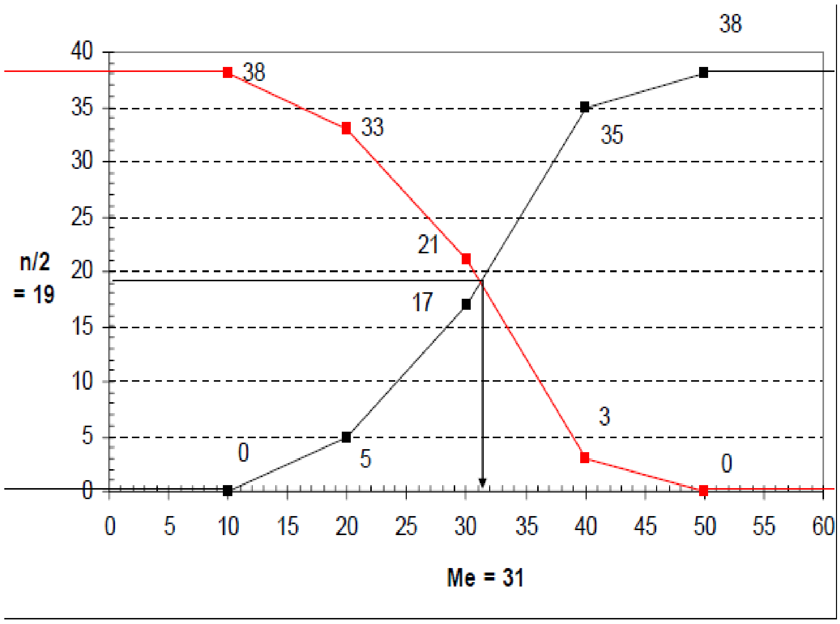
$N = 2p = 80$ que ce implique $\mathrm{p} = \frac{80}{2} = 40$

Donc $\mathrm{M_e} = (\mathrm{x}_{40} + \mathrm{x}_{41})/2 = (3 + 3)/2 = 3$

2) Médiane, pour les données regroupées

Si les données ont été regroupées en classes, on ne peut déterminer la valeur exacte de la médiane. On appellera classe médiane, La classe médiane est la première classe où la fréquence cumulée est supérieure ou égale à 0,5 ou l'effectif cumulé supérieur ou égale à $\frac{N}{2}$. On peut déterminer la valeur de la médiane grapgiquement ; voir les deux figures

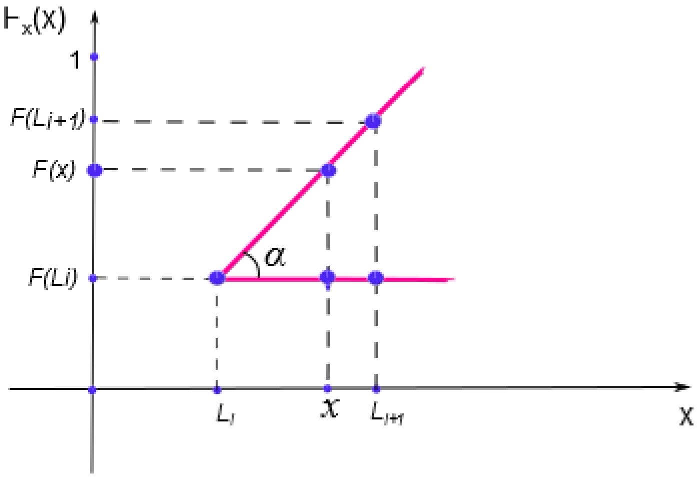

fréquences cumulées croissantes
50
médiane


40
35
30
25
20
15
10
5
0
n/2
= 19
38
33
21
17
3
0
0
5
35
38
0
5
10
15
20
25
30
35
40
45
50
55
60
Me = 31

● Méthode de calcul

Pour préciser la valeur de la médiane, il faut supposer que toutes les données sont réparties uniformément .

On repère la classe qui contient la médiane, puis on réalise une interpolation linéaire (voir figure)



Soit la classe médiane $[\ L_i,\ \ L_{i+1}[$ **c-à-d** $M_e \in [\ L_i,\ \ L_{i+1}[$ **, d'après la figure on a**

$$\text{tang}(\alpha) = \frac{F(M_e)-F(\ L_{i+1})}{F(L_{i+1})-F(\ L_i)} = \frac{M_e-\ L_i}{L_{i+1}-\ L_i},$$

Qui ce implique

$$M_e = L_i + \frac{0,5 - f^c \nearrow (L_i)}{f^c \nearrow (L_i) - f^c \nearrow (L_i)}(L_{i+1} - \ L_i).$$

Exemple 13

Classes	[0 ; 2[	[2 ; 4[	[4 ; 6[	[6 ; 8[
f_i	0,1	0,38	0,45	0,07
$f^c{}_i \nearrow$	0,1	0,48	0,93	1

$M_e \in [4\,;\ 6[$ car $f^c{}_i \nearrow 0{,}5$

$$M_e = 4 + \frac{0,5 - f\nearrow(4)}{f\nearrow(6) - f\nearrow(4)}(6-4)$$

$$= 4 + \frac{0{,}5 - 0{,}48}{0{,}93 - 0{,}48}(6-4) = \mathbf{4{,}088} \in [4\,;\ 6[$$

REMARQUE 11:

Quand on modifie les valeurs extrêmes d'une série, la moyenne change contrairement à la médiane qui ne change pas. On dit que la moyenne est "sensible aux valeurs extrêmes".

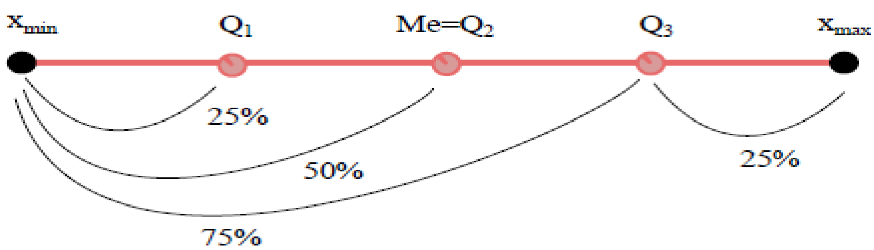
Nous généralisons la notion de la médiane dans la définition suivante.

Définition :

Pour $i \in \{1,2,3\}$, ***la quantité*** Q_i ***tel que*** $F(Q_i) = \frac{i}{4}$ **s'appelle le** $i^{èm}$ **quartile.**

REMARQUE 12

► La détermination ou le calcul de Q_i se fait exactement comme le calcul de la médiane (graphiquement ou analytiquement).



1.8.2 Paramètre de dispersion

Définition : On appelle dispersion statistique, la tendance qu'ont les valeurs de la distribution d'un caractère à s'étaler, à se disperser, de part et d'autre d'une valeur centrale.

a) Les paramètres de dispersion absolue

● Les paramètres de dispersion absolue indiquent de combien les valeurs d'une distribution s'écartent en général de la valeur centrale de référence.

● Un paramètre de dispersion absolue s'exprime toujours dans l'unité de mesure de la variable considérée.

● Les quatre paramètres de dispersion absolue les plus courants sont :

l'étendue,

► l'intervalle inter quantile (écarts inter quantiles),

► l'écart absolu moyen

► l'écart type.

b) L'étendue de la variation: l'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution , on la note *E*:

$$E = X_{max} - X_{min}$$

c) L'intervalle interquartile est l'étendue de la distribution sur laquelle se trouvent la moitié des éléments dont les valeurs de *X* sont les plus proches de la médiane.

On exclut alors de la distribution les 25% des valeurs les plus faibles et les 25 % des valeurs les plus fortes de *X*.

d) Variance et écart-type :

La variance et écart-type servent à évaluer la dispersion d'une distribution autour d'une valeur centrale : la moyenne

● **Variance** : La variance, notée S^2 est la moyenne du carré des écarts à la moyenne. La variance n'est pas un paramètre de dispersion absolue mais plutôt une mesure globale de la variation d'un caractère de part et d'autre de la moyenne arithmétique (quantité d'information). Pour obtenir un paramètre de dispersion absolue, on effectue la racine carrée de la variance, appelé écart-type et que l'on note S

$$S^2 = \frac{\sum_{i=1}^{r}(x_i - \overline{X})^2 n_i}{N} = \sum_{i=1}^{r}(x_i - \overline{X})^2 f_i$$

REMARQUE 13

La formule précédente est équivalente à la formule suivante :

$$S^2 = \frac{1}{N}\sum_{i=1}^{r} x_i^2\, n_i - \overline{X}^2 = \sum_{i=1}^{r} x_i^2\, f_i - \overline{X}^2$$

● **Ecart-type** : L'écart type, noté S est la racine carré de la moyenne du carré des écarts à la moyenne, c'est à dire la racine carrée de la variance.

$$S = \sqrt{S^2}$$

● **Le coefficient de variation**

Lorsqu'on compare des populations différentes, il est souvent plus commode de disposer d'indicateurs sans dimension. C'est pourquoi on complète l'étude de la dispersion d'une variable statistique par la donnée du coefficient de variation qui est égal au rapport de l'écart-type à la moyenne arithmétique :

$$CV = \frac{S}{\overline{X}}$$

Plus ce coefficient (qui ne peut être négatif) augmente, plus la dispersion de la série est importante autour de la valeur moyenne. On retiendra les ordres de grandeur suivants :

► $0 \leq CV \leq 0,2$ ou $0,3 \Rightarrow$ dispersion faible

► $0,3 \leq CV \leq 0,5$ ou $0,6 \Rightarrow$ dispersion moyenne

► $CV \geq 0,6 \Rightarrow$ dispersion forte

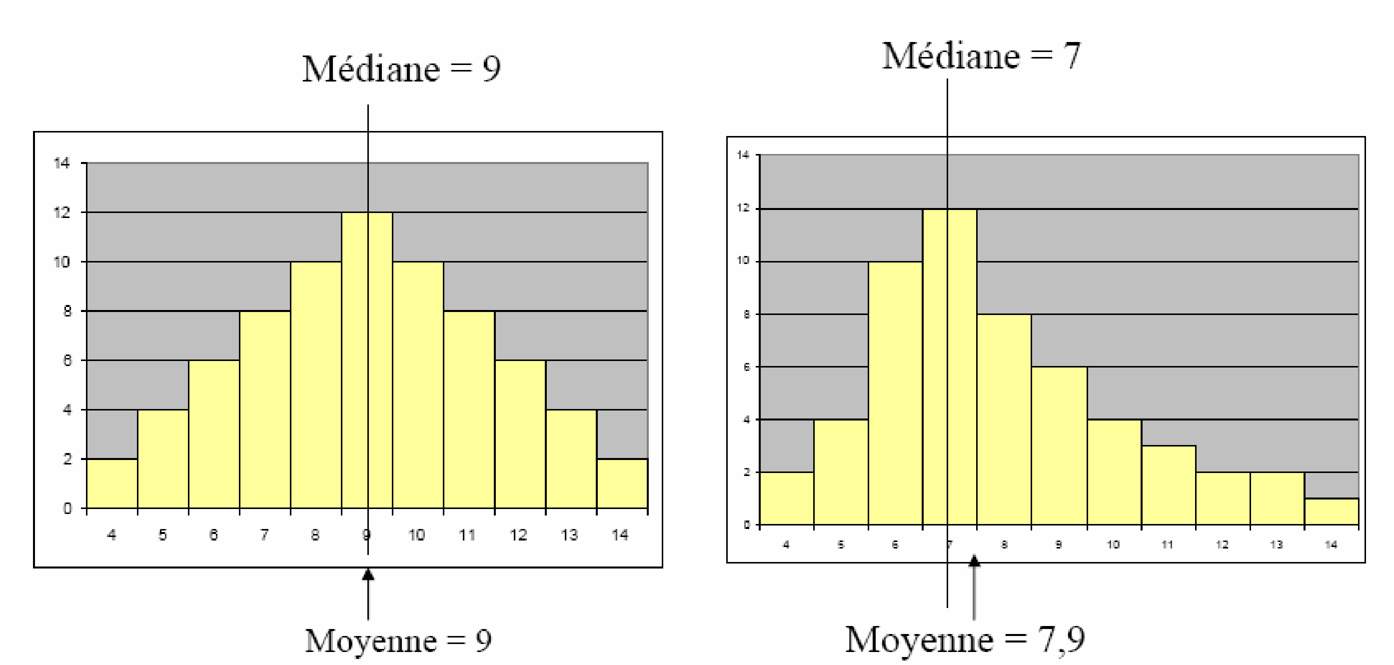
► $CV = 0$ signifie que toutes les valeurs de la série sont identiques.

► $CV > 1$ signifie que l'on a $> \overline{X}$, ce qui représente une dispersion des données extrêmement forte autour de la moyenne arithmétique.

1.8.3 . Paramètres de forme

Ces indicateurs viennent compléter la caractérisation d'une série statistique en donnant un ordre de grandeur de la physionomie de cette série, par le calcul d'un nombre unique, sans avoir besoin de tracer un graphique. Les paramètres de forme que nous aborderons sont :

1. **le coefficient d'asymétrie : . La distribution d'une variable est symétrique si les observations sont également dispersées de part et d'autre d'une valeur centrale. Ainsi, dans le cas de distributions symétriques, moyenne , médiane et le mode sont confondues, sinon elles sont distinctes**



Exemple de distributions symétrique et de dissymétrie

REMARQUE 14

- Dans le cas d'une dissymétrie positive on a généralement (partie droite plus longue que la partie gauche) : $M_o < M_e < \overline{X}$

- Dans le cas d'une dissymétrie négative on a généralement (partie gauche plus longue que la partie droite) : $M_o > M_e > \overline{X}$ **.**

a- Les coefficients d'asymétrie de Yule,

si la valeur centrale choisie est la médiane :Yule propose une mesure de l'asymétrie en comparant l'étalement vers la gauche et l'étalement vers la droite, tous deux repérés par la position des quartiles (Q_1, Médiane (Q_2))

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

Si : $\delta = 0 \Leftrightarrow$ symétrie parfaite

$\delta > 0 \Leftrightarrow$ oblique à gauche (ou étalement à droite) = dissymétrie à droite

$\delta < 0 \Leftrightarrow$ oblique à droite (ou étalement à gauche) = dissymétrie à gauche

b- coefficients d'asymétrie de Pearson

si les valeurs centrales choisies sont le mode et la moyenne. Pearson propose deux coefficients :

b.1. Le premier coefficient d'asymétrie de Pearson analyse la position de deux valeurs centrales (le mode et la moyenne arithmétique) relativisée par la dispersion de la série :

$$p = \frac{\bar{X} - M_O}{S}$$

Si : $p = 0 \Leftrightarrow$ symétrie parfaite

$p > 0 \Leftrightarrow$ oblique à gauche (ou étalement à droite) = dissymétrie à droite

$p < 0 \Leftrightarrow$ oblique à droite (ou étalement à gauche) = dissymétrie à gauche

b.2. Le second coefficient d'asymétrie de Pearson est plus élaboré : il s'appuie sur le calcul des moments centrés. Il s'écrit :

$$\beta_1 = \frac{u_2^3}{u_3^2}$$

De façon plus générale, on a :

► u_r moment centré d'ordre r est donné par $u_r = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{X})^2\, n_i$

► $m_r = \frac{1}{N}\sum_{i=1}^{k} x_i^r\, n_i$ moment d'ordre r.

► Pour $r = 2$, $u_2 = S^2$ et pour $r = 3$, u_3 moment centré d'ordre 3

Si : $\beta_1 = 0$ ⇔ symétrie

$\beta_1 > 0$ ⇔ oblique à gauche (ou étalement à droite) = dissymétrie à droite

$\beta_1 < 0$ ⇔ oblique à droite (ou étalement à gauche) = dissymétrie à gauche

c- **Les coefficients d'asymétrie de Fisher,**

si la valeur centrale choisie est la moyenne: Fisher propose

$$\gamma_1 = \frac{u_3}{S^3}$$

Si : $\gamma_1 = 0$ ⇔ symétrie

$\gamma_1 > 0$ ⇔ oblique à gauche (ou étalement à droite) = dissymétrie à droite

$\gamma < 0$ ⇔ oblique à droite (ou étalement à gauche) = dissymétrie à gauche

1. 9. Etude de deux variables statistiques- série statistique (bi variée)

1. 9.1 Présentation d'une série à deux variables

L'objectif de cette étude statistique est d'étudier sur une même population de N individus, deux caractères différents (ou modalités différentes) et de rechercher s'il existe un lien ou corrélation entre ces deux variables.

Exemple de relations possibles entre les variables suivantes : taille et âge ; diabète et poids ; taux de cholestérol et régime alimentaire ; niche écologique et population ; toxine et réaction métabolique ; survie et pollution ; effets et doses; organe 1 et 2 ; organe et fonction biologique ; …

Les caractères étudiés peuvent être aussi bien qualitatifs que quantitatifs.

Les résultats sont généralement représentés sous forme d'un tableau à double entrée, appelé tableau à deux dimensions, ou tableau croisé ou tableau de contingence, ou parfois tableau de corrélation.

Exemple de tableau de contingence

Sexe variable x	Effet 1	Effet 2	Effet 3	Total
H	43	36	3	Total des H : 82
F	49	12	12	Total des F : 73
Total	Total effet 1 92	Total effet 2 48	Total effet 3 15	Total des H et F :

1.9.2. Généralisation des représentation

Désignons par (X, Y) le couple de caractères étudiés.

A chaque observation conjointe (x_i, y_j) est associée le nombre d'individus ayant simultanément la valeur x_i pour le caractère X et la valeur y_j pour le caractère Y. Ce nombre est noté n_{ij} et appelé l'effectif associé à l'observation (x_i, y_j)

La ligne et la colonne total correspondent aux marges du tableau.

1.9 .3. Calcul des fréquences d'une statistique à deux

• Fréquences relatives partielles

La fréquence de l'observation (x_i, y_j) s'exprime par l'expression f_{ij}. Elle correspond à la proportion d'individus qui possèdent simultanément les valeurs x_i et y_j. Elle est obtenue par la formule suivante :

$$f_{ij} = \frac{n_{ij}}{N},$$

il est a remarquer que :

$$\sum_{i=1}^{k} \sum_{j=}^{z} f_{ij} = 1$$

• Fréquences relatives marginales $f_{i.}$ et $f_{.j}$

Il s'agit des fréquences relatives des distributions marginales.

$$f_{i.} = \frac{n_{i.}}{N} \quad \text{et} \quad f_{.j} = \frac{n_{.j}}{N},$$

Exemple 14

Effets de doses (variable Y : Effet 1, Effet 2 , Effet 3)
Sexe (variable X: H, F)

1) **Tableau de corrélation**

Sexe variable x	Effet 1	Effet 2	Effet 3	Total
H	43	36	3	Total des H : 82
F	49	12	12	Total des F : 73
Total	Total effet 1 92	Total effet 2 48	Total effet 3 15	Total des H et F :

2) Tableau des fréquences de l'effet des doses selon le sexe

Sexe	Effet1	Fffet2	Effet3	Total
H	0.28	0.23	0.02	0.53
F	0.32	0.08	0.08	0.47
Total	0.59	0.31	0.10	1

1.9.4. Calcul des moyennes marginales

Dans certaines distributions statistiques bidimensionnelles il est possible de calculer les moyennes, les variances et les écart-types marginaux. Nous expliciterons ces calculs à travers un exe

Soit la série statistique bidimensionnelle du couple (X, Y)

Calculer respectivement:

1- Les moyennes marginales de X puis de Y

2- Les variances et l'écart-type marginaux de X puis de

3- La moyenne conditionnelle de X quand $Y = 2$

4- La moyenne conditionnelle de Y quand $X = 3$

Réponce

1 et 2 - Les moyennes marginales, c'est le calcul des moyennes des effectifs marginaux.

- les variances et les écart-types marginaux se calculs aussi sur les effectifs marginaux.

Les formules respectives seront utilisées :

► **Pour les moyennes :**

a) $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{N}$ b) $\bar{Y} = \frac{\sum_{j=1}^{n} y_j}{N}$

► **Pour les variances :**

c) $S_X^2 = \frac{\sum_{j=1}^{n} x_i^2}{N} - \bar{X}^2$ d) $S_{Xj}^2 = \frac{\sum_{j=1}^{n} y_j^2}{N} - \bar{X}^2$

► **Dressons le tableau de distribution afin de faciliter les calculs**

X \ Y	-2	0	2	3	$n_{i.}$
2	3	4	0	6	13
3	4	3	3	2	12
4	2	3	3	2	10
$n_{.j}$	9	10	6	10	35

► **Distribution de la marginale X**

X	2	3	4	TOT
$n_{i.}$	13	12	10	35

► **Distribution de la marginale Y**

Y	-2	0	2	3	TOT
$n_{i.}$	9	10	6	10	35

a) $\overline{X} = \frac{\sum_{i=1}^{3} x_i}{N} = \frac{(2\times13)+(3\times12)+(4\times10)}{35} = \frac{102}{35} = 2.91$

b) $\overline{Y} = \frac{\sum_{j=1}^{4} y_j}{N} = \frac{(-2\times9)+(0\times10)+(5\times10)}{35} = \frac{24}{35} = 0.68$

c) $\delta_X^2 = \frac{\sum_{i=1}^{3} x_i^2 n_i}{N} - (\overline{X})^2 == 0.64$

d) $\delta_Y^2 = \frac{\sum_{j=1}^{4} x y_j^2 n_i}{N} - (\overline{Y})^2 = 3.81$

3- Pour déterminer la moyenne marginale X sachant que $Y = 2$, il sufit d'observer le comportement de X relatif à la colonne $Y = 2$

X	$y = 2$
2	0
3	3
4	3
TOT	6

► $\overline{X}_{y=2} = \frac{\sum_{i=1}^{3} x_i n_i}{N} = \frac{(2\times0)+(3\times3)+(4\times3)}{6} = \frac{21}{6} = 3.5$

- Pour déterminer la moyenne marginale de Y sachant que $X = 3$, il sufit d'observer le comportement de X relatif à la ligne $X = 3$

Y	-2	0	2	3	T0T
X=3	**4**	**3**	**3**	**2**	**12**

► $\overline{Y}_{X=2} = \frac{\sum_{j=1}^{3} Y_j \times n_j}{N} = \frac{(-2\times4)+(2\times3)+(3\times2)}{12} = \frac{4}{12} = 0.33$

1.9.5. Covariance

Une première approche entre de la relation éventuelle des valeurs d'une variable *X* avec des valeurs d'une variable *Y* est donnée par le calcul de la covariance.

● **Les formules pratiques de calculs de la covariance**

- Pour des données non groupées :

$$Cov(X,Y) = \overline{X \times Y} - \overline{X}\overline{Y}$$

$$= \sum_i^k \sum_j^l x_i y_j - \overline{X}\overline{Y}$$

- Pour des données groupées

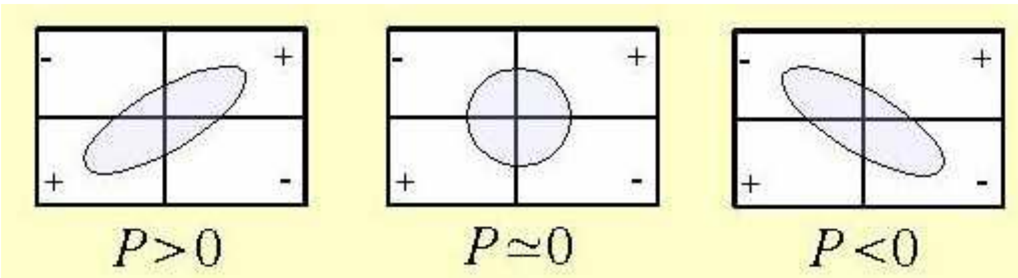
$$Cov(X,Y) = \sum_i^k \sum_j^l x_i y_j n_{ij} - \overline{X}\overline{Y}$$

● **Propriétés de la covariance**

● $Cov\,(X,X) = var\,(X)$

● $Cov(X\,,Y)\ \delta(X\,)\delta(Y)$

● Le signe de la **Cov** est un indicateur de la tendance de la relation sens positif ou Négatif (direction d'étirement du nuage de point)



● **Une covariance positive indique une tendance « croissante » des valeurs de Y en fonction de X, une covariance négative une tendance « décroissante »**

Exemple : **calcule de covariance**

Distribution bimodale dans un tableau contigence

X \ y	-2	6	2	n_i	$n_i x_i$
0	4	10	5	19	0
2	5	12	4	21	42
4	2	7	1	10	40
n_j	11	29	10	50	82
$n_j y_j$	-22				
Marg moy X	1.64				
Marg moy Y	-0.04				
Cov(X,Y)	-0.24				

1.9. 6. Coefficient de correlation

La covariance n'est pas un indicateur indépendant de l'ordre de grandeur des variables impliquées (de l'unité employée, par exemple).

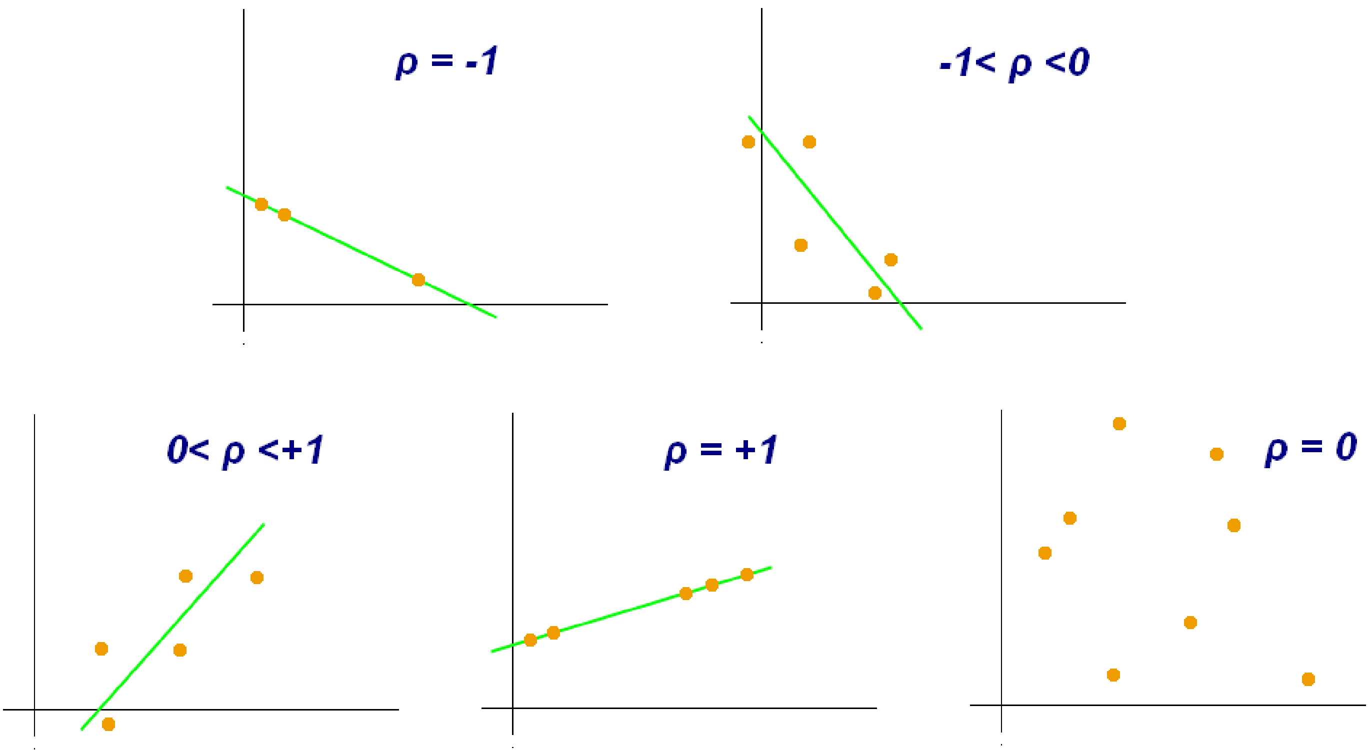
Le coefficient de corrélation, noté r, permet de résoudre cette difficulté. Ce coefficient pour le couple (X, Y) s'écrit selon la formule suivante :

$$r = \frac{Cov(X, Y)}{S_X S_Y}$$

Où S_X et S_Y désignent respectivement l'écart-type de la série statistique X et celui de la séries statistique Y

► **Propriété de r :**

- **r est toujours compris entre -1 et 1, c'est une covariance « réduite »**
- **quand ($|r| = 1$), les points représentatifs des couples (x_i, y_i), sont parfaitement alignés sur le graphique :**
- **quand ($|r|$ est voisin de 1), il existe une forte corrélation entre X et Y. Néanmoins (attention), ceci ne veut pas dire qu'il existe une relation de cause à effet entre elles.**
- **pour r = 1, la droite de la pente est croissante**
- **Si 0 < r < 1, la corrélation est positive, X et Y varient dans le même sens.**
- **Si -1 < r < 0, la corrélation est négative, X et Y varient dans le sens contraire.**
- **pour r = -1, la droite de la pente est décroissante**
- **quand (r = 0), aucune tendance** ne peut être déterminée



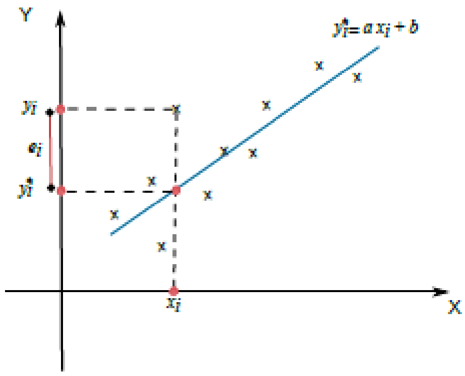
1.9.6. Doite de regression ou d'ajustement

► Importance de l'étude de corrélation entre 2 variables statistiques

L'une des méthodes simple d'étude de corrélation entre 2 variables consiste à rechercher une courbe d'équation $y = f(x)$ qui passe au plus proche de tous les points expérime expérimentaux. Une telle courbe permet d'avoir une idée sur la tendance de la relation entre les variables étudiées et de formuler d'éventuelles prévisions.

► Droite de régression linéaire

Une droite de régression linéaire s'écrit selon l'équation : « y = a x +b ». Cette approche de corrélation repose sur l'hypothèse que la relation entre deux variables est de nature linéaire



La droite la plus proche possible de chacun des points

- le coefficient a (pente) se détermine comme suit :

$$\hat{a} = \frac{\mathrm{Cov}(X, Y)}{S_X^2}$$

- le coefficient b (ordonnée à l'origine) se détermine comme suit :

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

Ainsi la droite de régression de y en fonction de x a pour équation :

$$\mathbf{y = a\,x + b = \frac{Cov(X,Y)}{S_X^2}(x - \bar{X}) + \bar{Y}}$$

Pour exprimer *X* en fonction de *Y*, il suffit d'inverser les rôles de *x* et *y*

$$\mathbf{x = \acute{a}\,x + \acute{b} = \frac{Cov(X,Y)}{S_X^2}(y - \bar{Y}) + \bar{X}}$$

Ces équations permettent de définir deux droites différentes de régression à l'intérieur dunuage de point. Néanmoins cette inversion, qui permet d'obtenir l'équation $x = \acute{a}y + \acute{b}$ (régression de *x* en *y*) n'est pas souvent intéressante, car en général, *Y* est une variable à exprimer et *X* est une variable potentiellement explicative.

► **Propriété de ces deux droites de régression :**

1) les deux droites de régression se coupent en un point qui a pour coordonnées les moyennes de *X* et de *Y*, point (x, y) , (en remplaçant dans l'équation *x* par sa moyenne, il est ainsi possible de retrouver *y* (qui correspond à la moyenne de *Y*)).

2) les coefficients $\acute{a}$ et (qui sont les pentes) sont toujours de même signe (soit – (corrélation négative) soit + (corrélation positive)), ainsi les deux droites sont orientées dans

$$\mathbf{Cov = \frac{\sum_{i=1}^{n} x_i y_i}{N}}$$

3) l'angle maximum des deux droites de régression est de 90° (droites perpendiculaires). Dans ce cas, les points sont dispersés dans tout le plan. La corrélation est nulle. Les droites sont respectivement parallèles à l'axe des *x* et à l'axe des *y*.

1.9.7. Corrélation

Définition : Corrélation C'est une relation positive ou négative entre deux phénomènes, mais elle n'est pas absolue. Ainsi, il y a une corrélation positive entre la taille et le poids des hommes : ceux qui mesurent un mètre quatre-vingt pèsent en général plus lourd que ceux dont la taille ne dépasse pas un mètre soixante. Mais il y a des petits gros et des grands maigres.

REMARQUE 15: Souvent, une corrélation est le signe d'une relation de cause à effet souvent comme la consommation de tabac qui provoque le cancer du poumon et non la prédisposition à ce cancer qui donne envie de fumer. Mais dans certains cas, les choses sont beaucoup moins évidentes. Et il peut arriver aussi que chacun des deux phénomènes soit à la fois cause et effet. En outre, il y a beaucoup de corrélations statistiques qui ne résultent aucunement d'une relation de cause à effet et qui sont de ce fait trompeuses. C'est notamment le cas pour les séries statistiques qui évoluent parallèlement dans le temps, avec le progrès économique et scientifique

●Application : AJUSTEMENT LINÉAIRE

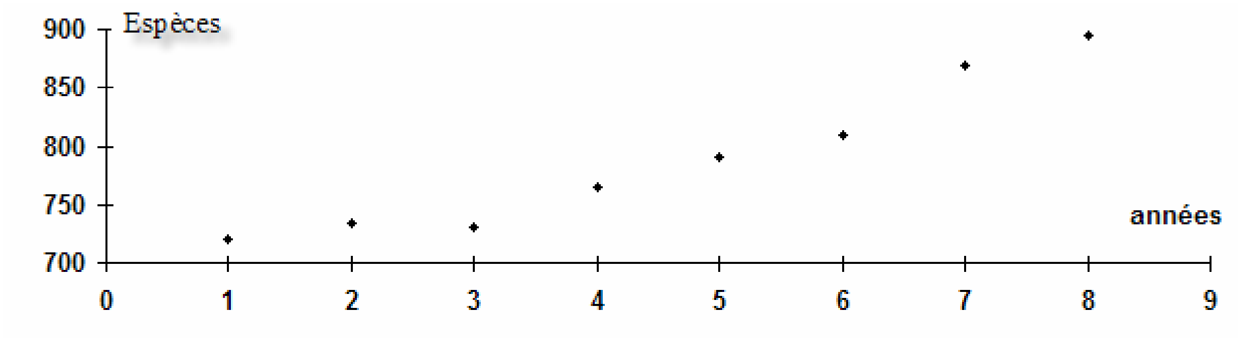
I. Savoir représenter graphiquement une série chronologique :

• Exemple 15 : On référence sur huit années, le nombre d'espèces affectées par une substance Toxique

Années	1	2	3	4	5	6	7	8
NB d'espèces	720	735	730	765	790	810	870	895

Représenter graphiquement cette série.

• Méthode : on porte en abscisse les numéros des années, $(1 \leq x_i \leq 8)$ et respectivement en ordonnées les effectifs des espèces affectées par cette toxine.



● **Ajuster la série par la méthode des moindres carrés :**

Exemple 16: ajuster la série précédente à l'aide d'une droite en utilisant la méthode moindres carrés.

● **Méthode : on commence par calculer la moyenne $\bar{x}$ des années (x_i) et la moyenne $\bar{y}$ des espèces (y_i).**

On obtient le point moyen *M* de la série, point par lequel passe la droite d'ajustement.

On calcule ensuite le coefficient directeur de la droite :

$$a = \frac{\sum X_i Y_i}{\sum X_i^2} \text{ ,}$$

Avec $X_i = x_i - \bar{X}$ et $Y_i = y_i - \bar{Y}$ (écarts par rapport aux moyennes respectives).

- le coefficient *b* (ordonnée à l'origine) se détermine comme suit :

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

Ainsi la droite de régression de *y* en *x* a pour équation :

$$y = a\,x + b = \frac{\mathrm{Cov}(X,Y)}{S_X^2}(x - \bar{X}) + \bar{Y}$$

Pour exprimer X en fonction de Y, il suffit d'inverser les rôles de x et y

$$x = \acute{a}\,x + \acute{b} = \frac{\mathrm{Cov}(X,Y)}{S_X^2}(y - \bar{Y}) + \bar{X}$$

Ces équations permettent de définir deux droites différentes de régression à l'intérieur dunuage de point. Néanmoins cette inversion, qui permet d'obtenir l'équation

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

(régression de x en y) n'est pas souvent intéressante, car en général, Y est une variable à exprimer et X est une variable potentiellement explicative.

Solution

Année x_i	Espèce y_i	X_i	Y_i	x_iy_i
1	720	-3.5	-69.38	242.81
2	735	-2.5	-54.38	135.94
3	730	-1.5	-59.38	89.06
4	765	-0.5	-24.38	12.19
5	790	0.5	0.63	0.31
6	810	1.5	20.63	30.94
7	870	2.5	80.63	201.56
8	895	3.5	105.63	396.69
36	6315			1082.50

On a $\overline{X} = 4.5$ et $\overline{Y}$ =789375 coordonnés du point moyen M

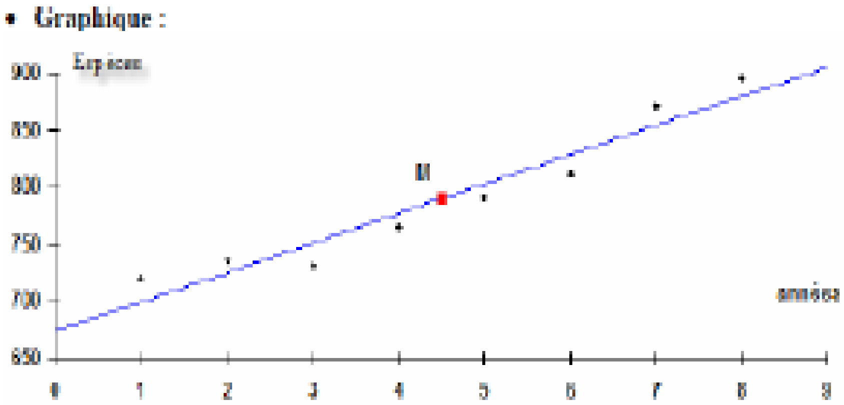
$$a = \frac{1082.50}{42} = 25.77$$

Coefficient dircteur de la droite sur M ; on a

$$\overline{Y} = a\,\overline{X} + b \Rightarrow b = 789375 - 25.77 \times 4.5 = 673.41$$

La droite a donc pour équation : $y = 25.77x + 673.41$.

Cela permet d'éffectué les prévisions. Par exemple pour l'année n° 9 le CA prévisionel sera $y_9 = 25.77 \times 9 + 673.41$=906.34 éspèces 907 (légèremeent différente de la méthode des points moyens



Partie 2
Analyse combinatoire

2. Analyse combinatoire

2.1 Introduction:

D'après François Le Lyonnais et son dictionnaire des mathématiques, la combinatoire est l'étude des configurations. Elle vise

1. **à mettre en lumière les propriétés intrinsèques de configurations données (nombres polygonaux, par exemple);**
2. **à démontrer l'existence ou la non-existence des configurations ayant des propriétés données (problème des sept ponts de Königsberg , construction des carrés magiques...);**
3. **à calculer le nombre des configurations, cet objectif définit l'analyse combinatoire proprement dite;**
4. **à énumérer ces configurations et à les classes au besoin;**
5. **éventuellement à les décomposer en sous-configurations.**

La combinatoire joue un rôle important en probabilités et en statistiques. Elle a fait progresser la théorie des algorithmes, la théorie des graphes, la théorie des jeux et la théorie de l'information. Ses domaines d'applications vont de la recherche opérationnelle à la chimie en passant par la physique et l'économie.

« L'analyse combinatoire s'emploie à étudier et à dénombrer divers groupements que l'on peut faire à partir d'ensembles finis. [...] L'analyse combinatoire est née de l'étude des jeux de hasard et s'est fortement développée sous l'influence du calcul des probabilités. » (*Atlas des mathématiques*, La Pochotèque, 1997)

2.2 Principe fondamental de dénombrement :

Soit une expérience aléatoire E composée de r expériences successives, la première pouvant produire un résultat quelconque parmi n_1 résultats possibles, la deuxième

produisant un résultat quelconque parmi n_2 résultats possibles,....., la r-ième pouvant produire un résultat quelconque parmi n_r résultats possibles. Le nombre total de résultats possibles pour l'expérience aléatoire E est le produit de

$$n_1 \times n_2 \times n_3 \times \ldots . \times n_r.$$

Exemple 1

Calculons le nombre de plaques minéralogiques distinctes disponibles par département quand la numérotation comprend 4 chiffres et 2 lettres.

Réponse :

En donnant une case à chaque chiffre ou lettre, donc on peut attribuer :

10 chiffres à la 1ère case

10 chiffres à la 2ème case

10 chiffres à la 3ème case

10 chiffres à la 4ème case

26 lettres à la 5ème case

26 lettres à la 6ème case

Le nombre de plaques différentes est donc :

$$N = 10.10.10.10.26.26 = 6750000 \, plaques.$$

2.2 Arrangement :

2.1.1. Arrangement sans répétition

Définition 1: On appelle arrangement sans répétition de p éléments pris parmi les n éléments de E, toute disposition ordonnée de p éléments de E.

Exemple 2 :

Les arrangements à 2 éléments de l'ensemble (1,2 ,3) sont (1,2), (1,3), (2,1), (2,3), (3,1), (3,2)

Proposition 1: le nombre d'arrangements sans répétition de p éléments pris dans un ensemble à n éléments est :

$$A_n^p = n\,(n-1)\,(n-2)\ldots\ldots\ldots\ldots\,(n-p+1)$$

Exemple 3 :

De combien de façon peut-on placer 3 dossiers différents dans 15 casiers vides à raison d'un dossier par casier.

Réponse :

On a : 15 façons différentes pour placer le 1er dossier

14 façons différentes pour placer le 2ème dossier

13 façons différentes pour placer le 3ème dossier

Au total on a : $N = 15.14.13 = 2730$ façons différentes.

► **Notion factorielle :**

$n! = n\,(n-1)\,(n-2)\,(n-3)\ldots\ldots\ldots\ldots\ldots 1$

$1! = 1,\ 2! = 2.1 = 2,\ 3! = 3.2.1 = 6$

Par convention : $0! = 1$

En appliquant cette notion factorielle à l'expression de A_n^p on trouve

$$A_n^p = n(n-1)(n-2)\ldots\ldots\ldots(n-p+1) = \frac{n!}{(n-p)!}$$

Exemple 4: l'ensemble (1,2 ,3)

On a: $n = 3$ et $p = 2$ donc

$$A_3^2 = \frac{3!}{(3-2)!}$$

Exemple 5: (3 dossiers)

On a: $n = 15$ et $p = 3$ donc

$$A_3^2 = \frac{3\,!}{(15-3)\,!} = 2730 \text{ façons différentes}$$

2.1.2. Arrangement avec répétition :

Lorsqu'un élément peut être choisi plusieurs fois dans un arrangement, le nombre d'arrangement avec répétition de p éléments pris parmi n, est alors :

$$A_n^p = n^p \quad avec \quad 1 \leq p \leq n$$

2.3 Permutations :

En itérant on vérifie qu'il y a n façons de choisir le pième élément de l'arrangement.de lettres identiques : L (3fois) et E (2fois)

2.3.1 Permutations sans répétition:

Définition 2 :

On appell e permutation des n éléments de l'ensemble E toute disposition ordonnées de ces n éléments. Le nombre de permutations de n éléments est noté P_n Les permutations de n éléments constituent un cas particulier des arrangements sans répétition : c'est le cas où $p = n$. Ainsi le nombre de permutation de n élément est :

$$P_n = A_n^n = n!$$

REMARQUE 15:

Deux permutations ne différent donc que par l'ordre des n éléments distincts qui la composent.

Exemple 6:

les permutations de l'ensemble (1,2,3) sont : (1,2,3) , (1,3,2), (2,1,3) , (2,3,1) , (3,1,2) , (3,2,1)

On a $p_3 = 3! = 6$

2. 3.2 Permutations avec répétition:

Dans le cas où il existerait plusieurs répétitions k d'un même élément parmi les n éléments, le nombre de permutations possibles des n éléments doit être rapporté aux nombres de permutations des k éléments identiques.

Le nombre de permutations de n éléments est alors :

$$p_n = \frac{n!}{k!}.$$

En effet, les permutations de k éléments identiques sont toutes identiques et ne comptent que pour une seule permutation.

Exemple 7 :

Considérons le mot « CELLULE ». Le nombre de mots possibles (avec ou sans signification) que l'on peut écrire en permutant ces 7 lettres est : $p_7 = \frac{7!}{2!\,3!} = 420$ **mots possibles. En considérant deux groupes de lettres identiques : L (3 fois) et E (2 fois)**

2.4 Combinaisons:

Définition 3:

On appelle combinaison de p éléments pris parmi les n éléments d'un ensemble E toute disposition non ordonnée de p éléments de E.

REMARQUE 16:

Deux combinaisons ne diffèrent que par la nature des éléments qui la composent, l'ordre de ces éléments est indifférent.

Exemple: 8

Les combinaisons à deux éléments de l'ensemble 1, 2, 3 sont : (1,2) ;(1,3) ;(2,3) .

Proposition 2:

Le nombre de combinaisons de p éléments pris dans un ensemble à n éléments est :

$$C_n^p = \frac{n!}{(n-p)!\,p!}$$

Exemple 9: (exemple précédent)

Les combinaisons à deux éléments de l'ensemble 1, 2,3

Réponse :

On a $n = 3$ et $p = 2$ donc $C_3^2 = \frac{3!}{(3-2)!2\,!} = 3$

Exemple 10 : **De combien de façons différentes peut-on choisir 2 délégués parmi 4 étudiants.**

Réponse :

On a $n = 4$ **et** $p = 2$ **donc** $C_4^2 = \frac{4!}{(4-2)!2\,!} = 6$

► **Propriétés :**

$C_n^n = 1\,; C_n^0 = 1$

$C_n^p = C_n^{n-p}; C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$

Formule du binôme de Newton :

$$(a+b)^n = C_n^0 a^n b^0 + C_n^1 a^{n-1} b^1 + C_n^p a^{n-2} b^2 + \cdots + C_n^n a^0 b^n$$

PARTIE 3
THÉORIE DE PROBABILITÉS

3. Théorie de probabilités

3.1. Introduction :

La théorie des probabilités est une branche bien établie des mathématiques qui trouve des applications dans tous les domaines de l'activité scientifique, de la musique à la physique, et dans l'expérience quotidienne, de la prévision météorologique à la prédiction des risques des nouveaux traitements médicaux.

Nous distinguons deux genres de phénomènes :

a) Ceux qui obéissent à des lois fixes. (loi de Newton de la pesanteur)

b) Ceux qu'on ne peut pas contrôler. Ils sont soumis au hasard.

Le résultat de l'expérience est variable, même si on répète l'expérience dans les mêmes conditions. Ce genre de phénomène est appelé «phénomène aléatoire».

Exemple 2:

- « Lancer un dé ».

- « Lancer une pièce de monnaie ».

La théorie des probabilités s'intéresse à l'étude de l'aspect aléatoire des phénomènes aléatoires.

3.2. Notions des probabilités

Définitions :

► Expérience aléatoire : c'est le mécanisme £permettant l'observation d'un phénomène aléatoire.

► Événement : on appelle événement tout ce qui peut se réaliser ou ne pas se réaliser, à la suite d'une expérience aléatoire.

►. Événement élémentaire : c'est un événement qui ne sera réalisé que par un seul résultat de l'expérience aléatoire.

Exemples :

- « obtenir un 6 » en jetant un dé est un événement élémentaire. - « Obtenir un nombre pair » en jetant un dé, n'est pas un événement élémentaire car il peut être réalisé par plusieurs résultats de l'expérience aléatoire qui sont : « obtenir 2 » ou « obtenir 4 » ou « obtenir 6 »

► **Ensemble fondamental** : l'ensemble Ω de tous les résultats possibles d'une expérience aléatoire donné s'appelle ensemble fondamental.

Exemples :

- « jeter une pièce de monnaie » $\Omega = \{P, F\}$ avec (P=pile ; F=face)
- « jeter un dé » $\Rightarrow$ $\Omega = \{1, 2, 3, 4, 5, 6\}$
- « jeter 2 pièces de monnaies » $\Rightarrow \Omega = \{PP, PF, FP, FF\}$

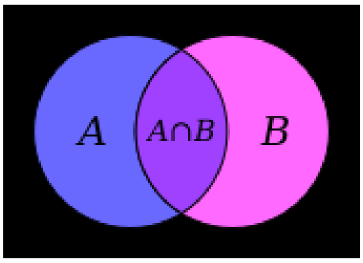
► Si A et B sont deux événements d'une expérience aléatoire alors $A \cap B$ et $A \cup B$ sont aussi des événements

► **L'événement « $A \cap B$ »** : est l'événement qui est réalisé si et seulement si les événements A et B sont tous les deux réalisés simultanément

Exemple 3:

Soit : $\Omega = \{1, 2, 3, 4, 5, 6\}$

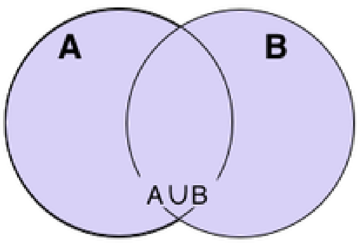
Et soient : $A = \{2, 4, 6\}$ et $B = \{1, 2, 4\} \Rightarrow A \cap B = \{2, 4\}$



► **L'événement « $A \cup B$ »** : est l'événement qui est réalisée si l'un au moins des événements A et B est réalisé.

Exemple 4:

Soit : $\Omega = \{1, 2, 3, 4, 5, 6\}$ et soient : $A = \{2, 4, 6\}$ et $B = \{1, 2, 4\} \Rightarrow AUB = \{1, 2, 4, 6\}$



3.3. Probabilité d'un événement:

Définition 1:

Si l'ensemble fondamental Ω contient N éléments équiprobables, et si l'événement A contient n éléments, alors la probabilité pour que A se réalise est:

$$P = P(A) = \frac{\mathrm{n}}{\mathrm{N}} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{\text{Cardinal}A}{\text{Cardinal}\Omega}$$

Exemple 5 :

- « lancer un dé » ⇨ $\Omega = \{1,2,3,4,5,6\}$, on a $N = 6$

Quelle est la probabilité d'obtenir un nombre de points pair.

« A » obtenir un nombre de points pair ⇨ $A = \{2,4,6\}$, on a $n = 3$

$$\Rightarrow P(A) = \frac{\mathrm{n}}{\mathrm{N}} = 36 = 0.5$$

Exemple 6 :

- « jeter une pièce de monnaie » $\Rightarrow$ $\Omega = \{P, F\}$ $\Rightarrow N = 2$

Quelle est la probabilité d'avoir face.

« A » avoir face $\Rightarrow$ $A = \{F\}$ $\Rightarrow$ $n = 1 \Rightarrow P(A) = \frac{\mathrm{n}}{\mathrm{N}} = \frac{1}{2} = 0.5$

Exercice 1:

On lance une pièce de monnaie deux fois consécutives, quelle est la probabilité d'avoir face au moins une seule fois ?

$\Omega = \{PP, PF, FP, FF\} \Rightarrow N = 4$,$A = \{PF, FP, FF\} \Rightarrow n = 3 \Rightarrow P(A) = \frac{\mathrm{n}}{\mathrm{N}} = \frac{3}{4}$

► **Événement contraire : L'événement contraire à A est $\bar{A}$, si parmi les cas possibles N, il y a n cas favorables à A, donc il reste $(N - n)$ cas favorables à$\bar{A}$.**

$$\Rightarrow P(\bar{A}) = N - \frac{\mathrm{n}}{\mathrm{N}} = 1 - \frac{\mathrm{n}}{\mathrm{N}}$$

$\Rightarrow \boldsymbol{P(\bar{A}) = 1 - P(A)}$; **Donc on aura :** $\boldsymbol{P(A) + P(\bar{A}) = 1}$

► **Impossibilité et certitude :**

➢ **Si le nombre de cas favorables à *A* est nul, l'événement *A* est impossible.**

$$P(A) = \frac{\mathbf{n}}{\mathbf{N}} = \frac{\mathbf{0}}{\mathbf{N}} = 0$$

Exemple 7: $\emptyset$ **est un événement impossible.**

➢ **Si le nombre de cas favorables à *A* est égal au nombre de cas possibles,**

L'événement *A* est certain. $\boldsymbol{P(A) = \frac{n}{N} = \frac{N}{N} = 1}$

Exemple 8: $\boldsymbol{\Omega}$ **est un événement certain.**

Entre ces deux extrêmes se situe toute une série d'événements probables.

REMARQUE 17:

La probabilité d'un événement est donc toujours comprise entre 0 et 1. :.

$\boldsymbol{0 \leq P(E) \leq 1}$

►. **Evénements incompatibles** **Les événements *A* et *B* sont dits incompatibles s'ils ne peuvent être réalisé en même temps, c'est à dire, si « A et B » est impossible donc A∩B=∅**

Exemple 9:

- « jet d'un dé » $\Rightarrow \boldsymbol{\Omega = \{1, 2, 3, 4, 5, 6\}}$

$\boldsymbol{A = \{1, 2\}}$ **et** $\boldsymbol{B = \{3, 4, 6\}}$

$\boldsymbol{A \cap B = \emptyset \Rightarrow A}$ **et** $\boldsymbol{B}$ **sont incompatibles.**

► **Evénements complémentaire :**

Deux événements *A* et *B* sont dits complémentaires si et seulement si :

➢ $\boldsymbol{A \cap B = \emptyset}$

➢ $\boldsymbol{A \cup B = \Omega}$

C'est-à-dire : $\boldsymbol{B = \Omega - A = \bar{A}}$

Donc $\boldsymbol{A}$ **et** $\boldsymbol{\bar{A}}$ **sont complémentaires.**

Exemple 10:

$\boldsymbol{\Omega = \{1,2,3,4,5,6\}}$

$\boldsymbol{A = \{1,3,5\}}$ **et** $\boldsymbol{B = \{2,4,6\}}$

$\boldsymbol{A \cap B = \emptyset}$

$\boldsymbol{AUB = \{1,2,3,4,5,6\} = \Omega}$ **Donc** $\boldsymbol{A}$ **et** $\boldsymbol{B}$ **sont complémentaires**

3.4. Opérations sur les événements :

1) $\boldsymbol{AUA = A}$

2) $\boldsymbol{AU\emptyset = A}$

3) $\boldsymbol{AU\,\Omega = \Omega}$

4) $\boldsymbol{AU\bar{A} = \Omega}$

5) $\boldsymbol{(\bar{\bar{A}}) = A}$

6) $\boldsymbol{AUB = BUA}$

7) $\boldsymbol{AU(BUC) = (AUB)UC}$

8) $\boldsymbol{AU\,(B \cap C) = (AUB) \cap (AUC)}$

9) $\boldsymbol{\overline{A \cup B} = \bar{A} \cap \bar{B}}$

10) $\boldsymbol{A \cap A = A}$

11) $\boldsymbol{A \cap \emptyset = \emptyset}$

12) $\boldsymbol{A \cap \Omega = A}$

13) $\boldsymbol{A \cap \bar{A} = \emptyset}$

14) $\boldsymbol{\overline{\Omega} = \emptyset, \overline{\emptyset} = \Omega}$

15) $\boldsymbol{A \cap B = B \cap A}$

16) $\mathbf{A \cap (B \cap C) = (A \cap B) \cap C}$

17) $\boldsymbol{A \cap (B \cup C) = (A \cap B) \cup (A \cap C)}$

18) $\boldsymbol{\overline{A \cap B} = \bar{A} \cup \bar{B}}$.

Exercice 2 :

Soient les deux événements $\boldsymbol{A}$ **et** $\boldsymbol{B}$ **avec :** $\boldsymbol{P(A) = 38;\ P(B) = 12\ et\ P(A \cap B) = 14}$

-Calculer : $\boldsymbol{P(A \cup B)}$ **,** $\boldsymbol{(\bar{A})}$ **,** $\mathbf{P(\bar{B})}$ **,** $\boldsymbol{P(\bar{A} \cap \bar{B})\ et\ P(\bar{A} \cup \bar{B})}$.

Solution :

$\boldsymbol{P(A \cup B) = P(A) + P(B) - P(A \cap B) = 38 + 12 - 14 = 58}$

$\boldsymbol{P(\bar{A}) = 1 - P(A) = 1 - 38 = 58}$;

$\boldsymbol{P(\bar{B}) = 1 - P(B) = 1 - 12 = 12}$

$\boldsymbol{P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 38}$

$\boldsymbol{P(\bar{A} \cup \bar{B}) = P\overline{(A \cap B)} = 1 - P(A \cap B)} = 34$

3.5. Probabilité conditionnelle

Définition 2

Soient A et B deux événements telle que que $P(B) \neq 0$.

On appelle probabilité conditionnelle de A relativement à B ou de A sachant B, la probabilité que l'événement A se réalise sachant que B est réalisé. Cette probabilité vaut

$$P_B(A) = \frac{P(A \cap B)}{P(B)}$$

REMARQUE 18: $P(B) \neq 0$ (car l'événement B est réalisé).

Déductions:

1) Si A et B sont incompatibles, alors $P(A/B) = 0$

2) $P(B/A) = P(A \cap B)P(A)$

3) $P(A \cap B) = P(A/B)\,P(B) = P(B/A)\,P(A)$

Exemple 11

Dans une école, $25\ \%$ des élèves échouent en maths, $15\ \%$ en chimie et $10\ \%$ à la fois en maths et en chimie.

On choisit un élève au hasard.

1) Si l'élève a échoué en chimie, quelle est la probabilité pour qu'il ait échoué en maths?

2) Si l'élève a échoué en maths, quelle est la probabilité pour qu'il ait échoué en chimie?

3) Quelle est la probabilité pour qu'il ait échoué en maths ou en chimie?

Solution :

M= « L'élève échoue en maths »

C = « L'élève échoue en chimie »

Données :

- $25\% = \frac{25}{100} = 0.25 = P(M)$

- $15\% = \frac{15}{100} = 0.15 = P(C)$

- $10\% = \frac{10}{100} = 0.10 = P(M \cap C)$

1- $P(M/C) = P(M \cap C)P(C) = 0.10 \times 0.15 = \frac{2}{3}$

2- $P(C/M) = P(M \cap C)P(M) = 0.10 \times 0.25 = \frac{2}{5}$

3- $P(M \cup C) = P(M) + P(C) - P(M \cap C) = 0.3$

3.5.1. Formule des probabilités totales

Définition 3 : (Partition d'un ensemble)

On dit que la suite $A_1, A_2, A_3; \dots ; A_n$ d'événements constitue une partition de Ω si :

1) $A_1 \cup A_2 \cup A_3 ; \dots \cup A_n = \Omega$

2) $A_i \cap A_j = \emptyset \; \forall \; i \neq j$

Proposition 1 :

Soit $A_1, A_2, A_3, \dots , A_n$ une suite d'événements constituant une partition de Ω, alors

$$P(A_1) + PA_2) + PA_3) \dots + P(A_n) = \sum_{i=1}^{n} \Sigma P(A_i) = 1$$

Exemple 12 :

A et $\bar{A}$ constituent une partition de Ω car :

1) $A \cup \bar{A} = \Omega$

2) $A \cap \bar{A} = \emptyset$

3.5.2. Théorème de Bayes

Soit $A_1, A_2, A_3; \dots ; A_n$ une suite d'événements constituant une partition de Ω

de probabilité non nulle. Alors, $\forall B \in$, on a :

$$P(A_i/B) = \frac{P(B/A_i)PA_i)}{\sum_{i=1}^{n} P(B/A_i)PA_i)}, \forall j = 1, 2, 3 \dots n$$

Exemple 13 :

Dans un atelier, deux machines M_1 et M_2 découpent des pièces métalliques identiques. M_1 fournit 60% de la production (parmi les quelles 6,3% sont défectueuses), le reste étant fourni par M_2 (dont 4% de la production est défectueuse).

La production du jour est constituée des pièces produites par les deux machines, et on en tire en fin de soirée une pièce au hasard (tous les prélèvements sont supposés équiprobables

1. La probabilité de prélever une pièce défectueuse, sachant qu'elle est produite par M_1 est : $P(D/M_1) = 0,063$.

2. La probabilité de prélever une pièce défectueuse, sachant qu'elle est produite par M_2 est : $P(D/M_2)) = 0,04$.

3. La probabilité de prélever une pièce défectueuse :

En utilisant la formule des probabilités totales, on a

$$P(D) = P(M_1) \times P(D/M_1) + P(M_2) \times P(D/M_2)$$

$$= 0,6 \times 0,063 + 0,4 \times 0,04 = 0,0538.$$

4. Si on prélève une pièce défectueuse, calculons la probabilité qu'elle soit produite par la machine M_1 :

En utilisant le théorème de Bayes, on a

$$P(M_1 \ D)) = \frac{P(M_1) \times P(D/M_1)}{P(D)} = \frac{0,063 \times 0,6}{0,0538} = 0,703.$$

PARTI 4

APPLICATIONS

Université de 8 Mai Guelma 1945 — Module : Probabilités et statistique

Faculté de science technique — Niveau : 2ème A

Département : ST — *Année : 2021-2022

Série 1 (statistique)

Exercice 1 : Classer les variables ci-dessous selon leur type :

Langue maternelle, Taille, Pays d'origine, Profession, Sexe, Nationalité, Poids, Pointure,

Race, Couleur des yeux, Dextérité Nombre d'enfants, Revenu mensuel, Taux d'endettement.

Exercice 2: Proposer des exemples de variable quantitative transformée en variable qualitative. Préciser les modalités de cette dernière.

Exercice 3: Compléter le tableau suivant :

						Total
Fréquences (...)	0.08	0.21				
Fréquence cumulées croissantes (…)			0.55	0.86		
Fréquence cumulées décroissantes (…)						

Exercice 4: Soit la liste suivante des prénoms d'un groupe d'étudiants suivis entre parenthèses d'une indication du nombre de livres lus dans l'année :(A = peu, B =

moyen, C = beaucoup, D =exceptionnel) Ahmed (C), Oussama (C), Yosra (A), Rabah (B), Nabil (A), Aridj (B), Hani (C), Salima (B), Farida (B), Dalila (C), Mehdi (D)

1. Quelle est la nature de la variable appétit de lecture ?

2. Construire le tableau représentatif de cette distribution.

3. Représenter cette distribution .

Exercice 5:

On a demandé aux enfants d'une classe : Combien y a-t-il d'enfants dans votre famille ?

La collecte des données nous fournit les données brutes :

1, 2, 1, 3, 1, 4, 2, 1, 3, 1, 2, 5, 2, 1, 1, 3, 2, 1, 2, 3, 1, 1, 1, 2, 4, 2, 1, 3.

1. **Présenter le tableau d'effectifs associé à cette série.**
2. **Donner la représentation graphique de cette distribution .**
3. **Déterminer la population, le caractère étudié et sa nature.**
4. **Calculer les fréquences cumulées croissantes et décroissantes.**
5. **Tracer la courbe cumulative croissante des effectifs.**
6. **Déterminer la fonction de répartition et tracer sa courbe.**

Université de 8 Mai Guelma 1945 Module : Probabilités et statistique

Faculté de science technique Niveau : $2^{ème}$A

Département : ST Année : 2021-2022

Série 2 (statistique)

Exercice 1 : On mesure la taille en centimetres de 50 elèves d'une classe

152 152 152 153 153

154 154 154 155 155

156 156 156 156 156

157 157 157 158 158

159 159 160 160 160

161 160 160 161 162

162 162 163 164 164

164 164 165 166 167

168 168 168 169 169

170 171 171 171 171

1. Determiner la population , le caractère étudié et sa nature .
2. Regrouper les données en classes .
3. Construire le tableau statistique.

Exercice 2 : La répartition des 100 exploitations agricoles selon leurs superficies en hectare (ha) se présente comme suit

Superficies en hectare	$[0,5[$	$[5,10[$	$[10,20[$	$[20,50[$	$[50,100[$
Nombre d'exploitations	5	24	38	26	5

1) **Calculer les fréquences, les centres , les amplitudes des classes.**
2) **Représenter l'histogramme et polygône des fréquences.**
3) **Calculer les fréquences cumulées croissante et décroissantes , et représenter leurs courbes dans le même repère.**
4) **Combien y-t-il des exploitations qui ont une superficie suérieure à 10 ha ?**
5) **Combien y-t-il des exploitations qui ont une superficie inférieure à 20 ha ?**
6) **Quelle le pourcentage des exploitations qui ont superficie supérieure à 10 ha ?**

<u>Exercice 3 :</u>On observe 100 fois le nombre d'arrivées (variable X) de clients à un bureau de poste pendant un intervalle de temps (10 minutes) et on obtient le tableau statistique

x_i	1	2	3	4	5	6
n_i	15	25	26	20	7	7

a-. Calculer les valeurs de tendance centrale de la distribution.

b- Calculer les valeurs de la dispersion de la distribution : variance, l'écart type et l'intervalle interquartile.

<u>Exercice 4</u> : Une société immobilière dispose de 600 appartements dont les surfaces sont données par le tableau suivant :

Surface (en mm2)	$[25, 50[$	$[50, 60[$	$[60, 80[$	$[80, 100[$	$[100, 120[$	$[120, 145[$
fréquence	0,02	0,15	0,13	0,22	0,28	0,20

a. **Compléter le tableau statistique suivant**

Classes	Centres c_i	Effectif n_i	Densités d_i	$n_i^c \nearrow$	f_i	$f_i^c \nearrow$	$f_i \times c_i$	$f_i \times c_{i\times} c_i$
$[25;50[$								
$[50;60[$								
$[60;80[$								
$[\_80;100[$								
$[100;120[$								
$[120;145[$								
total								

b. Calculer les indicateurs de position et ceux de dispersion et compléter le tableau suivant :

Moyenne =	**Q1 =**
Classe modale =	**Q2 =**
Mode =	**Q3 =**
Variance =	S^2 **=**
Ecart – type =	**S =**
Coefficient de variation =	**CV=**

c . calculer le coefficient d'asymétrie de Pearson et interpréter le résultat.

Université de 8 Mai Guelma 1945 **Module : Probabilités et statistique**

Faculté de science technique **Niveau : 2ème A**

Département : ST **Année : 2019-2020**

Série 3 (statistique) 2020

Exercice 1 : **Nous considérons 10 joueurs et soient :**

– Y la variable qui représente le nombre de jeux auquel un joueur joue.

– X la variable qui représente le gain ou perte (+1 s'il gagne 10 Da et −1 s'il perd 10 Da et 0 sinon).

Nous avons le tableau de contingence suivant,

X/Y	1	2	3	4	$n_{i.}$
-1	0	1	2	2	
0	1	1	0	1	
+1	0	1	1	0	
$n_{.j}$					

1. **Compléter le tableau si dessues**
2. **Calculer la covariance**

Exercice 2 : **Dans un TP de Micro-biologie , on a les données suivantes :**

x_i	0	0.5	1.1	1.5	1.9
m_i	0	10	20	30	40

La variable m_i représente les différentes masses appliquées comme dans le schéma ci-dessous et la variable x_i les hauteurs induits depuis l'état initial.

1. Déterminer la droite de régression D(m/x) et D(x/m).

2. Tracer le nuage de point et les deux droites. Représenter le point de coordonnée (x, y).

3. Peut-on déterminer x si m = 51.75 Kg ?

Exercice 3 :

- Le tableau de contingence suivant est entre le salaire mensuel X et l'ancienneté Y des ouvriers d'une entreprise.

X(×1000) \ Y	[0, 8[	[8, 16[	[16, 24[	[24, 32[
[20, 30[	5	6	1	0
[30, 40[	2	4	3	3
[40, 50[	0	2	4	10

1. Etudier les séries marginales.

2. Déterminer si les variables X et Y sont indépendantes.

3. Etudier les séries conditionnelles X/y_3 et Y/x_2 et présenter les résultats pour chaque Groupe de séries conditionnelles

Université de 8 Mai Guelma 1945 Module : Probabilités et statistique

Faculté de science technique Niveau : $2^{ème}$A

Département : ST Année : 2021-2022

Série 3 (Probabilités)

Exercice 1 :.

Combien peut-on former de numéros de téléphone à 8 chiffres ?

Combien peut-on former de numéros de téléphone à 8 chiffres ne comportant pas le chiffre 0

Exercice 2. :

d'un nombre de 3 chiffres distincts ou non.

Un clavier de 9 touches permet de composer le code d'entrée d'un immeuble, à l'aide d'une lettre suivie d'un nombre de 3 chiffres distincts ou non.

1	2	3
4	5	6
A	*B*	*C*

1) Combien de codes différents peut-on former ?

2) Combien y a-t-il de codes sans le chiffre 1 ?

3) Combien y a-t-il de codes comportant au moins une fois le chiffre 1 ?

4) Combien y a-t-il de codes comportant des chiffres distincts ?

5) Combien y a-t-il de codes comportant au moins deux chiffres identiques ?

Exercice3 :

Un sac contient 5 jetons verts (numérotés de 1 à 5) et 4 jetons rouges (numérotés de 1 à 4).

1) On tire successivement et au hasard 3 jetons du sac, sans remettre le jeton tiré.

Calculer les probabilités :

a) De ne tirer que 3 jetons verts ;

b) De ne tirer aucun jeton vert

c) De tirer au plus 2 jetons verts ;

d) De tirer exactement 1 jeton vert.

2) On tire simultanément et au hasard 3 jetons du sac. Reprendre alors les questions a), b), c) et d).

<u>Exercice 4 :</u>

Soient A et B deux événements tels que :

$P(A) = \frac{1}{5}, P(B) = \frac{1}{6}$ *et* $P(A \cup B) = \frac{1}{2}$.

1. Supposons que A et B soient incompatibles. Calculer $P(B)$.

2. Supposons que A et B soient indépendants. Calculer $P(B)$.

<u>Exercice 5 :</u>

3 machines automatiques produisent des pièces de voiture. La machine M_1 produit 40 % du total des pièces, la machine M_2 25 % et la machine M_3 produit 35 %.

En moyenne, les pourcentages des pièces non conformes aux critères imposées sont de 10 % pour la machine M_1 , de 5 % pour la machine M_2 et de 1% pour la machine M_3.

Une pièce est choisie au hasard dans la production totale des trois machines.

On constate qu'elle n'est pas conforme aux critères imposés.

1- Quelle est la probabilité qu'elle n'est pas conforme aux critères imposés.

2- Quelle est la probabilité qu'elle ait été produite par la machine M_1

Université de 8 Mai Guelma 1945 Niveau : $2^{ème}$A

Département : ST Année : 2022-2023

Série 3 (Probabiliés)

Exercice :1 traduire à l'aide des opérations sur les ensembles les expressions suivantes pour les 3 événements A, B, C

a. A seul se réalise ;
b. A et C se réalisent mais pas B
c. au moins l'un des trois événements se réalise ;
d. les trois événements se réalisent ;
e. aucun ne réalise ;
f. au plus l'un des trois se réalise ;
g. au plus deux des trois se réalisent.

Exercice :2 Soient A, B et C des événements

On pose $E_1 = A \cap (\overline{B} \cap \overline{C})$ et E_2=A$\cap$ $(B \cup C)$

1. Montrer que E_1 et E_2 sont incompatibles.

2. Déterminer l'ensemble $E_1 \cup E_2$.

3. On sait que P (A) = 0,6 , P (B) = 0,4 , P (C) = 0,3 , $P\ (B \cap C) = 0,1$, $P\ (A \cap C)\ 0,1$, $P\ (A \cap B) = 0,2$ et P $(A \cap B \cap C) = 0,05$.

- Calculer P (E_1) et P (E_2).

Exercice :3

A- Une urne contient sept boules : quatre rouges numérotées 1, 2, 3, 4 et trois vertes numérotées 1, 2, 3.

On tire deux boules au hasard, successivement et sans remise.

1. Quelle est la probabilité que la deuxième boule tirée soit rouge sachant que la première boule tirée est rouge ?

2. Quelle est la probabilité que les deux boules tirées soient rouges

B- On jet une pièce de monnaie deux fois de suites et on considèreles évènements :

- A : "On obtient pile au premier jet"
- B : "On obtient le même résultat dans les deux jets
- C : "On obtient pile dans les deux jettes"

A et B sont-ils indépendants ainsi A et C

Exercice :4 Dans une usine, deux lignes de montage fabriquent des composants électroniques.

La ligne A fabrique 70 % des composants. Le reste est fabriqué par la ligne B. La ligne A a un taux de composants défectueux en sortie de 0,5 %. Pour la ligne B, ce taux est de 0,4 %. On choisit au hasard un composant à la sortie de l'usine.

1. Calculez la probabilité que ce composante présente un défaut.
2. Quelle est la probabilité que ce composant ait été fabriqué par la ligne A sachant qu'il pésente un défaut .

Statistique déscriptive & théorie de probabilités

Université de 8 Mai Guelma 1945 Faculté de science technique Département : ST	Module : Probabilités et statistiqes Niveau : 2ème A Année : 2021-2022
Durée : 1H et 30min Goupe : Code :	
La note :	Code :

Examen final

Exercice 1 :

A) Cocher la bonne réponce

a- Quand les amplitudes sont inégales, pour dessiner l'histogramme

1. On calcule l'étendu.

2. On corrige les effectifs.

3. On calcule les effectifs cumulés.

b- L'intervalle interquartile :

1. Contient 50% des observations.

2. Est égal à Q3-Q1..

3. Est égal à Q1-Q3

c- La représentation graphique correspondant au cas quantitatif continu est :

1. L'histogramme.

2. Le diagramme en bâtons.

3. La courbe en escalier.

B) Parmi ces assertions, préciser celles qui sont vraies, celles qui sont fausses.(met V ou F)

1. La moyenne d'une série de valeurs distinctes peut être supérieure à la valeur maximale.

2. La variance peut être strictement négative.

3. L'écart type n'est jamais strictement inférieur e à zéro.

4. Les tableaux et graphiques sont utilisés pour donner une meilleure vue d'ensemble des données.

C) Classer ces statistiques selon leurs natures (indicateur de position ou de dispersion) Minimum ; Moyenne ;Écart-type ; Mode ; Médiane ; Premier quartile ; Etendu ; Coefficient de variation ; Variance ; Ecart interquartile.

Indicateur de position :……………………………………………………………………………. ….

Indicateur de dispersion :……………………………………………………………………………

Exercice 2 :

I) Soit la répartition des salaires journaliers des 620 employés d'une usine « A » :

1. **Déterminer la population, le caractère étudié et sa nature**
2. **Compléter le tableau**
3. **Déterminer la valeur du mode (par le calcul).**
4. **Déterminer la valeur de la médiane graphiquement et par le calcul, interpréter**
5. **Quel est le nombre d'employés qui perçoivent un salaire compris entre 600 et 800 DA par jour.**
6. **Evaluer la dispersion des salaires pour les employés dans cette usine.**

Réponse :

1 .la population statistique : le caractère étudié :

sa nature :……………………….

Les classes	n_i	f_i	$f_i^c \nearrow$	a_i	d_i	c_i	$c_i \times f_i$	$c_i^2 \times f_i$
$[5, 6[$	**100**							
$[6, 7[$	**80**							
$[7, 7.5[$	**240**							
$[7.5, 9[$	**160**							
$[9, 10[$								
TOTALE								

Statistique déscriptive & théorie de probabilités

Université de 8 Mai Guelma 1945 — Durée : 1H

Faculté de science technique — Niveau : $2^{ème}$A

Département : ST — Année : 2021-2022

Rattrapage de module probabilités &statistiques

Exercice 1 : (14 pts)

1. Commenter le diagramme ci-dessous.

2. Quelle est la variable représentée et déterminer ses valeurs ? Quel est l'individu statistique ?

3. Retrouver le tableau d'effectifs associé à ce diagramme et le complété par les effectifs cumulés croissants et tracer la courbe croissante des fréquences

4. Calculer la moyenne de joueurs sélectionnés par centre de formation

5. Calculer l'écart type du nombre de joueurs sélectionnés.

6. Déterminer le mode et le $2^{ème}$ quartile

Exercice 2 : (6 pts) À quels types de variable correspondent ces propriétés ?

1. **Ses valeurs ne possèdent pas d'ordre. Elles sont uniquement définies par des noms.**
2. **Elle s'exprime toujours à l'aide d'une unité de mesure.**
3. **Ses valeurs sont des noms mais correspondent à une hiérarchisation (c'est-à-dire possèdent un certain ordre)**
4. **Ses valeurs peuvent être n'importe quel nombre sur un intervalle.**

Statistique déscriptive & théorie de probabilités

Université de 8 Mai Guelma 1945 **Durée : 2H**

Faculté de science technique **Niveau : 2ème A**

Département : ST **Année : 2022-2023**

Examen final de module probabilités &statistiques

Exercice 1 :(12 pts)

Le tableau ci-dessous donne la répartition des salaires mensuels, en euros, des employés d'une entreprise :

Salaire	$[800; 900[$	$[900; 1000[$	$[1000; 1050[$	$[1050; 1150[$	$[1150; 1300[$
Effectif	**42**	**49**	**74**	**19**	**16**

1) Déterminer la population, le caractère étudié et sa nature.

2) Dans cette entreprise, combien d'employés gagnent au plus 1050 euros ?

3) Déterminer les trois quartiles (Q_1 , Q_2 , Q_3) graphiquement

4) Calculer de manière précise la médiane et les quartiles Q_1 et Q_3.

5) Calculer le salaire moyen dans cette entreprise, la distribution est-elle symétrique ?

6) Calculer l'écart type de cette série statistique.

7) Dans cette série statistiques se rajoute une sixième catégorie d'employés dont les salaires appartiennent à la classe [1300 ; 1500[. Quel est l'effectif de cette classe sachant que le salaire moyen au sein de cette entreprise est alors de 1200 euros.

Exercice 2 : (3pts)

On considère deux événements A et B tels que $P(A) = 0,4$ et $P(B) = 0,3$

1. **Calculer les probabilités de $A \cap B$ et de $A \cup B$ si A et B sont incompatibles .**
2. **Calculer les probabilités de $A \cap B$ et de $A \cup B$ si A et B sont indépendants .**

Exercice 3 : (5pts) Dans un groupe de 36 étudiants , il y a 15 garçons. 25 % des étudiants sont des filles qui font le devoir. Parmi les garçons , 20 % qui font le devoir . On choisit un étudiant au hasatrd dans cet groupe

1. **Calculer la probabilité que l'étudiant choisi fait le devoir**
2. **Q uelle est la probabilités que l'étudiant choisi est une fille sachant qu'elle fasse le devoir**
3. **Q uelle est la probabilités que l'étudiant choisi est un garçon et fait le devoir**
4. **<u>Micro –interrogation(10pts)</u> :Une entroprise dispose 153 machines . Un mois durant le service entretien de l'entreprise note tous les jours et pour chaque machine le nombre de pannes dans le tableau suivant, on prend 3 chiffres après la virgule.**

x_i	0	1	2	3	4	5	
n_i	69	41	19	13	8	3	

1. **Déterminer la population , la variable étudiée , les modalités et le type de caractère**
2. **Déterminer la fonction de répartition et tracer sa courbe**
3. **Calculer les paramètres de tendance centrale**
4. **Donner le pourcentage des machines qui ont au moins 2 pannes**

Partie 5

Solutions

Solutions de la série 1

EX 1 : **Les types possibles sont : qualitatif nominal (QN), qualitatif ordinal (QO), quantitatif discret (QD) et quantitatif continu (QC).**

Langue maternelle (QN) /Taille (QC) / Pays d'origine (QN) / Profession (QN)/

Sexe (QN) /Nationalité (QN) /Poids (QC) /Pointure (QD) /Race (QN)/ Couleur des yeux (QN)/ Dextérité (QO) /Nombre d'enfants (QD) / Revenu mensuel (QC) /Taux d'endettement (QC).

EX 2 : **Les variables quantitatives dans le tableau ci-dessous peuvent être transformées en variable qualitative ordinale. Les modalités de cette dernière sont précisées dans la seconde colonne.**

Variable quantitative	Modalités (QO)
1- Hauteur	1- Petit, Moyen, Grand
2- Poids	2- Très léger, Léger, Moyen, Lourd, Très lourd
3- Rendement	3- Faible, Moyen, Elevé
4- Chiffre d'affaire	4- Modéré, Moyen, Important, Très important
5- Cylindrée	5- Petite, Moyenne, Grosse

EX 3 :

						TOTALE
Fréquences (f_i)	**0.08**	**0.21**	**$a_{1.3}=f_3=0.26$**	**$a_{1.4}=f_4=0.31$**	**$a_{1.5}=f_5=0.14$**	1
Fréquences cummulées croissantes ($f_i^c\nearrow$)	**$a_{3.1}=0.08$**	**$a_{3.2}=0.29$**	**0.55**	**0.86**	**$a_{3.5}=1$**	
Fréquences cummulées décroissantes ($f_i^c\nearrow$)	**$a_{4.1}=1$**	**0.92**	**0.71**	**0.45**	**0.14**	

$\sum_{i=1}^{6} f_i$ *=1(total)*

$a_{1\,3}=0.08$ car le 1er $f^c_1\nearrow= f_1$

$a_{1.6}=1$ (la valeur maximale) et $a_{4.1}=1$

$a_{1.3}= f^c_2\nearrow= f_1+ f_2=0.08+0.21=0.29$

$a_{1.3}=f_3=0.26= 0.55-0.26= f^c_3\nearrow- f^c_2\nearrow$

$f_4=0.86-0.55= =f^c_4\nearrow- (f_1+ f_2+ f_3)$

f_5=1-0.86=0.14

$f^c{}_2$↘=1-0.08=0.92/ $f^c{}_3$↘=0.92-0.21=0.71 ....ect

- les valeurs sont évidentes / - avec calculs

<u>EX 4</u> :

1. **_L'appétit de lecture est une variable qualitative ordinale**
2. **Le tableau statistique**

Modalités	Effectifs	Fréquences $f_i=\frac{n_i}{N}$
Peu	**2**	**0 .18**
Moyen	**4**	**0.36**
Beaucoup	**4**	**0.36**
Exceptionnel	**1**	**0.09**
TOTAL	**11**	**1**

3. **Il-ya 2 représentation graphiques :**

le Tuyaux d'orgues qui est le plus simple (pas de calculs)

le diagramme circulaire (calculs des angles)

donc faire le plus simple

EX 5 :

x_i	1	2	3	4	5	TOTAL
n_i	12	8	5	2	1	28
f_i	0.43	0.3	0.18	0.07	0.04	1
${f_i}^c \nearrow = F(x)$	0.43	0.43+0.3=0.73	0.73+0.18=0.91	0.91+0.07=0.98	0.98+0.04=1.	
${n_i}^c \searrow$	28	28-12=16	16-8=8	8-5=3	3-2=1	

- **La population : Les enfants**
- **Le caractère : nbr d'enfants**
- **La nature : quantitative discret**
- **La représentation est un diagramme en bâtons.**

Solution de la série 2

EX 1 :

1)

La population : 50 élèves

Le caractère étudié : la taille

La nature : quantitatif continu

2) Pour regrouper les données en classes on va suivit les étapes suivantes :

a – Calculons l'etendu E=x_{max}-x_{min}= 171-152=19

b- Calcuons le nombre de calsses K$\simeq\sqrt{N}=\sqrt{50}$=7.07≃7

c- L'amplitude $a=\frac{E}{K}$=$\frac{19}{7}$=2.71≃3

D'où les classes sont :

[152 ;155[;[155;158[:[158 ;161[:[161;164[:[164 ;167 :[,[167 ;170[:[170;173[

3) Le tbleau statistique

Les classes	[152 ;155[	[155;158[	[158 ;161[	[161;164[	[164 ;167 [	[167 ;170[	[170;173[	ToT
Effectif	8	10	10	5	6	6	5	50

EX0 3

a) Le tableau statistique

X	ni	fi	Fi	xi*fi	xi²*fi
1	15	0.15	0.15	0.15	0.15
2	25	0.25	0.4	0.5	1
3	26	0.26	0.66	0.78	2.34
4	20	0.2	0.86	0.8	3.2
5	7	0.07	0.93	0.35	1.75
6	7	0.07	1	0.42	2.52
$\sum$	100	1		3	10.96

. Les valeurs de tendance centrale ou bien les paramètres de position

(la moyenne :$\overline{X}$, le mode :M_o ; la médiane M_e)

► **La moyenne** $\overline{X} = \frac{1}{N}\sum_{i=1}^{6} x_i n_i = \sum_{i=1}^{6} x_i f_i = 3$

► **Le mode est la valeur de X la plus fréquente c-à-d la modalité qui associée à l'effectif ou fréquence plus grande ; dans ce cas M_o= 3 (n=0.26≫)**

► **la médiane : M_e**

$F(M_e)=f^{c\nearrow}(M_e)=0.5 \Rightarrow M_e= 3$ (on choisit le x qu'il a le $f^{c\nearrow} \geq 0.5^{\circ}$)

ou bien on utilise la formule suivante si on utilise les effectifs cumulés croissant

$N=100=2p \Rightarrow p=\frac{100}{2}=50$, dans ce cas

$$M_e= \frac{x_p+x_{p+1}}{2}=\frac{x_{50}+x_{51}}{2}=\frac{3+3}{2}=3$$

Indice de Q1 est n/4= 25 % ⇒ Q1=2

Indice de Q2 est n/2=50% ⇒ Q2=3

Indice de Q3 est 3n/4=75 % ⇒Q3=4

c. Les valeurs de la dispersion de la distribution : variance, l'écart type et l'intervalle interquartile

Var(X)= 10.96 – 32= 1.96 $\Rightarrow s = \sqrt{var\ (X)} = 1.4$

IQ = Q3-Q1=4 – 2 = 2 Q1-1.5.IQ=2 - 1.5 . 2= -1 Q3+1.5 .

IQ= 4+1.5 . 2=7

EX 4 :

Classes	Centres c_i	Effectif n_i	Densités $d_i=\frac{n_i}{a_i}$	$n_i^c \nearrow$	f_i	$f_i^c \nearrow$	$f_i \times c_i$	$f_i \times c_{i\times}c_i$
$[25;50[$	37.5	12	0.48	12	9.02	0.02	0.75	28.125
$[50;60[$	55	90	9	102	0.15	0.17	28.125	453.75
$[60;80[$	70	78	3.9	180	0.13	0.3	9.1	9.1
$[\_80;100[$	90	132	6.6	312	0.22	0.52	19.8	19.8
$[100;120[$	110	168	8.4	480	0.28	0.80	30.8	3388
$[120;145[$	135.25	120	4.8	600	0.2	1	16.5	1361.25
total		600						7650.125

Calculons les indicateurs de position et ceux de dispersion

Moyenne = 85.2

Classe modale = 50;60

Mode = 55

Variance = 391.085

Ecart – type = 19.77587

Q_1**=72.12598**

Q_2 **=98.18182**

Q_3 **=116.4286**

$Q_3 - Q_1$ **= 44.30262**

$Q_1 - 1,5\ (Q_3 - Q_1) =$ **5.67205**

► $F(Q_1)$=0 .25 ⇒ $Q_1 \in [60,\ 80[$

$Q_1 = a_i + (a_{i+1}\text{-}a_i)\times\frac{0.25-F(a_i)}{F(a_{i+1})-F(a_i)}$ $\quad (a_i = 60,\ a_{i+1} = 80)$

$Q_1 = 60 + (80\text{-}60)\times\frac{0.25-0.173}{0.3-0.173} \cong 72.13$

► $F(Q_2)$=0 .5 ⇒ $Q_2 \in [80,\ 100[$

$Q_2 = 80 + (100\text{-}80)\times\frac{0.5-0.3}{0.52-0.3} \cong 98.18$

► $F(Q_3)$=0 .75 ⇒ $Q_3 \in [100,\ 120[$

$Q_3 = 100 + (120\text{-}100)\times\frac{0.75-0.52}{0.80-0.52} \cong 116.43$

S^2=7650 .125-85.2^2=391.085

Solution de la série 3

<u>EX 1</u> : Les lois marginales sont données dans ce tableau

X \ Y	1	2	3	4	$n_{i\bullet}$
-1	0	1	2	2	5
0	1	1	0	1	3
1	0	1	1	0	2
$n_{\bullet j}$	1	3	3	3	N=10

La covariance est calculer à partir de

$$Cov(X,Y) = \overline{XY} - \overline{X}\overline{Y}$$

Nous avons

$\overline{X} = \frac{1}{N}\sum_{i=1}^{3} x_i\, n_{i\bullet} = -0.3$, **et** $\overline{Y} = \frac{1}{N}\sum_{i=1}^{4} y_i\, n_{\bullet j} = 2.8$

De plus , nous avons $\overline{XY} = \frac{1}{N}\sum_{i=1}^{3}\sum_{i=1}^{4} x_i y_i\, n_{ij} = -1$

Donc : $Cov(X,Y) = -0.16$

<u>EX 2</u> :

Nous déterminons facilement les moyennes $\overline{X} = 1$ **et** $\overline{m} = -1$ **. De plus nous avons**

$Cov(X,m) = \overline{Xm} - \overline{X}\overline{m}$ **= 29.6-20 = 0.6.**

Après calculs nous avons aussi

$S^2{}_X = \overline{X^2} - (\overline{X})^2$**=0.464 et** $S^2{}_m = \overline{m^2} - (\overline{m})^2$**=200** $\Rightarrow S_X$**= 0.681 et** S_m**=14.14.**

Dans ce cas les coéfficients de la droite sont donnés par

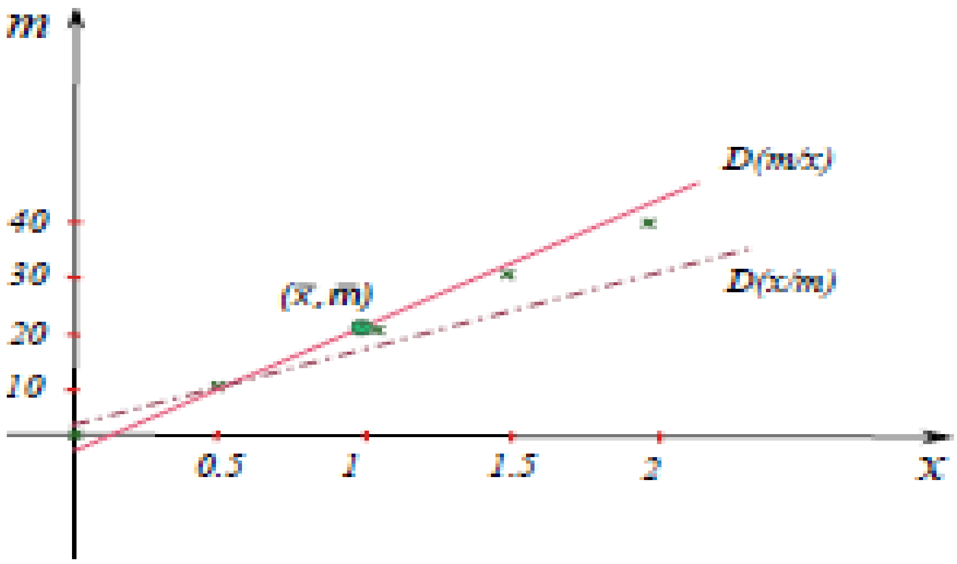
$a = \frac{Cov(X,Y)}{S^2{}_X}$=20.69 et $b = \bar{m} - a\bar{X}$=-0.69 .

Alors $D_{X/m}$: $\mathbf{m = 20.69x - 0.69.}$

En renversant les axes, nous obtenons

$D_{m/X}$: $\mathbf{x = 0.048m - 0.04.}$

Nous traçons les deux droites dans la figure ci-dessous (*D*(*x/m*) est la symétrie de *D*(*m/x*) par rapport à la première bissectrice).



EX 3: Nous complétent le tableau comme suit

Y / X($\times$ 100)	$[0,8[$	$[8,16[$	$[16.24[$	$[24,32[$	$n_{i\bullet}$	$f_{i\bullet}$
$[20,30[$	5	6	1	0	12	0.3
$[30,40[$	2	4	3	3	12	0.3
$[40,50[$	0	2	4	10	16	0.4
$n_{\bullet j}$	7	12	8	13	40	1
$f_{\bullet j}$	7\40	12\40	8\40	13\40	1	

Les moyennes après les calculs : $\overline{X} = 36(\times 1000)$ **et** $\overline{Y} = 17.4$.

La variance et l'écart-type de X : $S^2{}_X = 69$, $S_X = 8.310$

La variance et l'écart-type de Y : $S^2{}_Y = 78.04$ **et** $S_Y = 8.84$

2- Si on choisit i **=3 et** $j = 1$**, nous obtenons**

$$N \times n_{31} = 40 \times 0 = 0 ,$$

et

$$n_{3\bullet} \times n_{\bullet 1} = 16 \times 7 = 112,$$

qui sont bien évidement non égaux. Par conséquent, d'existe i et j tel que

$$N \times n_{ij} \neq n_{i\bullet} \times n_{\bullet j} .$$

Donc X et Y ne somt pas indépendants.

3- L série $X|Y_3$ est la série conditionnelle de X sachant $Y=y_3$ (j=3) sa moyenne est donnée par $\overline{x_3}$= 83.75

3- L série $Y|x_2$ est la série conditionnelle de Y sachant $X=x_2$ (i=2) sa moyenne est donnée par $\overline{y_2}$= 16.67.

Solution de la série 3 (2021-2022)

Exercice 1

Un numéro de téléphone à 8 chiffres est une 8-liste d'éléments choisis dans l'ensemble $\Omega = \{0;1;2;3;4;5;6;7;8;9\}$.

On applique le principe fondamental de dénombrement un arrangement avec répétition

On peut former 10^8 numéros de téléphone à 8 chiffres

Un numéro de téléphone à 8 chiffres ne comportant pas le chiffre 0 est une 8-liste d'éléments choisis dans l'ensemble $\Omega' = \{1;2;3;4;5;6;7;8;9\}$.

On peut ainsi former 9^8 =43046721 numéros de téléphone à 8 chiffres ne comportant pas le chiffre 0

Exercice n:2

1) Un code est un élément du produit cartésien entre un élément de l'ensemble {A, B ; C}, de cardinal 3, et de l'ensemble des 3-listes d'éléments de {1 ;2 ;3 ;4 ;5 ;6}, de cardinal $6^3 = 216$

Il y a donc $3\times6^3 = 3\times 216 = 648$ codes possibles.

2) Si le code ne doit pas contenir de chiffre 1, alors les 3-listes sont constituées d'éléments de {2 ;3 ;4 ;5 ;6}. Il y en a

donc 5^3 =125, et le nombre de codes vaut alors $3\times5^3 = 3\times125 = 375$

3) Le contraire de « le code contient au moins une fois le chiffre 1 » est « le code ne contient aucun chiffre 1 »

Le nombre de codes contenant au moins une fois le chiffre 1 est donc égal au nombre total de codes diminué du nombre de codes ne contenant pas le chiffre 1. Ces deux nombres ayant été calculés dans les deux questions précédentes, on conclut que le nombre de codes contenant au moins une fois le chiffre 1 est égal à 648-375=273

Exercice n :3

1) Tirages successifs sans remise de 3 jetons parmi 9. Il y a 3

A_9^3= 504 possibilités

a) Notons A l'événement « Tirer 3 jetons verts ». On a card(A)= A_5^3

P(A) = A_5^3 / A_9^3=5/42

b) Notons B l'événement « Ne tirer aucun jeton vert ». On a card(B)= A_4^3

P(A) = A_4^3 / A_9^3=1/21

c) Notons C l'événement « Tirer au plus 2 jetons verts »

1ère méthode

P(C)=P($\bar{A}$)=1-- P(A)=1-(5/42)=37/42

2ème méthode

P(C)= (A_4^3+3 A_5^1 A_4^2+ 3A_5^2 A_4^1)/ A_9^3)=37/42

c) Soit D l'événement « Tirer exactement 1 jeton vert ».:

P(D)= 3A_5^2 A_4^1/ A_9^3=5/14

2) Tirages simultanés de 3 jetons parmi 9. Il y a C_5^3=84

P(A) = C_5^3 / C_9^3=5/42

Le même principe pour les autres

Exercice 4 :

1) A et B incompatible donc $A \cap B = \emptyset$ d'où $P(A \cup B) = P(A) + P(B) \Rightarrow P(B) = \frac{1}{2}-\frac{1}{5} = \frac{3}{10}$.

2) A et B indépendants : $P(A \cap B) = P(A) \times P(B) \Rightarrow \frac{1}{2}=\frac{1}{5} + P(B)-\frac{1}{5}P(B) \Rightarrow$

$\frac{4}{5}P(B) = \frac{3}{10} \Rightarrow P(B)= \frac{3}{8}$

Exercice 5 :

Les événements

$M1$**: « Pièces produites par la machine** $M1$ **»**

$M2$**:« Pièces produites par la machine** $M2$ **»**

$M3$**:« Pièces produites par la machine** $M3$ **»**

C**:« Pièces conformes aux critères »**

$\overline{C}$ **« Pièces non conformes aux critères »**

Les données :

$P\ (M1) = 40\% = 0.4$ $\qquad$ $P\overline{C}/M1) = 0.10$

$P\ (M2) = 25\% = \ 0.25$ $\qquad$ $P(\overline{C}/M2) = 0.05$

$P\ (M3) = 35\% = 0.35$ $\qquad$ $P(\overline{C}/M3) = 0.01$

Conditions de Bayes

1°) $M1 \cup M2 \cup M3 = 100\% = \Omega$

2°) $M1 \cap M2 = M1 \cap M3 = M2 \cap M3 = \emptyset$

1) **Calculons la probabilité qu'elle n'est pas conforme aux critères imposés.**

c-â-d $P(\overline{C}) = ??$

On applique la formule des la probabilités totales

$$P\ (\overline{C}\ ) = P(\overline{C}/\ M_1)P(\ M_1)\ +\ P(\overline{C}/\ M_2)P(\ M_2)\ +\ P(\overline{C}/\ M_3)P(\ M_3)$$

$$= 0.10 \times 0.4 + 0.05 \times 0.25 + 0.01 \times 0.35$$

$$= 0.0560$$

2) **Calculons la probabilité qu'elle ait été produite par la machine** M_3

On applique la formule des la probabilités totales la formule de Bayes car l'inconnu Est un élément de la partition $\{A_1, A_2, A_3\}$

D'ou

$$\mathrm{P}(\ \mathrm{M_1}/\overline{\mathrm{C}}) = \frac{\mathrm{P}(\overline{\mathrm{C}}/\ \mathrm{M_1})\mathrm{P}(\ \mathrm{M_1})}{\mathrm{P}(\overline{\mathrm{C}}/\ \mathrm{M_1})\mathrm{P}(\ \mathrm{M_1})\ +\ \mathrm{P}(\overline{\mathrm{C}}/\ \mathrm{M_2})\mathrm{P}(\ \mathrm{M_2})) + \mathrm{P}(\overline{\mathrm{C}}/\ \mathrm{M_3})\mathrm{P}(\ \mathrm{M_3}))} = \frac{0.04}{0.056} = 0.7$$

Solution de la série 3 (2022-2023)

Exercice 1

a. A seul se réalise : $A \cap \bar{B} \cap \bar{C}$

b. A et C se réalisent mais pas B : $A \cap C \cap \bar{B}$

c. au moins l'un des trois événements se réalise : $A \cup B \cup C$

d. les trois événements se réalisent : $A \cap B \cap C$

e. aucun ne réalise : $\bar{A} \cap \bar{B} \cap \bar{C}$

f. au plus l'un des trois se réalise :

$\overline{(A \cap B) \cup (A \cap C) \cup (B \cap C)}$

g. au plus deux des trois se réalisent. $\overline{A \cap B \cap C} = \bar{A} \cup \bar{B} \cup \bar{C}$

Exercice 2

1. E1 $\cap$ E2 = (A $\cap \bar{B} \cap \bar{C}$) $\cap$ (A $\cap$ (B$\cup$C)) = (A $\cap \bar{B} \cap \bar{C} \cap$ B) $\cup$ (A $\cap \bar{B} \cap \bar{C}$ $\cap$ C) = $\varnothing \cup \varnothing = \varnothing$.

2. A $\cap$ B $\cap$ C = A $\cap$ ($\overline{B \cup C}$) donc en appelant K = B $\cup$ C , on a E1 $\cup$ E2 = (A $\cap \bar{K}$) $\cup$ (A $\cap$ K) = A.

3. On calcule P (B$\cup$C) = 0,4 + 0,3 – 0,1 = 0,6 , P ($\overline{B \cup C}$) = 0,4 ; P (E1) + P (E2) = P (A) = 0,6 .

P (A$\cup$ K) = P (A$\cup$ B$\cup$C)

= P (A$\cup$ B$\cup$C) = P (A) + P (B) + P (C) – P (A$\cap$ B) – P (B$\cap$C) – P (C $\cap$ A) + P (A$\cap$ B$\cap$C) .

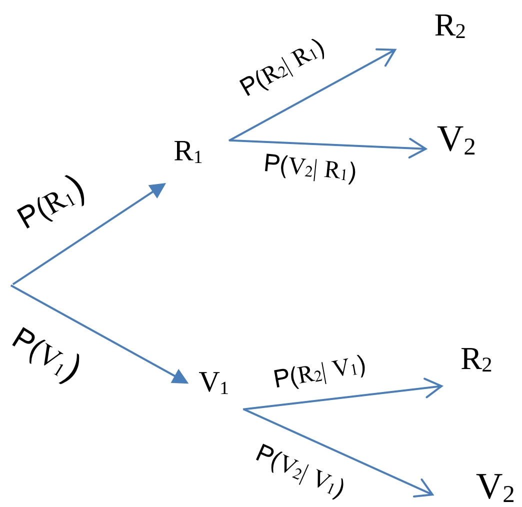
0,6 + 0,4 + 0,3 – 0,1 – 0,1 – 0,2 + 0,05 = 0,95 ;

par ailleurs

P (A∪ K) = P (A) + P (K) – P (A∩ K)⇒ 0,95 = 0,6 + 0,6 – P (E2) ⇒ P (E2) = 0,25

et enfin P (E1) = 0,6 – 0,25 = 0,35 .

Exercice 3 : Partie A



Soient les événement suivants :

R_1 : « On obtient une boule rouge dans le premier tirage »

R_2 : « On obtient une boule rouge dans le deuxième tirage »

1. **La probabilité que la 2ème boule est rouge sachant que la première est rouge c'est**

$P(R_2| R_1)$

Les cas possibles pour obtenir une boule rouge dans le premier tirage c'est 4 parmi le total 7, donc il reste 3 boules rouges parmi 6 dans le 2ème tirage

Qui ce implique $P(R2|\ R1)=\frac{3}{6}=\frac{1}{2}$

1. **La probabilité que ce soit les deux boules rouges c'est**

$$P(R1 \cap R2) = \mathrm{P(R2|\ R1)} \times \mathrm{P(R1)} = \frac{4}{7} \times \frac{1}{2} = \frac{2}{7}$$

Partie B :

$\Omega = \{PP, PF, FP, FF\}$, card$(\Omega) = 4$

$A = \{PP, PF\}$, card$(A) = 2$ $\quad A \cap B = \{PP\}$, $card(A \cap B) = 1$

$B = \{PP, FP\}$, card$(B) = 2$

$C = \{PP\}$, card$(C) = 1$ $\quad A \cap C = \{PP\}$, $card(A \cap c) = 1$

$$P(A \cap B) = \frac{card(A \cap B}{\mathrm{card}(\Omega)} = \frac{1}{4}$$

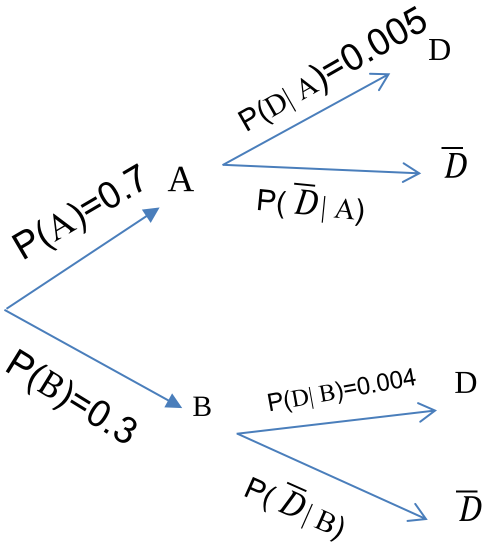
$$\frac{1}{4} = P(A) \times P(B) = \frac{2}{4} \times \frac{2}{4} = \frac{1}{4}$$

Donc A et B sont indépendants

$$P(A \cap C) = \frac{card(A\cap C)}{\mathrm{card}(\Omega)} = \frac{1}{4}; P(A) \times P(C) = \frac{2}{4} \times \frac{1}{4} = \frac{1}{8}$$

$P(A \cap C) \neq P(A) \times P(C)$; **donc A et C ne sont pas indépendants**

Exercice 4



Soient les événements suivants :

A: « Le composant est fabriqué par la ligne A »

B : « Le composant est fabriqué par la ligne B »

D : « Le composant présente un défaut »

Les données :

$P(A)=0.7$; $P(B)=0.3$; $P(D|A)=0.005$; $P(D|B)=0.004$

1. **La probabilité que le composant présente un défaut c'est $P(D)$**

On appulique la formule de la probabilité totale

$$P(D) = P(D|A) \times P(A) + P(D|B) \times P(B)$$

$$= 0.005 \times 0.7 + 0.004 \times 0.3 = 0.0047$$

1. **La probabilité que le compsant est fabriqué par la ligne A sachant présente un défaut**

C'est $P(A|D)$, ;on applique la formule de Bayes

$$P(A|D) = \frac{P(D|A) \times P(A)}{P(D)} = \frac{0.005 \times 0.7}{0.0047} = 0.74$$

Corrigé-type d'examen

2021-2022

EXO 1 : (6 pt)

a- Quand les amplitudes sont inégales, pour dessiner l'histogramme

2. On corrige les effectifs (0.5 pt)

b- L'intervalle interquartile :

1. Contient 50% des observations (0.5 pt)

c- La représentation graphique correspondant au cas quantitatif continu est :

1. L'histogramme. (0.5 pt)

B) 1→F, 2→F, 3→V, 4→V (0.5×4 pt)

C) ►Indicateur de position : Minimum ; Moyenne ; Premier quartile ; Mode ; Médiane (0.25× 5 pt)

►Indicateur de dispersion ! Ecart-type ; Etendu ; Coefficient de variation ; Variance ; Ecart interquartile (0.25× 5 pt)

EX02 : (14 pt) La population est 620 employés (0.25 pt)

Le caractère étudié est le salaire journalier (0.25 pt)

La nature est quantitatif continu (0.25+0.25)

Les classes	n_i	f_i (0.5 pt)	$f_i^c \nearrow$ (1 pt)	a_i (0.25pt)	d_i (0.5 pt)	c_i (0.5pt)	$c_i \times f_i$ (1 pt)	$c_i^2 \times f_i$ (1 pt)
$[5,6[$	100	0.16	0.16	1	100	5.5	0.88	4.84
$[6,7[$	80	0.13	0.29	1	80	6.5	0.845	5.49
$[7,7.5[$	240	0.39	0.68	0.5	480	7.25	2.827	20.49
$[7.5,9[$	160	0.26	0.94	1.5	106.7	8.25	2.145	17.69
$[9,10[$	40 (0.25 pt)	0.06	1	1	40	9.5	0.57	5.41
TOTALE	620 (0.25 pt)						7.27	54

Réponse 3 : (0.5 pt) Déterminons le mode

La classe modale est $[7,7.5[$ car $d_{i\_} = 480 \gg$ (0.25 pt)

Qui ce implique $M_o = \frac{7+7.5}{2}$=7.25 (0.25 pt)

Réponse 4 (4.25pt)

► Déterminons la médiane graphiquement

R.Q : La médiane est la valeur de x qui relative au moitié de la population. Donc pour la trouvé graphiquement on trace les 2 courbes croissante et décroissante dans ce cas la médiane est la projection du point d'intersection sur l'axe des abscisses ou bien médiane est la projection du point d'intersection de la courbe cumulative et la droite $y = 0.5$

RQ : (noter sur le graphe) (1.5 pt)

Interprétation (1 5pt) : La moitié (50%) des employés de l'usine «A» perçoivent moins de 7.27×100=727 DA par jour et l'autre moitié (50%) perçoivent plus de 727 DA par jour

► **Déterminons la médiane par calcul**

$f_i^c \nearrow (M_e) = 0.5 \Rightarrow M_e \in [7, 7.5[$ **(0.25 pt)**

Par interpolation linéaire on a

$\frac{0.5-f_i^c\nearrow(7)}{f_i^c\nearrow(7.5)-f_i^c\nearrow(7)} = \frac{M_e-7}{7.5-7} \Rightarrow M_e = \frac{0.5-0.29}{0.68-0.29} \times 0.5 + 7 = 7.27 \in [7, 7.5[$ **(0.5+0.5 pt)**

Réponse 5 : (1. pt)

Le nombre d'employés qui perçoivent un salaire compris entre 600 et 800 DA par jour

C- à- d : $n_{[6,8]} = n_{[6,7[} + n_{[7,7.5[} + n_{[7.5,8]} = 80 + 240 + n_{[7.5,8]} = 320 + n_{[7.5,8]}$ **(0.25 pt)**

$n_{[7.5,8]} = ?$

$n_{[7.5,9]} \rightarrow a = 2.5 \rightarrow 160$

$n_{[7.5,8]} \rightarrow à = 1.5 \rightarrow X$ $\Rightarrow X = \frac{1.5\times160}{2.5} = 96$ **(0.5 pt)**

$\Rightarrow n_{[6,8]} = 320 + 96 =$ 416 employés **(0.25pt)**

Réponse 6 : (2pt) Pour la dispersion des salaires ; il faut calculer le coefficient de variation CV tel qu $CV = \frac{S}{\overline{X}}$ **(0.25pt)**

$\overline{X} = \sum_{i=1}^{5} c_i \; f_i = 7.27$ **(0.25+0.25pt)**

$S^2 = \sum_{i=1}^{5} c_i{}^2 f_i - (\overline{X})^2 = 54 - (7.27)^2 = 54 - 53 = 1$ **(0.5+0.5pt)**

$\Rightarrow S = 1$

$\Rightarrow CV = \frac{1}{7.27} \simeq 0.13$ **(0.25pt)**

Solution de rattrapage

Exercice 1

1. Il s'agit d'un diagramme en bâtons représentant la variable Nombre de joueurs sélectionnés. Les valeurs de cette variable vont de 0 à 6.

2. La variable représentée est le nombre de joueurs sélectionnés.

L'individu statistique est le centre de formation.

3. Les modalités sont : 0 , 1 , 2, 3 , 4, 5, 6.

Les Effectifs sont : 2, 7 , 12 , 16, 25, 12, 4

4. La moyenne de joueurs sélectionnés par centre de formation est

$\overline{X}=\sum_{i=1}^{7} x_i n_i=\frac{263}{78}=3,37$

5. Pour calculer l'écart type du nombre de joueurs sélectionnés, on calcule d'abord la variance. Une formule pour la variance est

$S^2=\sum_{i=1}^{7} {x_i}^2 n_i-\overline{X}^2$**; donc la variance vaut 13; 37 - (3; 37)^2= 2; 01 et l'écart type vaut**

$S=\sqrt{2;\ 01}$ **= 1; 42.**

6. Le mode M_o=4

Déterminons le 2ème quartile (la médiane) :

$$N=78=2p\Rightarrow p=\frac{78}{2}=39\Rightarrow Q_2=\frac{x_{39}+x_{40}}{2}=\frac{4+4}{2}=4$$

Exercice 2

Les types possibles sont : qualitatif nominal (QN), qualitatif ordinal (QO), quantitatif discret (QD) et quantitatif continu (QC).

1. QN

2. QD ou QC

3. QO

4. QC

Corrigé-type d'examen (2022-2023)

Exercice :1 (12pt)

RQ : SVP accepter les résultats si l'étudiant utilise les fréquences au lieu les effectifs

Salaire	$[800; 900[$	$[900; 1000[$	$[1000; 1050[$	$[1050; 1150[$	$[1150; 1300[$
Effectif	42	49	74	19	16
$n_i^c \nearrow$ (0.5 pt)	42	91	165	184	200
a_i (0.25pt)	100	100	50	100	150
c_i (0.5 pt)	850	950	1025	1100	1225
d_i (0.5 pt)	0.42	0.49	1.48	0.19	0.10

1)

- La population : les 200 employés ……………………………….. (0.25pt)

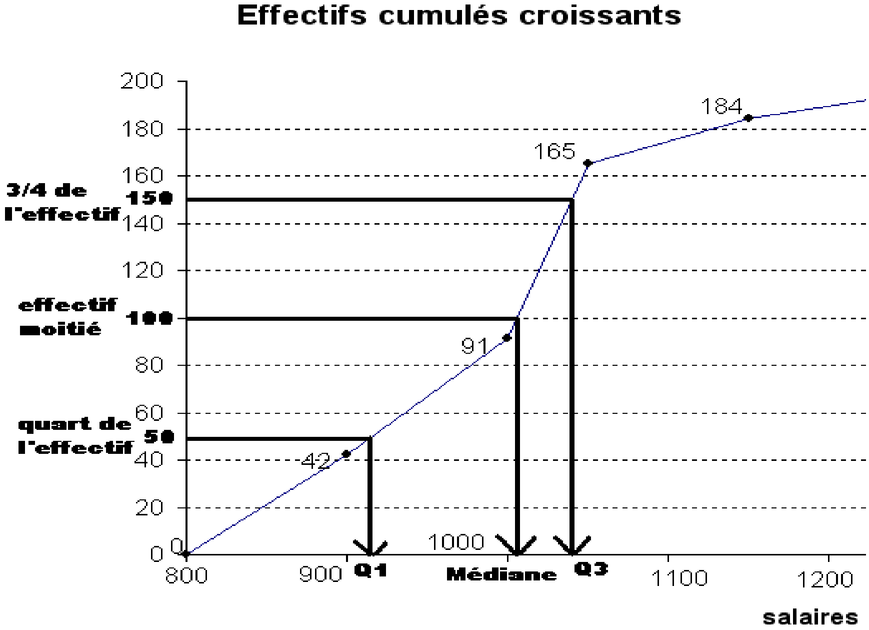
- **Le caractère étudié : les salaires mensuels …………… (0.25pt)**
- **Sa nature : quantitatif continu ……………………………. (0.25pt)**

2) Le nombre d'employés gagnent au plus 1050 euros c'est

$n_{x<1050} = 42 + 49 + 74 = 165$ **employés ………………………..(0.5pt)**

Ou bien $n_i^c \nearrow_{x<1050} = 165$ **employés**

3) Détermination des quartiles graphiquement (1.25 pt =0.25 pour chaque quartile + 0.5 pour la courbe)



4) **Calculons les quartiles par l'interpolation linéaire**

a- **Le 1er quartile Q_1**

$n_i^c \nearrow (Q_1) = \frac{1}{4}N = 50 \Rightarrow Q_1 \in [900, 1000[$ ....(0.25 pt)

$\frac{Q_1-900}{1000-900} = \frac{n_i^c\nearrow(Q_1)-n_i^c\nearrow(900)}{n_i^c\nearrow(1000)-n_i^c\nearrow(900)} = \frac{50-42}{91-42} = \frac{8}{49} = 0.16$ ....(0.75 pt)

$\Rightarrow Q_1 = 0.16\times100+900 = 916$ euro ....(0.25 pt)

b- **La médiane M_e**

$n_i^c \nearrow (M_2) = \frac{N}{2} = 100 \Rightarrow \text{Me} \in [1000, 1050[$ ....(0.25 pt)

$\frac{Q_1-1000}{1050-1000} = \frac{n_i^c\nearrow(Q_1)-n_i^c\nearrow(1000)}{n_i^c\nearrow(1050)-n_i^c\nearrow(1000)} = \frac{100-91}{165-91} = \frac{9}{74} = 0.12$ ....(0.75 pt)

$\Rightarrow M_2 = 0.12\times 50 + 1000 = 1006$ ....(0.25 pt)

c- **Le 3ème quartile Q_3**

$n_i^c \nearrow (Q_3) = \frac{3N}{4} = 150 \Rightarrow Q_3 \in$

$[1000, 1050[$...(0.25pt)

$$\frac{Q_3-1000}{1050-1000}=\frac{n_i^c\nearrow(Q_{13})-n_i^c\nearrow(1000)}{n_i^c\nearrow(1050)-n_i^c\nearrow(1000)}=\frac{150-91}{165-91}=\frac{59}{74}=0.8$$ ....(0.75 pt)

$\Rightarrow Q_3$=0.8$\times$ $50+1000=1040$ ....(0.25pt)

5) **Calculons le salaire moyen**

$\overline{X}=\frac{1}{N}\sum_{i=1}^{5} c_i n_i=\frac{42\times 850+49\times 950+\cdots 16\times 1225}{200}=\frac{198600}{200}=993.$ ....(0.5pt)

Pour que la distribution soit symétrique il suffit de vérifier que

$$\overline{X}=M_e=M_o$$;................ (0.25pt)

donc il reste de calculer M_o

La classe modale c'est $[1000;1050[$ **car** d_i**=1.48** $\gg$ **(les amplitudes sont différentes)** ......(0.25pt)

$\Rightarrow M_o=\frac{1000+1050}{2}=1025$ . ...(0.25pt)

On trouve $\overline{X}\neq M_o\neq M_e$ **, on déduire que la distribution est antisymétrique**.......(0.25pt)

6) **Calculons l'écart –type**

$S=\sqrt{S^2}$ et $S^2=\frac{1}{N}\sum_{i=1}^{5} c_i^2\, n_i-(\overline{X})^2=\frac{42\times 850^2+49\times 950^2+\cdots 16\times 1225^2}{200}-(993)^2$

$=$ **10519,75**.............................(0.25pt)

7) **Calculons l'effectif** n **pour la classe** $[1300;1500[$ **tel que le salaire moyen devient 1200 euros**

$\overline{Y}=1200=\frac{993\times200+1400\times n}{200+n}\Leftrightarrow 198600+1400n=240000+1200\,n\Leftrightarrow 200n=41400\Leftrightarrow n=207$

Il y aura donc 207 personnes dont le revenu appartient à la tranche [1300 ; 1500[. ......(1.25pt)

Exercice 2 (3pts)

1. **A et B sont incompatibles** $\Leftrightarrow A \cap B = \emptyset$ ….. …… $(0.5pt)$

   ► $P(A \cap B)= P(\emptyset)=0$ ….. …… $(0.5pt)$

   ► $P(A \cup B)= P(A) + P(B) - P(A \cap B)=0.3+0.4-0=0.7$ ….. …… $(0.5pt)$

2. **A et B sont indépendants** $\Leftrightarrow P(A \cap B) = P(A) \times P(B)$ ….. …… $(0.5pt)$

   $P(A \cap B) = P(A) \times P(B)=0.3\times0.7=0.21$ ….. …… $(0.5pt)$

   $P(A \cup B)= P(A) + P(B) - P(A) \times P(B)=0.7-0.21=0.49$ ….. …… $(0.5pt)$

Exercice 3 :

Soient les événements suivants :

G : « L'étudiant choisi est un garçon » $P(G) = \frac{15}{36}$

F : « L'étudiant choisi est une fille » $P(F) = 1 - \frac{15}{36}=\frac{21}{36}$

D : « L'étudiant choisi fait le devoir » $P(D|G) = 0.2$;

$P(D|F)$ =0 .25

1. **Calculons la probabilité que l'étudiant choisi fait le devoir**

   On applique la formule des probabilités totale suivante

$$P(D) = P(D|G)\times P(G) + P(D|F)\times P(F) = \frac{2}{10}\times\frac{15}{36}+\frac{1}{4}\times\frac{21}{36}\cong 0.23$$

2. **Calculons la probabilité que l'étudiant choisi est une fille sachant qu'elle fasse le devoir**

   On applique la formule de bayes

$$P(F|D) = \frac{P(D|F)\times P(F)}{P(D)}=\frac{0.25\times0.145}{0.23}=0.63$$

3. **Calculons la probabilité que l'étudiant choisi est un garçon et fait le devoir**

$$P(G \cap D) = P(D|G) \times P(G) = \frac{2}{10} \times \frac{15}{36} = 0.083$$

Micro-interrogation

x_i	0	1	2	3	4	5	TOT	
n_i	69	41	19	13	8	3	153	
f_i	0.451	0.268	0.124	0.085	0.052	0.02	1	0.5pt
F(x)	0 .451	0.719	0.843	0.928	0.98	1		0.5pt
$n_i^c\nearrow$	69	110	129	142	150	153		0.5pt

1)

-La population : 153 machines (0.25 pt

-La variable étudié est le nombre de pannes 0.25 pt

-Les modalités : 0, 1, 2, 3, 4, 5 0.25 pt

-Le type de caractère est quantitatif discret 0.25 +0.25pt

2) 2 pt

$$F(x)=\begin{cases} 0 & si \quad x < 0 \\ 0.451 & si \quad 0 \le x < 1 \\ 0.719 & si \quad 1 \le x < 2 \\ 0.843 & si \quad 2 \le x < 3 \\ 0.928 & si \quad 3 \le x < 4 \\ 0.98 & si \quad 4 \le x < 5 \\ 1 & si \quad x > 5 \end{cases}$$

Ou bien $F(M_e)=0.5 \Rightarrow M_e$

3) Les paramètre de tendance centrale :

► Le mode M_o =0 car n=69≫ 0.5 pt

►La médiane M_e

N=153 impair =2p+1⇒P=$\frac{152}{2}$=76

$M_e=x_{p+1}=x_{77}=1$ **1pt**

Ou bien $F(M_e)=0.5 \Rightarrow M_e=1$

►La moyenne $\overline{X}$ **1pt**

$\overline{X} = \frac{1}{N}\sum_{i=1}^{6} x_i n_i = \frac{1}{153}(0 \times 69 + \cdots 5 \times 3) \cong 1.08$

4)

Le pourcentage des machines qui ont au moins deux pannes

$f_{x\geq 2}$=124+0.085+0.052+0.02=1-(0.451+0.268)=0.281 **1pt**

P=0.281×100=28.1% **0.25pt**

La courbe cumulative **1 .5pt**

Bibliographie

[1] G. Chauvat and J.-P. Reau, ***Statistiques descriptives, Armand Colin, 2002.***

[2] M. Tenenhaus, ***statistique : Méthodes pour décrire, expliquer et prévoir, Dunod, 2006.***

[3] J.-J. Droesbeke, ***Éléments de statistiques, Ellipses, 2001.***

[4] G. Calot, ***Cours de statistique descriptive, Dunod, 1969.***

[5] J. Vaillant, ***Eléments de Statistique descriptive, 2015.***

[6] L. Leboucher and M.-J. Voisin, ***Introduction à la statistique descriptive, 2013.***

[7] F. Mazerolle, ***Statistique descriptive, 2009.***

[8] B. Oukacha and M. Benmessaoud, ***Statistique descriptive et calcul des probabilités, 2013.***