République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Abdelhamid ibn badis Mostaganem

Faculté des Sciences exactes et de l'informatique
Département de mathématique

Polycopié de cours intitulé :
Introduction aux probabilités et statistique descriptive

Présenté par

Bouanani Oussama

Année universitaire 2021-2022

Information sur la matière

Faculté des Sciences exactes et de l'informatique (FSEI).

Département : Mathématiques et informatique.

Public cible : Première année licence de Mathématiques et Informatique (MI).

Intitulé du cours : Introduction aux probabilités et statistique descriptive.

Crédit :03

Coefficient :02

Durée : 14 semaines

Horaire :

Jeudi : 8h30-13h30

Salle : Amphi2

Enseignant : ***Cours, TD*** :Dr .O Bouanani

Contact :

1. oussama.bouanani@univ-saida.dz
2. bouananio70a@gmail.com

Disponibilité :

Au Laboratoire : Dimanche(matin), lundi, Mercredi et jeudi (toute la journée)

Présentation du cours

Ce cours intitulé ⧫ Introduction aux probabilités et statistique descriptive ⧫ a pour objectif d'initier des étudiants de la première année mathématique et informatique (MI) les principes fondamentaux en probabilités et en séries statistiques à une variable (unidimensionnelle). Notons que la statistique est définie comme un ensemble de méthodes scientifiques où un ensemble de techniques basées sur le recueil, l'organisation, la présentation, l'analyse et interpréter les données afin de rendre les données compréhensibles par tous. Il nous permet aussi de prendre des décisions éclairées dans des conditions d'incertitude. Cette science est un outil essentiel pour la compréhension et la gestion des phénomènes complexes. En conséquence, ce cours est nécessaire que les étudiants des autres filières en général et les étudiants (MI) en particulier, doivent acquérir auprès d'eux les concepts de base de manière à les aider à prendre les décisions appropriées dans leur domaine de spécialisation.

Les unités d'apprentissage qui constituent ce cours permettent d'acquérir un ensemble de compétences. En d'autre terme, l'étudiant devrait être capable, à la fin du cours, de :

- Apprendre le vocabulaire utilisé dans la statistique descriptive à une série statistique simple.
- Figurer des données sous forme de tableaux statistiques.
- Représenter des données par un graphique approprié.
- Analyser des données statistiques par des paramètres statistiques appropriés.
- Apprendre les techniques et méthodes de dénombrement.
- S'initier aux notions de bases de calcul des probabilités

Mots-clés : Statistique descriptive- Paramètres de position - Paramètres de dispersion
Dénombrement- Probabilités

Contenu

La figure ci-dessous (page-05-) présente les différentes unités d'apprentissages et les séquences pédagogiques qui les correspondants :

Pré-requis

Pour pouvoir tirer le maximum de ce cours il faut connaître

- ▶ Connaissances de mathématiques de terminale
- ▶ Les notions de base relatives aux calculs les sommes
- ▶ Techniques ensemblistes : opérations ensemblistes, relations (Algèbre1. S1)

Visées d'apprentissage

Dans un contexte global, la compétence visée par ce cours est d'être capable de présenter et résumer les données statistiques, tirer des conclusions sur la population étudiée et d'aider à la prise de décision et modéliser un phénomène aléatoire, le quel peut se décrire par une loi de la variable aléatoire.
En d'autre terme, le cours introduction aux probabilités et statistique descriptive vise à vous :

♠ **En termes de connaissances**

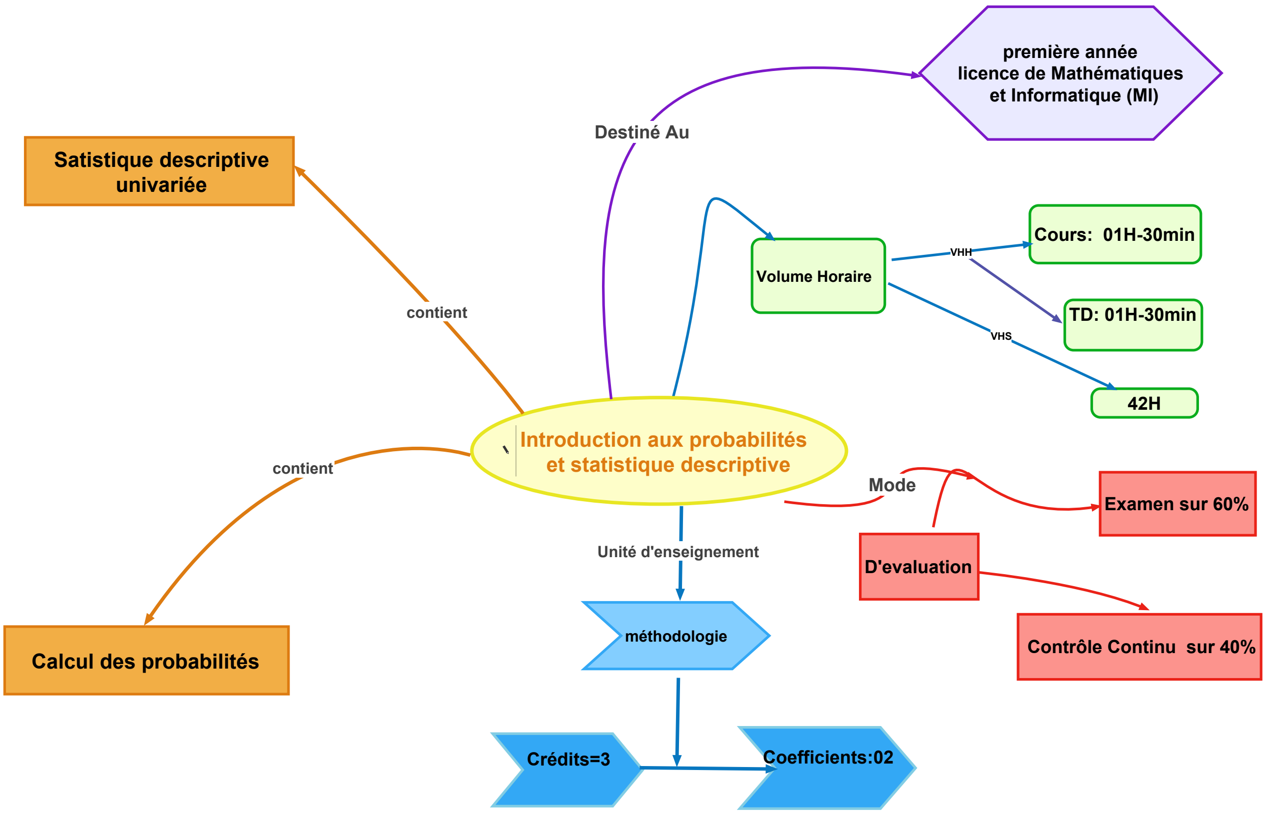
- ▶ Apprendre les notions de base en probabilités et statistique descriptive, lesquelles permettant la modéliser (solution) des problèmes aléatoires issus de différentes disciplines.

♠ **En termes de savoir-faire**

- ▶ Entrainer à l'application des outils probabiliste et les techniques statistiques pour la résolution des problèmes qui présentent des phénomènes non déterministes.
- ▶ Orienter vers la phase de la programmation avec langage R statistique

♠ **En termes de savoir-être**

- ▶ Sensibiliser au respect des exigences d'un commanditaire (spécificités du système, contraintes de fonctionnement,...).



Carte conceptuelle de cours : Stat-proba
Réaliser par :O. BOUANANI

Mathématiques et Informatique	Faculté des Sciences Exactes et del'Informatique

Modalités d'évaluation des apprentissages

L'évaluation se fera au moyen d'un examen final et d'une note de contrôle continu dont la pondération est de 60% pour l'examen, 40% pour le contrôle continu (voir Figure 2 (a)). De amples détails sont décrits plus bas.

- **Examen final** : Il porte sur tout ce qui a été vu dans le cours le long du semestre, il contient :
  1. Des problèmes similaires ou proches à ceux traité lors des TD, Tests et interrogations.
  2. Des exercices et questions de réflexion lesquelles ont été déjà traité oralement lors des séances de cours.
- **Contrôle continue** : Il comporte deux examens partiels et un devoir collectif dont la pondération est de 40% pour chaque examen partiel et 20% pour le devoir collectif (voir Figure 2 (b)).

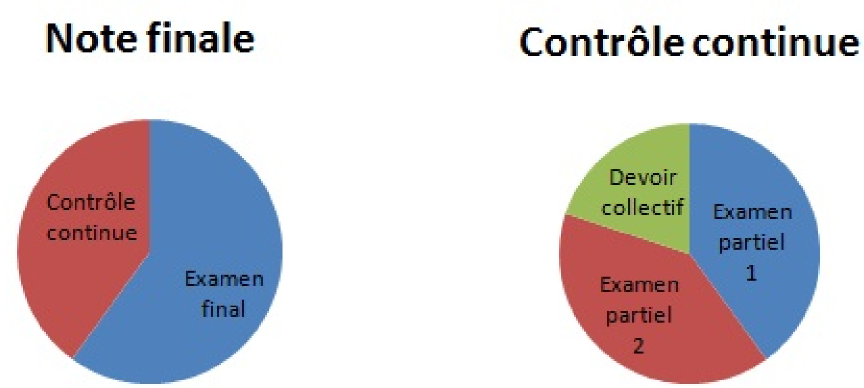


Figure 2 : Modalité d'évaluation.
(a). Evaluation note finale, (b). Evaluation contrôle continue.

Activités d'enseignement-apprentissage

Pour une bonne assimilation des concepts de la matière introduction aux probabilités et statistique descriptive, le cours propose plusieurs méthodes et manières d'apprentissages ayant chacune leurs spécificités et avantages.

- **Cours magistrale** : (1h-30 par semaine) Les connaissances et les La théorie, les méthodes et les exemples d'illustrations sont présentés et étayés.
  1- Une prise de note est obligatoire. Elle servira à maîtriser les concepts indispensables et à réaliser les activités d'apprentissages proposées.
  2- Un débat sera ouvert à la fin de chaque séance. La participation de tous les étudiants est souhaitée dans le but de répondre librement et s'échanger les connaissances et les points vue, de s'auto évaluer et se corriger les notions mal assimiler.
- **Travaux dirigés** : (1h-30 par semaine) Les savoir-faire sont transmis à travers des travaux dirigés :
  1- Un ensemble de différents exercices est programmées à la fin de chaque unité pour pouvoir vérifier la capacité à mobiliser les savoirs dans la résolution des problèmes et exercices proposées.
  2- Des devoirs et exercices individuels seront proposés pour permettre le développement de l'autonomie des étudiants et la réflexion de chacun.
- **Devoirs collectifs** : Les savoir-être sont transmis à travers des devoir collectifs :
  1- Des devoirs collectifs sont proposés pour vous permettre de travailler sur de nouveaux problèmes ainsi de développer les compétences
  2- Le travail en équipe permettra d'échanger les idées, de corriger les notions mal assimilés et d'assurer une bonne qualité de production qui servira et sera utiles dans la vie professionnelle.

Le tableau ci-dessous résumé les activités d'enseignement-apprentissage.

Modalités	Objectifs / Méthodes
Savoir	Cours magistral-Prise de notes-Débats en séance de cours
Savoir-faire	TD -Projet individuel
Savoir-être	Projet collectifs

Bibliographie et Ressources

- Plusieurs ressources sont mises à votre disposition au niveau de la bibliothèque universitaire.
- Les références suivantes peuvent être utiles mais ne sont pas obligatoires :

1 A. Ayache et J. Hamonier. Cours : Statistique Descriptive et Calcul de Probabilités. Université de Lille. France, 2014.

2 B. Tilière et D. Godhino. Cours : Statistique descriptive et probabilités. Université Paris-Est-Créteil. ESIPE https ://www.ceremade.dauphine.fr.

3 B. Etienne , J.Claude et R. Gillet . Livre : Statistique descriptive. Pearson France. ISBN 27440749427. (2010)

4 C. Suquet. Introduction au calcul de probabilités. Université des Sciences et Technologies de Lille U.F.R. de Mathématiques Pures et Appliquées. France, (2003).

5 D. Fredon, M. M. -Bertrand, F. Bertrand. Mathématiques Statistique et probabilités en 30 fiches. Dunod, Paris, ISBN 978-2-10-054257-4, (2009).

6 J. Escoffier. Livre : Probabilités et statistiques pour le CAPES externe et l'Agrégation interne de Mathématiques .ISBN 978-2-7298-5850-6.(2010).

7 J. Pierre. Statistique et probabilités ; 6e édition.Dunod, 11 rue Paul Bert, 92240 Malakoff, www.dunod.com ISBN 978-2-10-075259-1, (2016).

Table des figures

Chapitre 1

Statistique descriptive univariée

Sommaire

Ce chapitre est divisé en trois sections. Dans la première section, on présente terminologie de la statistique descriptive (vocabulaire statistique). une représentation graphique d'une variable est donnée dans la deuxième section. La troisième section est consacrée aux les paramètres de position, paramètres de dispersion et paramètres de forme

1.1 Concepts de base des statistiques

1.1.1 Population

- Le terme population (univers statistique) désigne l'ensemble sur lequel porte notre étude statistique ou l'ensemble sur lequel on étudie un phénomène statistique.
- Les éléments de la population sont appelés individus ou unités statistiques.
- Ensemble des individus (populations) observés sont généralement désignés par Ω.
- Le nombre d'éléments ($\omega \in \Omega$) de la population est notée N (effectif total de la population).

Exemple 1.1.1. ♣

1. *L'ensemble des étudiants de l'Université Mostaganem de la filière (MI) .*
2. *Matériel informatique produit par l'entreprise.*

Remarques 1. ⋆
L'ensemble Ω peut être un groupe de personnes, d'entreprises, de familles, de choses ou des animaux.

1.1.2 Individu (unité statistique)

Chaque composante (membre, élément) d'une population est appelée unité statistique.

1.1.3 Échantillon

C'est un sous-ensemble d'une population. Le nombre des éléments d'un échantillon est noté par n.

1.1.4 Variable statistique(Caractère)

Un caractère statistique est une caractéristique possédée par les éléments(unités statistiques) d'une population. Alors une variable statistique (V. S) était vue comme une application

$$X : \Omega \longmapsto C$$

.

L'ensemble C est appelé : Un ensemble de valeurs pour la variable statistique X.

Exemple 1.1.2. ♣
Les notes des étudiants à l'examen de statistique, Mentions, Taille, température, nationalité, couleur des yeux, situation familiale, salaire... etc.

1.1.5 Modalités

Les modalités d'un caractère statistique correspondent aux diverses situations (valeurs) qui peuvent être supposées.

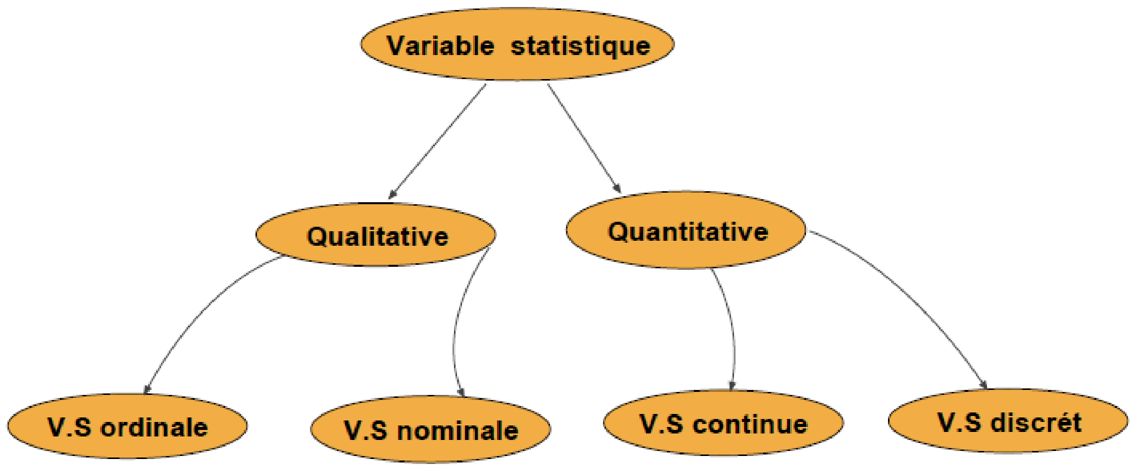
Exemple 1.1.3. ♣
– *Caractère est " le groupe sanguin d'un individu".*
– *Modalités sont " A, B, AB, O"*
– *Caractère est " Nombre d'absences dS étudiants "*
– *Modalités sont " 0 1,3,4 ".*

1.1.6 Types des caractères

On distingue les variables selon qu'elles prennent des valeurs numériques ou non. On dit qu'une variable est

1- **Qualitative** lorsque ses valeurs ne sont pas des nombres (admettent des modalités non mesurables). On distingue deux types de variables qualitatives :
   – Variable qualitative **ordinale** lorsque ses modalités peuvent être ordonnées, par exemple : niveau d'éducation, les grades militaires, les taux de réussite, etc.
   – Variable qualitative **nominale** lorsque ses les modalités ne peuvent pas être ordonnées , tels que la nationalité, la situation familiale, le sexe, la couleur.

2- **Quantitatif** lorsque ses valeurs sont des nombres (admettent des modalités mesurables). On distingue deux types de variables :
   – Variable quantitative discrète : Si l'ensemble des valeurs possibles est dénombrable (ne prendre que des valeurs isolées), par exemple : nombre d'enfants dans la famille, nombre de pièces de la maison, nombre d'accidents de la circulation.
   – Variable quantitative continue : Si l'ensemble des valeurs possibles est continu (peuvent prendre toute valeur dans un intervalle), par exemple : taille, âge, poids,..,etc.



1.2 Étude d'une variable statistique quantitative discrète

1.2.1 Série statistique

Soit X une variable mesurée sur une population Ω de taille N. L'ensemble E qui regroupe toutes les observations de la variable sur la population est appelé : série statistique. Cet ensemble s'écrit :

$$E = \{x_i, i = 1, n\}.$$

Les techniques qui permettent :

- D'expliquer (décrire) cet ensemble (forme, variation,asymétrie, concentration, ...etc.)
- De résumer l'information contenue dans cet ensemble par des nombres, par un tableau statistique ou par des graphiques sont regroupées dans ce qu'on appelle : **statistique descriptive à une variable (unidimensionnelle).**

1.2.2 Distributions statistiques

Une distribution statistique est une représentation des données collectées dans un tableau où figurent la variable, l'effectif et la fréquence associée à chaque valeur prise par cette variable.

1.2.2.1 Effectif et fréquences relatives d'une modalité

- On appelle effectif d'une modalité x_i le nombre d'individus dans la population qui possèdent cette modalité x_i. Elle est notée n_i et donnée par :

$$n_i = card\{w \in \Omega, X(w) = x_i\}.$$

où card(A) :=nombre d'éléments de A.

- On appelle la fréquence relative ou (effectif relatif) d'une modalité x_i le rapport de l'effectif cette modalité par à l'effectif total(N). Elle est notée f_i et donnée par : $f_i = \dfrac{n_i}{N}$, (en % $f_i \times 100\%$) où effectif total est donné par

$$N = \text{card}(\Omega) = \sum_{i=1}^{k} n_i \text{ et } \sum_{i=1}^{k} f_i = 1$$

1.2.2.2 Effectif et Fréquence cumulée croissante d'une modalité

On appelle effectif cumulée croissante (respective, fréquence cumulée croissante) d'une modalité x_i la somme des effectifs (respectifs, fréquence) de toutes les valeurs inférieures ou égales à x_i.Elle est notée $N_{ic}^{\nearrow}$ (respective, $f_{ic}^{\nearrow}$) et donnée par : $N_{ic}^{\nearrow} = \displaystyle\sum_{k=1}^{i} n_k$, (respective, $f_{ic}^{\nearrow} = \displaystyle\sum_{k=1}^{i} f_k$.)

1.2.2.3 Effectif et Fréquence cumulée décroissante d'une modalité

On appelle effectif cumulée décroissante (respective, fréquence) d'une modalité x_i la somme des effectifs (respectifs, fréquence) de toutes les valeurs supérieures ou égales à x_i. Elle est notée $N_{ic}^{\searrow}$ (respective, $f_{ic}^{\searrow}$ et donnée par : $N_{ic}^{\searrow} = \displaystyle\sum_{j=i}^{n} n_j$, (respective, $f_{ic}^{\searrow} = \displaystyle\sum_{j=i}^{n} f_j$.)

1.2.2.4 Fonction de répartition empirique d'un caractère discret

Une fonction de répartition empirique représentait les fréquences cumulées croissantes, qui à tout nombre réel x associe la proportion des objets de la population pour lesquels la valeur du caractère X inférieure ou égal x. Par conséquent la fonction de répartition empirique FX est définie de $\mathbb{R}$ dans $[0, 1]$ et est donné par :

- $\forall x \in]-\infty, x_1[,\ F_X(x) = 0,$
- $\forall i \in \{1, .., n-1\},\ \forall x \in [x_i, x_{i+1}[, F_X(x) = f_1 + ... + f_i,$
- $\forall x \in [x_n, +\infty[,\ F_X(x) = 1.$
- Fonction de répartition empirique d'un caractère quantitative discrète est une fonction croissante, continue à droite.

1.2.2.5 Fonction de répartition empirique cas continue

On suppose que le caractère X est quantitatif continue et a pour classes

$$[a_0, a_1[,\ [a_0, a_1[, ... , [a_{n-1}, a_n[.$$

La fonction de répartition empirique F_X est définie de $\mathbb{R}$ dans$[0,1]$ et est donné par :
- $\forall x \in]-\infty, x_1[,\ F_X(x) = 0,$
- $\forall i \in \{1, .., n-1\},\ \forall x \in [a_i, a_{i+1}[,\ F_X(x) = \sum_{j=1}^{i} f_j + f_{i+1}\left(\frac{(x-a_i)}{a_{i+1}-a_i}\right)$
- $\forall x \in [x_n, +\infty[,\ F_X(x) = 1.$
- Fonction de répartition empirique d'une variable X quantitative continue est une fonction croissante et continue sur $\mathbb{R}$

1.2.3 Présentation des données statistiques

1.2.3.1 Tableaux statistiques

Les tableaux statistiques permet de décrire et de synthétiser des données relatives à un seul phénomène, qu'il soit qualitatif ou quantitatif.

1. **Tableau d'une variable statistique qualitative ordinale**

Modalités	n_i	$N_{ic}^{\nearrow}$	$N_{ic}^{\searrow}$	f_i	$f_{ic}^{\nearrow}$	$f_{ic}^{\searrow}$
Catégorie 1	n_1	n_1	N	f_1	f_1	1
Catégorie i	n_i	$\sum_{k=1}^{i} n_k$	$\sum_{j=i}^{m} n_j$	$f_i = n_i/N$	$\sum_{k=1}^{i} f_k$	$\sum_{j=i}^{m} f_j$
Catégorie m	n_m	N	n_1	f_m	$\sum_{k=1}^{m} f_k$	f_m
Total Σ	$N = \sum_{i=1}^{m} n_i$	.	.	$\sum_{i=1}^{m} f_i = 1$	1	1

Exemple 1.2.1. ♣ *On interroge 50 personnes sur leur le niveau d'éducation. On a obtenu la série statistique suivant : Moyen(15), Secondaire(15), Études supérieures(2), Primaire(13), Universitaire(5)*
- ***Population*** *: 50 Personnes.*
- ***Individu*** *: Personne.*
- ***Caractère étudie (V.S)*** *: Le niveau d'éducation (primaire-moyen-secondaire-universitaire-Études Supérieur) est une variable qualitative ordinale.*

Distribution de niveau d'éducation sur un échantillon de 50 personnes

Modalités	n_i	$N_{ic}^{\nearrow}$	$N_{ic}^{\searrow}$	f_i	$p_i\%$	$f_{ic}^{\nearrow}$	$f_{ic}^{\searrow}$
Primaire	*13*	*13*	*50*	13/50	26%	*13/50*	*1*
Moyen	*15*	*28*	*37*	15/50	30%	*28/50*	*37/50*
Secondaire	*15*	*43*	*22*	15/50	30%	*43/50*	*22/50*
Universitaire	*5*	*48*	*7*	5/50	10%	*48/50*	*7/50*
Études Supérieur	*2*	*50*	*2*	2/50	4%	*1*	*2/50*
Total Σ	*50*	.	.	*1*	*100*		

Lorsque la variable est nominale, n'y figurent pas les effectifs et fréquences cumulées

2. **Tableau d'une variable statistique quantitatif discret**

Exemple 1.2.2. ♣ *: La série statistique suivant représenter le nombre d'enfants dans une famille pour un échantillon de 24 familles :*

1, 4, 5, 2, 6, 4, 1, 2, 6, 2, 3, 2, 1, 6, 3, 3, 3, 3, 1, 4, 2, 4, 5, 5, 1, 6, 2, 4, 2, 2, 1, 3, 2, 6, 5, 1, 1, 5, 5, 3, 2, 5.

Tableau statistique

x_i	*1*	*2*	*3*	*4*	*5*	*6*
n_i	*8*	*10*	*7*	*5*	*7*	*5*
$N_{ic}^{\nearrow}$	*8*	*18*	*25*	*30*	*37*	*42*
f_i	*8/42*	*10/42*	*7/42*	*5/42*	*7/42*	*5/42*
$f_{ic}^{\nearrow}$	*8/42*	*18/42*	*25/42*	*30/42*	*37/42*	*1*
$f_{ic}^{\searrow}$	*1*	*34/42*	*24/42*	*17/42*	*12/42*	*5/42*

3. **Tableau d'une variable statistique quantitative continue**
   – Comme mentionné précédemment, dans la variable continue quantitative, le domaine d'étude comprend une infinité de valeurs, parce qu'il n'est pas possible de développer toutes ces valeurs, le domaine d'étude est divisé sous forme des intervalles. Cet intervalle s'appelle une classe et le nombre des classes est déterminé selon la taille de l'échantillon et selon la répartition des unités statistiques sur le domaine d'étudier. Pour construire ces classes, on respecte les règles suivantes :

1. Le nombre de classes est compris entre 5 et 20 (de préférence entre 6 et 12).

2. Chaque fois que cela est possible, les amplitudes des classes sont égales.
3. Chaque classe (sauf la dernière) contient sa borne inférieure mais pas sa borne supérieure. La répartition en classes des données nécessite de définir a priori le nombre de classes K et donc l'amplitude de chaque classe. En règle générale, on choisit au moins cinq classes de même amplitude. Cependant, il existe des formules qui nous permettent d'établir le nombre de classes et l'intervalle de classe (l'amplitude) pour une série d'observations de taille N
4. La règle de Sturge :$K = 1 + (3.3\, log_{10}(N))$.
5. La règle de Yule : $K = 2.5\sqrt[4]{N}$.

   L'intervalle de classe est obtenue ensuite de la manière suivante : $L = \dfrac{b-a}{K}$, où b (resp. a) désigne la plus grande (resp. la plus petite) valeur observée.

   Exemple 1.2.3. ♣ *Les données suivantes présente la quantité de production journalière pour une période de 30 jours dans un établissement industriel : 35, 34, 44, 33, 45, 36, 34, 33, 33, 35, 45, 44, 33, 33, 34, 43, 41, 43, 35, 34, 33, 33, 41, 42, 41, 43, 33, 34, 33, 36. La quantité de production journalière est une variable quantitative continue, et afin de figurer les données dans un tableau de fréquence. Nous suivons les étapes suivantes :*

   (a) *Calculer l'étendue de la population : e=$x_{min} - x_{max}$=45-33=12.*
   (b) *Calculer le nombre de la classe selon l'équation de Sturges :* $K = 1 + (3.3\, log_{10}(30)) \simeq 6$. *Donc, le nombre de classes égale 6.*
   (c) *Calculer L'amplitude de la classe :* $L = 12/6 = 2$. *Lors de la détermination de l'amplitude de la classe, l'inégalité suivante doit être prise en compte ;* $e \leq K \times L$.

   La classe commence par une valeur appelée la borne inférieure, et se termine par une valeur appelée la borne supérieure, puis nous trouvons que la borne inférieure de la première classe est inférieur ou égal à la plus petite valeur des données.

   Par conséquent, la borne inférieure de la première classe égale 33. D'une façon générale, désignons les extrémités des classes par $a_0, a_1, ..., a_n$ on a alors :

$$C_1 = [a_0, a_1[,\ C_2 = [a_0, a_1[,\ ..,C_n = [a_{n-1}, a_n[,$$

   – $a_0 = min(X(\Omega))$.
   – $a_i = a_{i-1} + L$ *avec* $i \in \{1, 2, .., n\}$.
   – *Le centre de la i-ème classe, elle est notée c_i et donnée par $c_i = (a_i - a_{i-1})/2$.*

Distribution des établissements selon la quantité de production journalière

Classes	n_i	$N_{ic}^{\nearrow}$	$N_{ic}^{\searrow}$	f_i	$p_i\%$	f_{ic}	f_{id}
$[33, 35[$	10	10	30	$1/3$	33%	$1/3$	1
$[35, 37[$	5	15	20	$1/6$	16%	$3/6$	$20/30$
$[37, 39[$	3	18	15	$1/10$	10%	$18/30$	$15/30$
$[39, 41[$	3	21	12	$1/10$	10%	$21/30$	$12/30$
$[41, 43[$	4	25	9	$2/15$	15%	$25/30$	$9/30$
$[43, 45]$	5	30	5	$1/6$	16%	1	$1/6$
Total Σ	30	.	.	1	100		

1.2.3.2 Représentation graphique

Les données statistiques peuvent être décrites et résumées à l'aide de graphiques et de formes géométrique. Cette dernière permet une analyse rapide du phénomène étudié et en utilise différentes type de graphique dépend du type de variable étudiée.

– **Une variable statistique qualitative**

  On utilise Deux diagrammes pour représenter une série d'observations d'une variable qualitative ; le diagramme par tuyaux d'orgue (diagramme à barres)(Voir 1.1) et le diagramme à circulaire (dit camembert)(Voir 1.2). Le principe de diagramme à barres est de représenter des aires proportionnelles aux fréquences (effectifs) de la variable statistique. Pour partager le diagramme circulaire en tranche ou secteur correspondant aux modalités observées, on utilise la formule suivante $\alpha_i = \dfrac{n_i \, 360°}{N} = f_i.\, 360°$

FIGURE 1.1 – Diagramme en tuyaux d'orgue

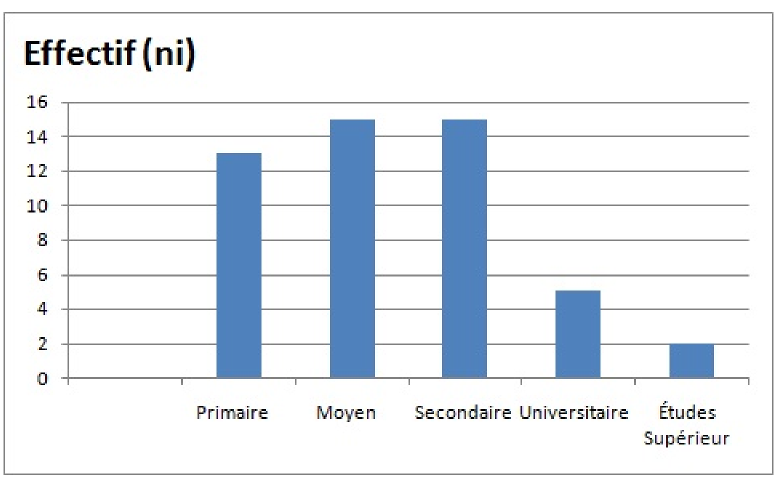
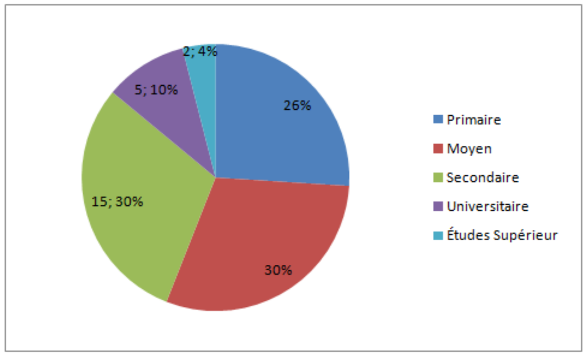


FIGURE 1.2 – Diagramme en secteurs



Pour partager le diagramme circulaire en tranche ou secteur correspondant aux modalités observées, on utilise la formule suivante $\alpha_i = \dfrac{n_i\,360°}{N} = f_i.\,360°$

– **Une variable statistique quantitative**

Nous distinguons deux types deux types de représentation graphique d'une distribution statistique à caractère quantitatif

– Diagramme différentiel : Concerner à la représentation les données du caractère X selon la fonction de distribution (effectif, fréquence).

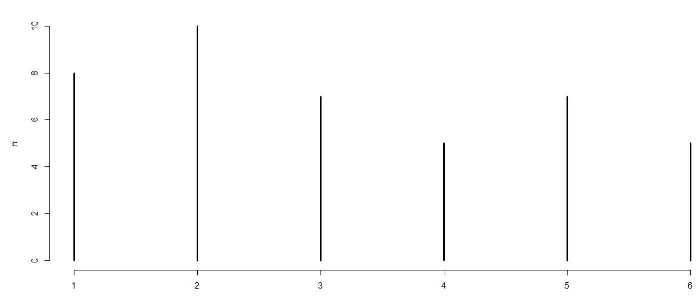
– Diagramme intégral : Concerner à la représentation les données du caractère X selon la fonction de distribution cumulative (effectif cumulée, fréquence cumulée).

1.2.4 Graphiques pour variables discrètes

a) Diagramme différentiel

– **Diagramme en bâtons :** Sur l'axe des abscisses on représente les modalités (les valeurs de la variable), alors que sur l'axe des ordonnées on représente les effectifs ou les fréquences (voir 1.3)

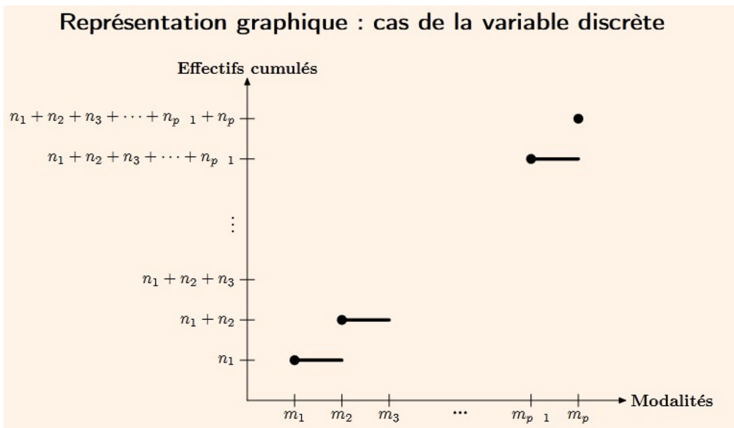
FIGURE 1.3 – Diagramme en bâtons



– **Polygone des fréquences (ou des effectifs)** : Le polygone des fréquences est ligne brisée joignant les sommets des bâtons

b) Diagramme intégral

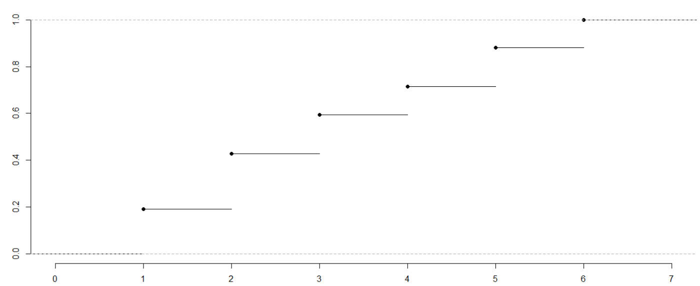
– **Diagramme des fréquences (ou des effectifs)cumulés croissants (resp. décroissant)** c'est une courbe en escalier qui représente la fonction de répartition empirique dans le cas de diagramme des fréquences cumulées croissantes.



Exemple 1.2.4. *Pour les données de l'exemple (1.2.2) la fonction de répartition de la variable statistique X est donnée par*

$$F(x) = \begin{cases} 0, & si\ x < 1 \\ 8/42, & si\ 1 \leqslant x < 2 \\ 18/42, & si\ 2 \leqslant x < 3 \\ 25/42, & si\ 3 \leqslant x < 4 \\ 30/42, & si\ 4 \leqslant x < 5 \\ 37/42, & si\ 5 \leqslant x < 6 \\ 1, & si\ x \geqslant 6 \end{cases}$$

FIGURE 1.4 – Courbe des fréquences cumulées croissante

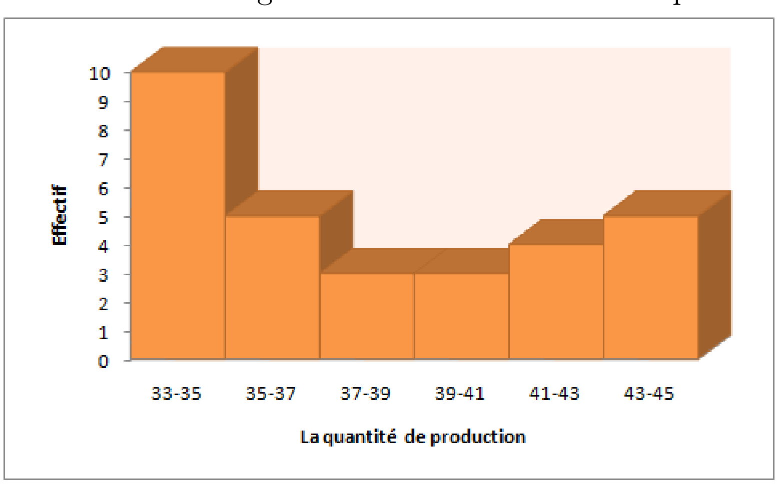


Entre les valeurs x_i et x_{i+1} la courbe est un segment de droite.

Représentations Graphiques pour variables continue

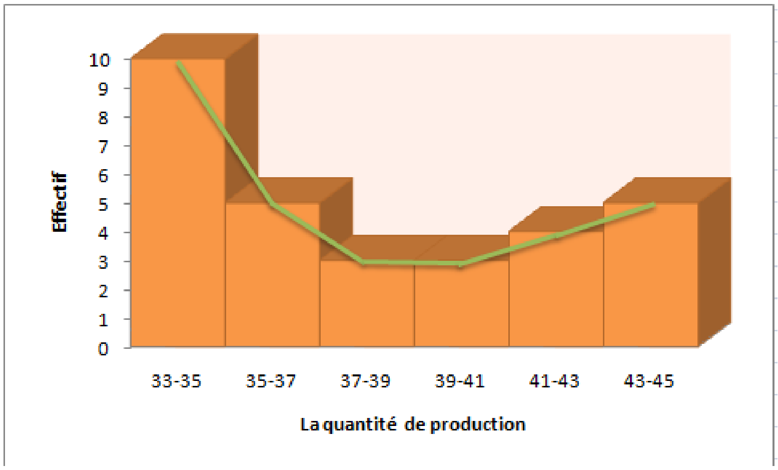
1. **Histogramme** : Ce sont des rectangles dont les bases sont les classes et les hauteurs sont les fréquences (ou les effectifs)correspondantes.
   (a) On représente les valeurs des bornes des classes sur l'axe des abscisses.
   (b) Dans le cas d'un caractère quantitatif continu, on définit la densité d'effectif d_i d'une classe d'effectif n_i et d'amplitude L_i par : $d_{ni} = n_i/L_i$ (ou, dans le cas des fréquences, $d_{fi} = f_i/L_i$).
   (c) Dans le cas des amplitudes égales (resp. inégale) . Sur l'axe des ordonnées. On représente sur chaque classe $[a_i, a_{i+1}[$ un rectangle de largeur l'amplitude de la classe $L_i = a_{i+1} - a_i$ et de hauteur correspondra aux effectifs (ou fréquences) de la classe $[a_i, a_{i+1}[$. (resp. effectif corrigé ($n_{ic} = d_i.min(L_i)$ ou fréquence corrigée ($f_{ic} = d_{fi}.min(L_i)$)

FIGURE 1.5 – Histogramme des effectifs de l'exemple 1.2.3



On obtient le **polygone des effectifs** (ou des fréquences (voir 1.6)) en reliant les milieux des bases supérieures des rectangles.

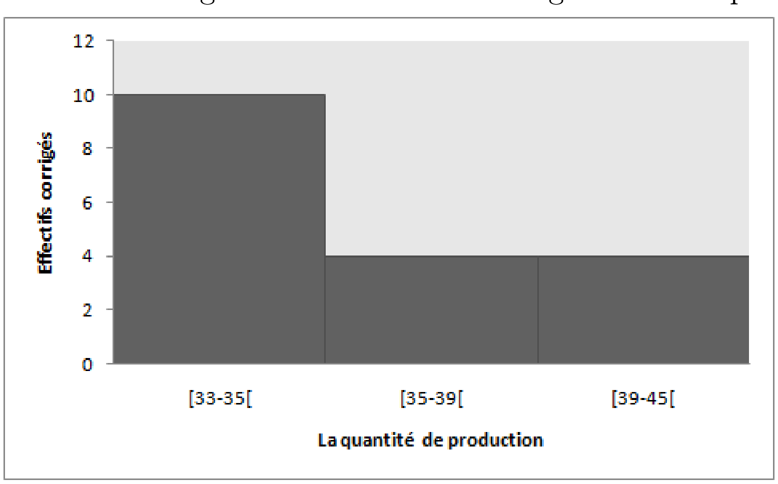
FIGURE 1.6 – Le polygone des effectifs de l'exemple 1.2.3



Exemple 1.2.5. *(Dans le cas des amplitudes inégales)*
Supposons que l'on regroupe les données de l'exemple précédent en classe d'amplitudes inégales. ***Distribution des établissements selon la quantité de production journalière***

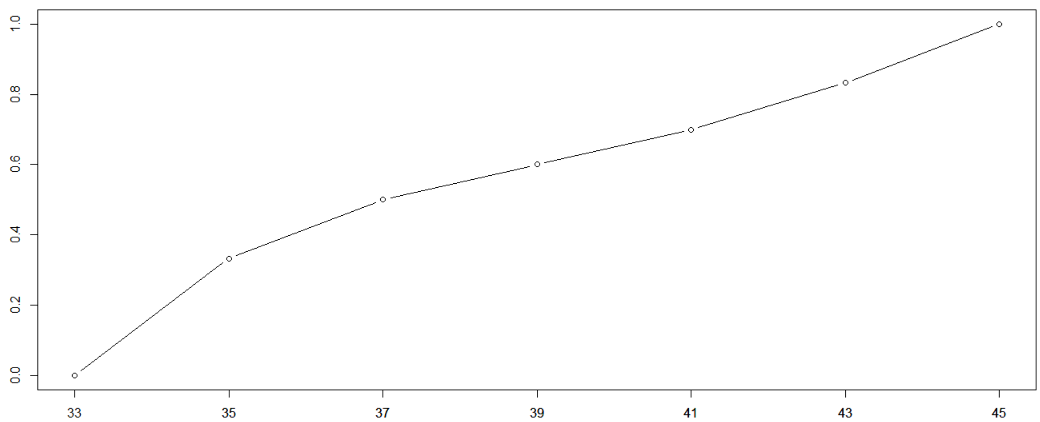
Classes	n_i	*densité* d_{in}	*Effectifs corrigés*
$[33, 35[$	*10*	*5*	*10*
$[35, 39[$	*8*	*2*	*4*
$[39, 45[$	*12*	*2*	*4*
Total Σ	*30*	.	

FIGURE 1.7 – Histogramme des effectifs corrigés de l'exemple 1.2.3



2. La courbe cumulative (ou polygone des fréquences cumulées) : c'est une ligne brisée joignant les points $(a_0, 0), (a_1, F_{1c}), ..., (a_n, 1)$ ou $(a_0, 0), (a_1, N_{1c}), ..., (a_n, 1)$ par des segments de droite.

FIGURE 1.8 – Courbe des fréquence cumulées croissante de l'exemple 1.2.3



Résumé : les graphiques d'une série statistique :

La détermination du graphique dépend à la nature du caractère étudié

Caractère qualitatif	Caractère quantitative discrète
Diagramme en tuyaux d'orgues	Diagramme en bâtons
Diagramme circulaire	polygone des fréquences
	courbe des fréquence cumulées
	courbe des effectifs cumulées

Caractère quantitative continu
Histogramme
polygone des fréquences
courbe des fréquences cumulées
courbe des effectifs cumulées

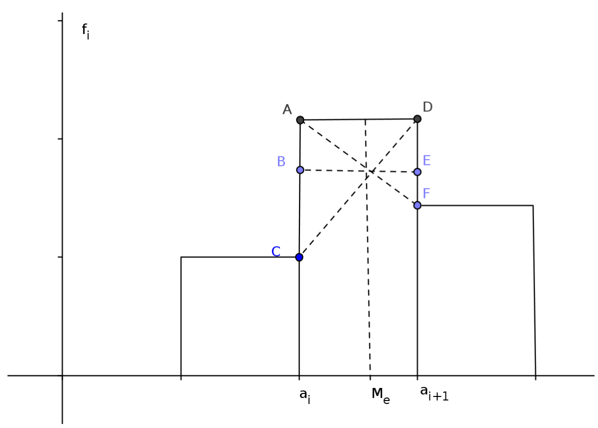
1.3 Paramètres de position (Les caractéristiques de tendance centrale)

Le mode

Le mode noté par M_o associe à la modalité la plus fréquente (resp. fréquence la plus élevée). Pour une variable statistique continue dans lequel les données sont groupées en classes, la classe modale correspond à l'effectif (resp. l'effectif (corrigé)) le plus élevé, dans ce cas le mode est calculé à partir de la formule suivante ; $M_o \in [a_i, a_{i+1}[$:

$$M_o = a_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right).(a_{i+1} - a_i)$$

Δ_1= Différence entre l'effectif de la classe modale et l'effectif de la classe précédente (la classe qui précède la classe modale) .
Δ_2= Différence entre l'effectif de la classe modale et l'effectif de la classe qui suite de la classe modale . Graphiquement, le mode d'une série statistique discrète, est la valeur qui correspond au maximum du diagramme en bâton, et dans le cas continu voire la figure (??)



Remarques 2. *Une série statistique peut admettre uniques ou plusieurs modes. Si la distribution observée possède un seul mode, on dit que la distribution est uni-modale sinon la distribution multimodale ou plurimodale...etc.*

Exemple 1.3.1. *Avec les données de l'exemple (1.2.2)(cas discret) ; le mode est égal à $X = 2$.*
Avec les données de l'exemple (1.2.3) (cas continue) :
la classe modale est donc [33,35[correspond à l'effectif corrigé le plus élevé $n_1 = 10$.

Alors, la valeur du mode est égale à $M_o = 33 + (10/10 + 5).2 \simeq 34,33$ *(unité de la quantité de production)*

Les moyennes

Il existe différentes type de moyennes autres que la moyenne arithmétique. Leur utilité sera vue en exercice.

Moyenne arithmétique :

La moyenne arithmétique est la somme de toutes les données de la série statistique divisées par le nombre des éléments de l'échantillon de taille N. La moyenne arithmétique est notée $\overline{X}$ et donnée par

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Si les données sont présentées dans un tableau statistique dans le quel chaque modalité est associée à fréquence simple ou relative alors on calcule la moyenne arithmétique pondérée par la formule suivante :

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

Propriétés 1.3.1. ♠

- $\forall \alpha, \beta \in \mathbb{R}$ *on a :* $\overline{\beta X + \alpha} = \beta\overline{X} + \alpha$ *et* $\overline{\beta X} = \beta\overline{X}$.
- *Soit* $\overline{X}$ *la moyenne arithmétique de la série statistique* $E = (x_1, n_1), (x_1, n_1), .., (x_k, n_l)$ *et* $\overline{Y}$ *la moyenne arithmétique de la série statistique* $Q = (Y_1, m_1), (Y_1, m_1), .., (Y_1, m_r)$. *La moyenne arithmétique des deux séries statistiques E et Q sont égale*

$$\overline{X}_{(E\cup Q)} = \frac{n\overline{X} + n\overline{Y}}{n+m}, \quad avec \ n = \sum_{i=1}^{k} = n_i, \quad et \quad m = \sum_{i=1}^{r} = m_i$$

Moyenne géométrique

La moyenne géométrique d'une série statistique positive, notée $\overline{X}_g$ est donnée par :

$$\overline{X}_g = \sqrt[N]{\prod_{i=1}^{n} x_i}$$

.

Si les données avec d'effectifs n_i, la moyenne géométrique pondérée est calculée par la formule suivante ;

$$\overline{X}_g = \sqrt[N]{\prod_{i=1}^{n} x_i^{n_i}} = exp\left(\frac{1}{N}\sum_{i=1}^{n} n_i ln x_i\right)$$

Ce type de moyenne peut être utilisée dans le calcul la moyenne de taux d'intérêt.

Exercice 1.3.1. *Un client d'une banque dépose une somme d'argent X pour 5 ans. L'intérêt de la banque est réactualisé chaque année. Ainsi, l'intérêt établi pour les 2 premières années est de τ_1, les 2 années suivantes de τ_2 et la dernière année de τ_3. Déterminer le taux d'intérêt moyenne pour les 5 ans. Soit S_i la somme cumulée à la fin de l'année i. on a*

- *A l'issue de la deuxième année : $S_2 = X(1+\tau_1)^2$.*
- *A l'issue de la quatrième année : $S_4 = X(1+\tau_1)^2(1+\tau_2)^2$.*
- *A l'issue de la cinquième année : $S_5 = X(1+\tau_1)^2(1+\tau_2)^2(1+\tau_3)$.*
- *Le taux moyen sur 5 ans égale $(1+\tau_{moy})$ et son la somme cumulée égale $S_5 = X(1+\tau_{moy})^5 = X(1+\tau_1)^2(1+\tau_2)^2(1+\tau_3) \Rightarrow 1+\tau_{moy} = \left((1+\tau_1)^2(1+\tau_2)^2(1+\tau_3)\right)^{1/5}$.*
  Alors $1+\tau_{moy}$ est la moyenne géométrique des taux d'intérêt

Moyenne Harmonique

La moyenne harmonique d'une série statistique non nulle, notée $\overline{X}_H$ est la moyenne de l'inverse de la variable X, ou l'inverse de la moyenne arithmétique des inverses de la somme de ses valeurs

$$\overline{X}_H = N.\left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1}.$$

Si les données avec d'effectifs n_i la moyenne harmonique est égale à :

$$\overline{X}_H = N.\left(\sum_{i=1}^{n}\frac{n_i}{x_i}\right)^{-1}.$$

Ce type de moyenne peut être utilisée dans le calcul la moyenne de vitesse.

Moyenne quadratique

La moyenne quadratique d'une série statistique X, notée $\overline{X}_q$ est défini par

$$\overline{X}_q = \left(\frac{1}{N}\sum_{i=1}^{k} n_i x_i^2\right)^{1/2} = \left(\sum_{i=1}^{k} f_i x_i^2\right)^{1/2}.$$

Remarques 3. $\star$

- *Dans le cas d'un caractère continu on remplace x_i par le centre c_i de la classe.*
- *Les différents types des moyennes calculées pour un caractère doivent vérifier l'inégalité suivante*

$$x_{min} \le \overline{X}_H \le \overline{X}_g \le \overline{X} \le x_{max}.$$

Médiane

La médiane d'une série statistique X, notée M_e où $x_{(1/2)}$ est défini comme étant la valeur qui partage les individus de la série statistique en deux groupes de même effectifs.

- **(a) Caractère statistique discret :**

  La détermination de la médiane nécessite que les termes de la série soient classés par ordres croissants (ou décroissant).

  Si la valeur de l'effectif total N est impaire, c-à-d : $N = 2k + 1$ alors la médiane M_e est la valeur qui se trouve à l'ordre $k + 1$ (ou $(N + 1)/2$).

$$M_e = x_{k+1} = x_{(N+1)/2}$$

  .

  Si la valeur de l'effectif total N est paire, c-à-d : $N = 2k$, alors la médiane M_e est la valeur qui se trouve à l'ordre k et $k + 1$ (ou $N/2$ et $(N/2) + 1$)

$$M_e = (x_{k+1} + x_k)/2 = (x_{N/2} + x_{(N/2)+1})$$

  .

- La médiane vérifie la relation suivante

$$M_e = min\{x_i,\ \ i \in \{1, 2, ..., k\},\ \ F(x_i) \ge 0.5\}.$$

Exemples 1.3.1. *Avec les données de l'exemple 1.2.3. La médiane est la valeur de la variable qui divise la population de la série statistique en deux parties égales. Nous avons, $N = 42$ est un nombre pair donc $Me = \frac{x_{(21)}+x_{(22)}}{2} = 3$.*

FIGURE 1.9 – La médiane si n est impair

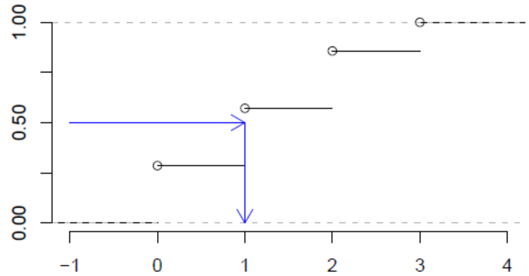
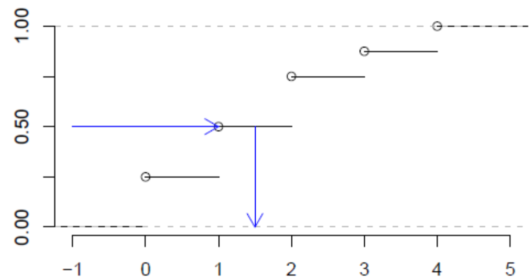


FIGURE 1.10 – La médiane si n est pair



– **(b) Caractère statistique continu**
la médiane est la modalité x qu'elle vérifie :

$$F(x) = 0.5.$$

Pour calculer la médiane on doit déterminer la classe médiane à partir des effectifs (repe. fréquences) cumulés croissants (c'est une classe de rang N/2), puis on calcule la valeur ponctuelle de la médiane selon la formule suivante : si $M_e \in [a_i, a_{i+1}[$, alors :

$$M_e = a_i + \frac{0.5 - F_{-1}}{f}(a_{i+1} - a_i).$$

– F_{-1} : Somme des fréquences correspondant à toutes les classes qui précède la classe médiane .
– f : Fréquence de la classe médiane.

Nous généralisons la notion de la médiane dans la section suivante

Quantiles

Soit $p \in]0, 1]$, les quantiles sont des indicateurs qui divisent la distribution en p parties égales. Si le caractère X est discret, le quantile d'ordre p notée q_p où x_p est l'unique valeur de la série statistique vérifie la relation suivante :

$$q_p = min\{x_i, \;\; i \in \{1, 2, ..., k\}, \;\; F(x_i) \geq p\}.$$

Si $(N.p) \in \mathbb{N}^*$, alors :

$$q_p = \frac{x_{(NP)} + x_{(NP+1)}}{2}.$$

Si $(N.p) \notin \mathbb{N}^*$, alors :

$$q_p = x_{[NP]+1}.$$

où $[NP]$ représente la partie entière de NP.
Si le caractère X est continu : le quantile d'ordre p est la modalité x qu'elle vérifie :

$$F(x) = p.$$

Pour calculer le quantile d'ordre p on doit déterminer la classe de quantile à partir des effectifs (repe. fréquences) cumulés croissants (c'est une classe dont le rang (la position) est le plus petit entier qui suit Np), puis on calcule la valeur ponctuelle du quantile selon la formule suivante : si $q_P \in [a_i, a_{i+1}[$, alors :

$$q_p = a_i + \frac{P - F_{-1}}{f}(a_{i+1} - a_i).$$

F_{-1} : Somme des fréquences correspondant à toutes les classes qui précède la classe $[a_i, a_{i+1}[$

Remarque 1.3.1. ♣

- *Pour* $p = 1/2$, *on parle de deuxième quartile* $Q_2 = M_e$ *est la médiane.*
- *Pour* $p = 1/4$, *on parle de premier quartile* (Q_1)
- *Pour* $p = 3/4$, *on parle de troisième quartile* (Q_3)
- *Pour* $p = k/10$ *(avec* $k \in [[1; 9]]$*), on parle de déciles* (D_k)
- *Pourp* $= k/100$ *(avec* $k \in [[1; 99]]$*), on parle de centiles* (C_k)

1.4 Paramètres de dispersion

Soient X et Y deux ensembles d'observations suivants :

$$X = \{6, 6, 7, 7, \underbrace{8}_{\overline{X} = \, M_e}, 9, 9, 10, 10\} \quad et \quad Y = \{1, 2, 4, 6, \underbrace{8}_{\overline{Y} = \, M_e}, 10, 12, 14, 15\}.$$

Les deux séries statistiques ayant la même moyenne et la même médiane, mais ils sont différents. Le premier ensemble est moins dispersé que le deuxième.

L'étendue

L'étendue d'une série statistique X notée e est défini comme un paramètre qui mesure l'écart entre la valeur maximum et minimum de la distribution statistique :

$$e = x_{max} - x_{min}$$

Interquartile

L'intervalle interquartile est le segment $[Q_1, Q_3]$, cet intervalle contient 50% des observations. L'écart interquartile est la longueur de segment interquartile :

$$IQ = Q_3 - Q_1.$$

Variance et L'écart type

La variance d'une série statistique X notée Var(X) est défini comme la somme des carrés des écarts à la moyenne divisée par l'effectif N. L'écart-type notée σ_X est égal à la racine de la variance.

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

et

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\sum_{i=1}^{n} f_i(x_i - \overline{X})^2}$$

Propriétés 1.4.1. $\star$

- *L'écart type est un paramètre de dispersion le plus utilisé et le plus simple à interpréter. Il donne la variation moyenne de la distribution autour de la moyenne arithmétique.*
- $Var(X) \geq 0$ *(la variance d'une variable est une valeur positive)*
- $\forall \alpha \in \mathbb{R},\ \ Var(\alpha) = 0.$
- $\forall \alpha,\ \beta \in \mathbb{R},\ \ V(\alpha X + \beta) = \alpha^2 Var(X).$
- $Var(X) = \frac{1}{N}\sum_{i=1}^{n} n_i(x_i)^2 - \overline{X}^2 = \sum_{i=1}^{n} f_i(x_i)^2 - \overline{X}^2.$

L'écart moyen absolu

L'écart moyen absolu d'une série statistique X la quantité notée $e_{\overline{X}}$ et définie comme la somme des valeurs absolues des écarts à la moyenne divisée par la taille de population N :

$$e_{\overline{X}} = \frac{1}{N} \sum_{i=1}^{n} n_i |x_i - \overline{X}|.$$

Coefficient de variation

Quand on veut comparer la dispersion de deux séries statistique qui n'ont pas le même ordre de grandeur ou qui portent sur des variables différentes, ont utilisé Le coefficient de variation se définit comme le rapport de l'écart type divisé par la moyenne, exprimé en pourcentage, c-a-d : $CV = \frac{\sigma_X}{\overline{X}}$ avec $\overline{X} \neq 0$.

Remarques 4. $\star$

– *Plus la dispersion est forte plus le coefficient de variation est élevé.*

– *Si* $CV < 15\%$ *indique une bonne homogénéité de la distribution.*

Moments

Soit $r \in \mathbb{N}$. Le moment d'ordre r d'une série statistique X le nombre noté noté M_r est défini par la formule suivante :

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

Le moment centré d'ordre r d'une série statistique X est donné par :

$$M_r^{'} = \frac{1}{N} \sum_{i=1}^{k} n_i \left(x_i - \overline{X}\right)^r = \sum_{i=1}^{k} f_i \left(x_i - \overline{X}\right)^r.$$

Variable centrée réduite

Une variable X est dite centrée (resp. réduite), si elle est de moyenne arithmétique nulle et de variance égale à 1 .

Exemple 1.4.1. *Soient X et Y deux variables, tel que la variable Y est définie par l'expression suivante :*

$$Y = \frac{X - \overline{X}}{\sigma_X}$$

. Démontrer que la Y associé à variable X est centré-réduite ? Il suffit de montrer que $\overline{Y} = 0$ et $Var(Y) = 1$. Grâce à les propriétés (1.3.1) et (1.4.1), on a :

$$\overline{Y} = \frac{\overline{X} - \overline{X}}{\sigma_X} = 0 \quad et \quad Var(Y) = \frac{1}{Var(X)} Var(X)$$

1.5 Paramètres de forme

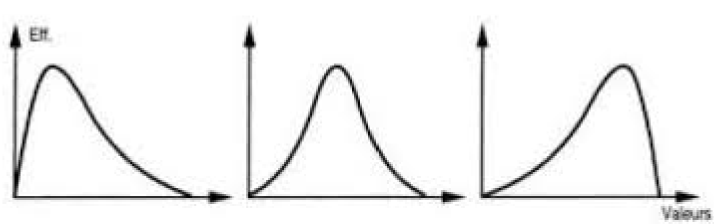
Pour déterminer la forme de la distribution de la variable statistique X, il existe des indicateurs de forme pour évaluer l'asymétrie (skewness en anglais) et l'aplatissement (kurtosis en anglais) d'une distribution statistique.

- **Coefficients d'asymétrie** : coefficients d'asymétrie de Ficher, coefficients d'asymétrie de Yule et coefficients d'asymétrie de pearson.
- **Coefficients d'aplatissement** : coefficients d'aplatissement de pearson et coefficients d'aplatissement de Ficher

L'asymétrie ou dissymétrie

Si la moyenne, la médiane et le mode sont égaux à ($\overline{X} = M_e = M_o$), alors La courbe de la distribution de fréquence ou effectif est symétrique. Si la moyenne, la médiane et le mode sont différents, la distribution est asymétrique à droite ($\overline{X} > M_e > M_o$) ou à gauche ($\overline{X} < M_e < M_o$)

FIGURE 1.11 – A gauche : Asymétrie à droite, Centre : Symétrie, A droite : Asymétrie à gauche



Coefficient d'asymétrie de Fisher :

$$\gamma_1 = \frac{M'_3(X)}{\sigma^3(X)}$$

1. Si $\gamma_1 < 0$, la distribution est asymétrique(étalée) à droite,
2. Si $\gamma_1 = 0$, la distribution est symétrique
3. Si $\gamma_1 > 0$, la distribution est asymétrique (étalée) à gauche.

Coefficient d'asymétrie de Pearson :

$$\beta_1 = \frac{\overline{X} - M_o}{\sigma(X)}$$

1. Si $\beta_1 < 0$, la distribution est asymétrique (étalée) à droite.
2. Si $\beta_1 > 0$, la distribution est asymétrique(étalée) à gauche.
3. Si $\beta_1 = 0$, la distribution est symétrique.

Remarque 1.5.1. *Si la distribution admet plusieurs mode, le coefficient d'asymétrie de Pearson inutilisable.*

Coefficient d'asymétrie de Yule :

$$C_Y = \frac{Q_3 + Q_1 + 2M_e}{Q_3 - Q_1}$$

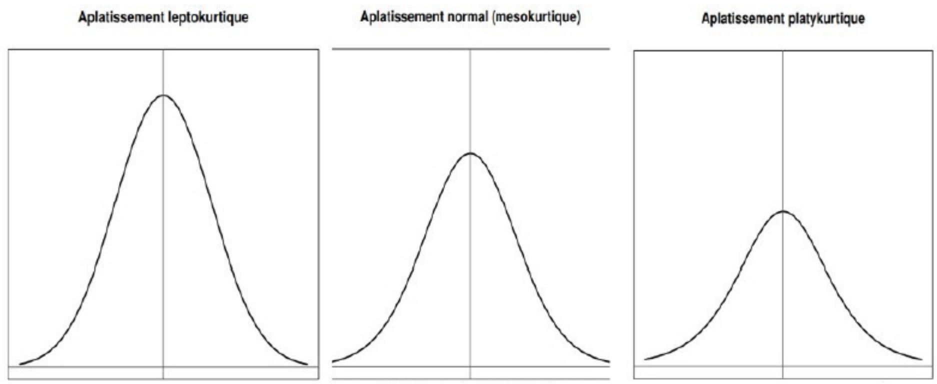
1. Si $C_Y < 0$, la distribution est asymétrique(étalée) à droite.
2. Si $C_Y > 0$, la distribution est asymétrique (étalée) à gauche.
3. Si $C_Y = 0$, la distribution est symétrique.

L'aplatissement

L'aplatissement désigne la forme de la courbe de la distribution (plus Aplati, moins aplati ou normal) ou bien signifie la comparaison de sommet de la courbe par rapport au sommet de la courbe de distribution normale

- Une distribution est platicurtique (hyponormale) si la courbe est moins aplatie que la courbe normale
- Une distribution est leptocurtique (hyperonormale) si la courbe est plus aplatie que la courbe normale.

– Une distribution est mesokurtique (normale) si la courbe est normale.



Remarque 1.5.2. ▶
Paramètres de forme sont principalement utilisés pour vérifier la distribution Proche de la distribution normale (loi normale) ; en effet, pour une telle loi, le coefficient d'aplatissement est égal 3 et le coefficient d'asymétrie égal 0.

Coefficient d'aplatissement de Pearson

$$\beta_2 = \frac{M_4'(X)}{\sigma^4(X)}.$$

Coefficient d'aplatissement de Fisher ou coefficient de Yule :

$$\gamma_2 = \beta_2 - 3.$$

– Si $\gamma_2 = 0$, une distribution est dite mésokurtique.
– Si $\gamma_2 > 0$,une distribution est dite leptokurtique.
– Si $\gamma_2 < 0$, une distribution est dite platykurtique.

UNIVERSITE Abdelhamid Ibn Badis MOSTAGANEM

Département de Mathématique et de l'Informatique

L1 MI — Stat -Proba — Février 2022

TD N1-Séries statistiques à une variable (simples

Rappels sur les Opérations de Base

Exercice 1

Soit : $a_1, a_2, ..., a_n$ des nombres réels :

- Développer les expressions suivantes :

$$\sum_{i=1}^{4} a_i, \quad \sum_{i=1, i\neq 4}^{6} a_i, \quad \sum_{i=1, i\neq 2, i\, pair}^{6} a_i, \quad \prod_{i=1}^{4} a_i, \quad \prod_{i=1}^{n} a_1, \quad \prod_{i=1, i\, impair}^{5} a_i.$$

$$\sum_{i=1}^{n} a_1, \quad \left(\sum_{i=1}^{n} a_i\right)^2 \quad \sum_{i=1}^{n} (\alpha a_i + \beta a_{i+1})^2$$

Vocabulaire de la statistique descriptive

Exercice 2

1. Pour chacune des variables suivantes préciser si elle est :Qualitative nominale, Qualitative ordinale, Quantitative discrète, Quantitative continue

   (a) Citoyenneté.
   (b) Taille.
   (c) Nombre de maison vendues par ville.
   (d) Niveau de la scolarité.
   (e) Chiffre d'affaire d'une entreprise.
   (f) Nombre de langues parlées.
   (g) Couleur des yeux.
   (h) Lieu de résidence.
   (l) Poids

2. Proposer les modalités pour chaque variable quantitative transformée en variable qualitative :

   Hauteur, Poids, Chiffre d'affaire.

3. Pour les sujets d'étude qui suivent, spécifier : Population statistique- Unité statistique-Variable statistique-Type de variable Statistique
   1. Les salaires annuels des enseignants d'une école.
   2. Le sport pratiqué par chacun des élèves d'une classe.
   3. La taille des joueurs de l'équipe nationale de handball
   4. La durée de vie des tubes de télévisions fabriquées par une usine.

Représentations graphiques-Distributions statistiques

Exercice 3
Le tableau suivant donne la répartition selon le groupe sanguin de 40 individus pris au hasard dans une population,

Groupe sanguins	A	B	AB	O
L'effectif	20	10	n	5

1. Déterminer la variable statistique et son type.
2. Déterminer l'effectif des personnes ayant un groupe sanguin B.
3. Donner toutes les représentations graphiques possibles de cette distribution.

Exercice 4
La liste suivante est composée des personnes d'un groupe d'étudiants, suivis entre parenthèses du nombre de films que chacun d'entre eux a vus au cours du mois dernier : Mohamed (3), anes (2), riad (2), karim (3), Abdelah (1), oussama (2), rahmani(0), Farida (2), ibtihel (2), khanssa (0).

1. Déterminez : la population étudiée, la variable étudiée, la nature de la variable et les modalités de la variable
2. Construisez le tableau statistique en commençant par ordonner de manière croissante la série, puis en cherchant les effectifs cumulés (croissants, décroissants) et les fréquences cumulées (croissantes, décroissantes).
3. Tracer le diagramme en bâtons et le polygone des fréquences

Exercice 5 Le tableau suivant représente le nombre d'absences de 233 étudiants

x_i	0	1	2	3	4
n_i	90	45	60	20	18

1. Représenter le tableau statistique où figure les effectifs et les fréquences cumulées croissantes et décroissante.
2. Quelle est la proportion des étudiants qui ont 3 absences au minimum ?
3. . Que représente la valeur : $1 - (f_1 + f_2)$?
4. Tracer la courbe cumulatives des fréquences croissantes

Exercice 6
L'étude du poids de 50 personnes a donné les résultats suivants en kg représentés dans une série statistique ordonnée :
37 43 47 50 52 54 55 56 58 58 61 62 63 64 65 66 66 67 68 69 69 69 70 71 72 72 72 73 73 74 74 75 76 76 77 77 79 79 80 82 82 84 86 87 88 90 92 93 97 97.

1. Calculer l'étendue de la population
2. Regrouper les données en des classes selon la règle de Sturge
2. Représenter le tableau statistique où figure les effectifs et les fréquences cumulées croissantes et décroissantes.
3. Donner la proportion en pourcentage des personnes qui ont un poids X supérieur ou égale à 80 kg ?
4. Tracer l'histogramme des effectifs et le polygone des effectifs.
5. Tracer les courbes cumulatives des effectifs croissants et décroissants.

UNIVERSITE
Abdelhamid Ibn Badis
MOSTAGANEM

Département de Mathématique et de l'Informatique

L1 MI | Stat -Proba | Novembre 2021

TD N1-Statistique descriptive à une variable

Rappels sur les Opérations de Base

Solution de l'exercice 01 :
Soit : $a_1, a_2, ..., a_n$ des nombres réels :

1- Développer les expressions suivantes :

- $\sum_{i=1}^{4} a_i = a_1 + a_2 + a_3 + a_4$
- $\sum_{i=1, i\neq 4}^{6} a_i = a_1 + a_2 + a_3 + a_5 + a_6$
- $\sum_{i=1, i\neq 2, i\, pair}^{6} a_i = a_4 + a_6$
- $\prod_{i=1}^{4} a_i = a_1 \times a_2 \times a_3 \times a_4$
- $\prod_{i=1}^{n} a_1 = a_1 \underbrace{\times a_2 \times ... \times a_n}_{n fois} = a_1^n$
- $\prod_{i=1, i\, impair}^{5} a_i = a_1 \times a_3 \times a_5$
- $\sum_{i=1}^{n} a_1 = \underbrace{a_1 + a_2 + ... + a_n}_{n fois} = na_1$
- $\left(\sum_{i=1}^{n} a_i\right)^2 = \sum_{i=1}^{n} a_i^2 + 2\sum_{i=1}^{n}\sum_{j=1, i\neq j}^{n} a_i a_j$
- $\sum_{i=1}^{n} (\alpha a_i + \beta a_{i+1})^2 =$

. $\alpha^2 a_1^2 + (\alpha^2 + \beta^2)\sum_{i=2}^{n} a_i^2 + \beta^2 a_{n+1}^2 + 2\alpha\beta \sum_{i=1}^{n} a_i a_{i+1}$

Vocabulaire de la statistique descriptive

Exercice 2

- **Caractère qualitatif nominale** : (a)-(g)-(h)
- **Caractère qualitatif ordinale** : (d)
- **Caractère quantitatif discret** (c)-(e)-(f)
- **Caractère quantitatif continu** : (b)-(i)

2/ Les modalités de la variable statistique quantitative **"Hauteur"** transformée en variable qualitative : Petit-Moyen-Grand.

- Les modalités de la variable statistique quantitative **"Pois"** transformée en variable qualitative : Très léger- Léger-Moyen-Lourd-Très lourd.
- Les modalités de la variable statistique quantitative **"Chiffre d'affaire"** transformée en variable qualitative : Modéré-Moyen-Important-Très important

Sujet de l'étude	Population.S	Unité.S	Variable.S	Type de V.S
1	Les enseignants d'une école	Un enseignant	Les salaires annuels	quantitative continue
2	Les élèves d'une classe	Un élève	le sport pratiqué	qualitative nominale
3	Les joueurs de l'équipe nationale de handball	Un joueur	la taille de joueurs	quantitative continue
4	Les télévisions fabriquées par une usine	Un tube de TV	la durées de vie des tubes des TV	quantitative continue

Solution de l'exercice 03 :

1- La population dans cette étude est les 40 personnes. Donc $N = 40$. La variable statistique est le groupe sanguin des individus et elle est qualitative.

2- L'effectif total est égal à 40. Par conséquent, $N = \sum_{i=1}^{4} n_i$. Alors $40 = 20 + 10 + n_3 + 5$. Ce qui implique que $n_3 = 5$.

3- Nous avons deux représentations possibles "Tyaux d'orgue" et "Diagramme en secteur".

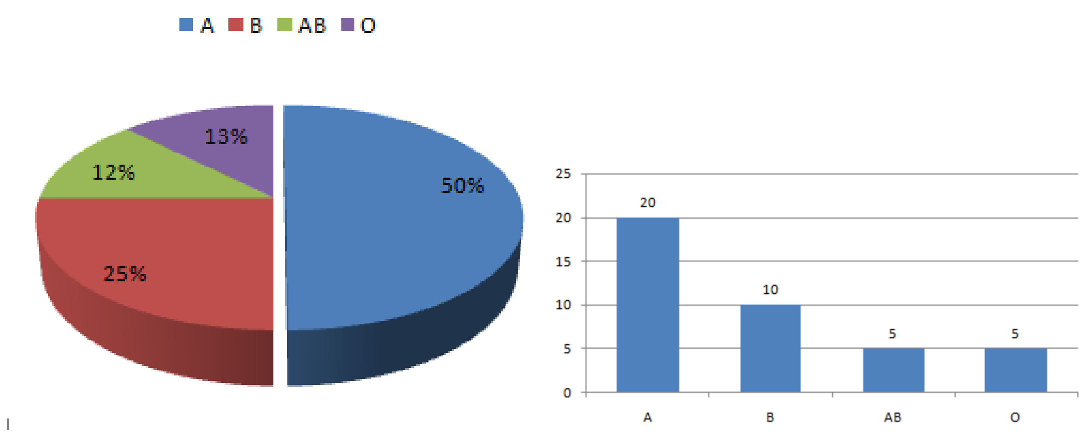


Figure 1 – A gauche "Diagramme en secteur" et à droite "Tyaux d'orgue"

Solution de l'exercice 4 :

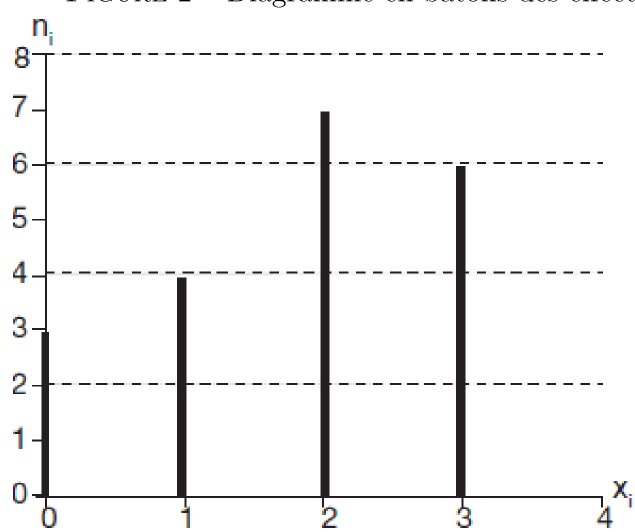
– La population étudiée est le groupe d'étudiants.

- La variable étudiée est X = " nombre de films que chacun d'entre eux a vus au cours du mois dernier ".
- La variable étudiée est quantitative discrète
- L'ensemble X des modalités est $X = \{0, 1, 2, 3\}$

Tableau statistique

x_i	n_i	$N_{ic}^{\nearrow}$	$N_{ic}^{\searrow}$	f_i	$f_{ic}^{\nearrow}$	$f_{ic}^{\searrow}$
0	2	2	10	1/5	1/5	1
1	1	3	8	1/10	3/10	4/5
2	5	8	7	1/2	4/5	7/10
3	2	10	2	1/5	1	1/5
Total Σ	10			1		

FIGURE 2 – Diagramme en bâtons des effectifs



<u>Solution de l'exercice 5</u> :
Tableau statistique

x_i	n_i	$N_{ic}^{\nearrow}$	$N_{ic}^{\searrow}$	f_i	$f_{ic}^{\nearrow}$	$f_{ic}^{\searrow}$
0	90	90	233	90/233	90/233	1
1	45	135	143	45/233	135/233	143/233
2	60	195	98	60/233	195/233	98/233
3	20	215	38	20/233	215/233	38/233
4	18	233	18	18/233	1	18/233
Total Σ	233			1		

1/ La proportion des étudiants qui ont 3 absences au minimum égale :

$$f_{4c}^{\searrow} = (f_4 + f_5) = (0.085 + 0.075) = 0.162 = 16.2\%$$

2/ $1-(f_1+f_2) = f_3+f_4+f_5 = f_{3c}^{\searrow} = (98/233)\times 100\% = 42,1\%$, c'est la proportion des étudiants qui ont 2 absence au minimum (plus d'une absence)

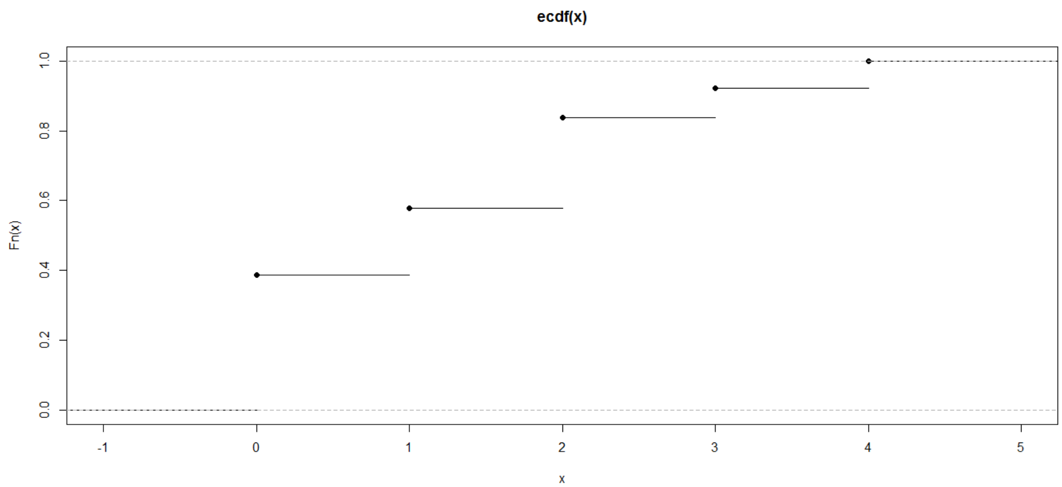


FIGURE 3 – Courbe cumulatives des fréquences croissantes

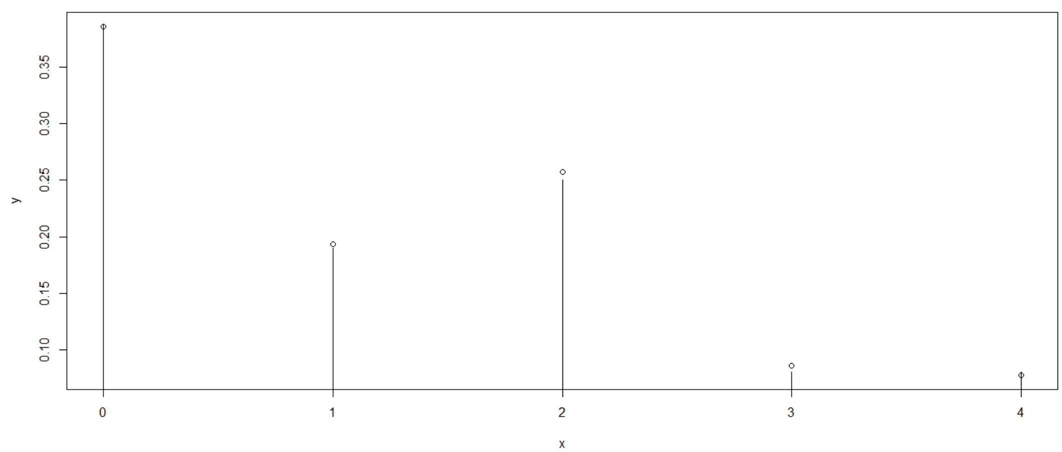


FIGURE 4 – Diagramme en bâtons

<u>**Solution de l'exercice 6**</u> :

1 On sait que l'étendu est égale au maximum moins le minimum. Ainsi

$$e = x_{max} - x_{min} = 97 - 37 = 60.$$

2 Par la méthode de Yule, nous avons

$$k = 2.5\sqrt[4]{N} = 2.5\sqrt[4]{50} \simeq 6.68 \simeq 7.$$

3 Par la méthode de Sturge, nous avons

$$k = 1 + 3.3\,log_{10}(N) = 1 + 3.3\,log_{10}(50) \simeq 6.6 \simeq 7.$$

Nous avons donc l'amplitude qui égale

$$L_i = \frac{e}{k} \simeq 8.7$$

4 Tableau statistique

Classe	n_i	$N_{ic}^{\nearrow}$	$N_{ic}^{\searrow}$	f_i	$f_{ic}^{\nearrow}$	$f_{ic}^{\searrow}$
$[37 - 45.6[$	2	2	50	2/50	2/50	1
$[45.6 - 54.2[$	4	6	48	4/50	6/50	48/50
$[54.2 - 62.8[$	6	12	44	6/50	12/50	44/50
$[62.8 - 71.4[$	12	24	38	12/50	24/50	38/50
$[71.4 - 80[$	15	39	26	15/50	39/50	26/50
$[80 - 88.6[$	6	45	11	6/50	45/50	11/50
$[88.6 - 97.2]$	5	50	5	5/50	1	5/50
Total	50			1		

5 La proportion en pourcentage des personnes qui ont un poids X supérieure ou égale à 80 kg.
On a :

$$(X \succeq 80) = (X \in [80 - 88.6[) \cup (X \in [88.6 - 97.2[)$$

On note n_1 (resp. n_2) effectif des personnes qui ont un poids $X \in [80 - 88.6[$ (resp. $X \in [88.6 - 97.2[$), d'où la proportion en pourcentage des personnes qui ont un poids $X \geq 80$ est égale

$$((n_1 + n_2)/N).100\% = 22\%,$$

<u>2-méthode</u>

La proportion en pourcentage des personnes qui ont un poids$X \geq 80$ est égale $f_{6c}^{\searrow} \times 100\%$. D'où :

$$f_{6c}^{\searrow} \times 100\% = (f_6 + f_7) \times 100\% = 22\%.$$

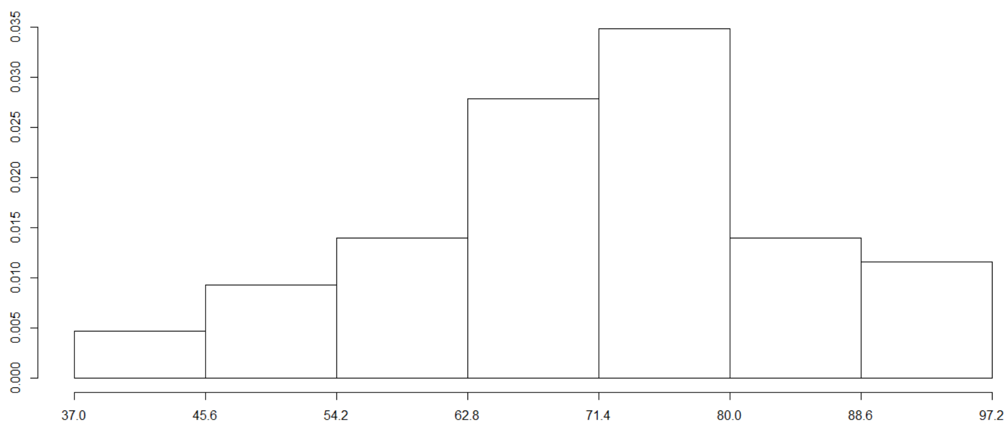


FIGURE 5 – Histogramme des effectifs

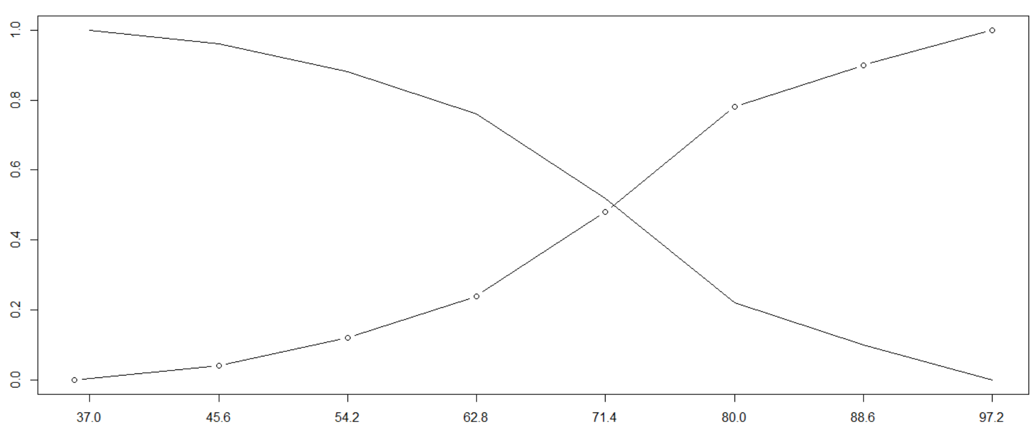


FIGURE 6 – Courbe cumulatives des effectifs croissants et décrissantes

UNIVERSITE
Abdelhamid Ibn Badis
MOSTAGANEM

Département de Mathématique et de l'Informatique

L1 MI | Stat -Proba | Mars 2022

TD N2-Séries statistiques à une variable (simples

Paramètres de position-Paramètres de dispersion

Exercice 1

Le gérant d'un magasin vendant des articles de consommation courante a relevé pour un article particulier qui semble connaître une très forte popularité, le nombre d'articles vendus par jour. Son relevé a porté sur les ventes des mois de Mars et Avril, ce qui correspond à 52 jours de vente. Le relevé des observations se présente comme suit :
7; 13; 8; 10; 9; 12; 10; 8; 9; 10; 6; 14; 7; 15; 9; 11; 12; 11; 12; 5; 14; 11; 8; 10; 14; 12; 8; 5; 7; 13; 12 ; 16; 11; 9; 11; ; 11; 12; 12; 15; 14; 5; 14; 9; 9; 14; 13; 11; 10; 11; 12; 9; 15.

1. Quel est le type de la variable statistique étudiée.
2. Déterminer le tableau statistique en fonction des effectifs, des fréquences, des effectifs cumulés et des fréquences cumulés.
3. Tracer le diagramme des bâtonnés associé à la variable X.
4. Soit F_X la fonction de répartition. Déterminer F_X.
5. Calculer le mode M_o et la moyenne arithmétique $\bar{X}$.
6. Déterminer à partir du tableau puis à partir du graphe, la valeur de la médiane M_e.
7. Calculer la variance et l'écart-type.
8. Déterminer les quartiles et l'intervalle interquartiles.

Exercice 2

Chez un fabriquant de tubes de plastique, on a prélevé un échantillon de 100 tubes dont on a mesuré le diamètre en décimètre. 1.94- 2.20 -2.33- 2.39- 2.45- 2.50- 2.54 -2.61- 2.66- 2.85 1.96 2.21 2.33 2.40 2.46 2.51 2.54 2.62 2.68 2.87 2.07 2.26 2.34 2.40 2.47 2.52 2.55 2.62 2.68 2.90 2.09 2.26 2.34 2.40 2.47 2.52 2.55 2.62 2.68 2.91 2.09 2.28 2.35 2.40 2.48 2.52 2.56 2.62 2.71 2.94 2.12 2.29 2.36 2.41 2.49 2.52 2.56 2.63 2.73 2.95 2.13 2.30 2.37 2.42 2.49 2.53 2.57 2.63 2.75 2.99 2.14 2.31 2.38 2.42 2.49 2.53 2.57 2.65 2.76 2.99 2.19 2.31 2.38 2.42 2.49 2.53 2.59 2.66 2.77 3.09 2.19 2.31 2.38 2.42 2.50 2.54 2.59 2.66 2.78 3.12

1. Déterminez : la population étudiée, la variable étudiée, la nature de la variable et les modalités de la variable.

2. En utilisant la méthode de Yule puis de Sturge, établir le tableau statistique (Faites débuter la première classe par la valeur 1.94).
3. Déterminer par le calcul la valeur du diamètre au-dessous de laquelle se trouvent 50% des tubes de plastique. Que représente cette valeur.
4. Tracer l'histogramme et la courbe des effectifs cumulés croissants, en déduire le mode et la médiane
5. Déterminer par le calcul le pourcentage de tubes ayant un diamètre inférieur à 2.58.
6. Calculer les moyennes arithmétique, géométrique, harmonique et quadratique
7. Calculer le mode, les quartiles et l'écart type

Exercice 3

On considère deux groupes d'étudiants. Nous relevons leurs notes d'examens dans les deux tableaux suivants

Note (groupe A)	8	9	10	11
Effectif	2	2	1	1

Note (groupe B)	6	8	9	13	14
Effectif	2	2	2	1	1

- Calculer la moyenne, l'écart type et coefficient de variation de chaque groupe. Comparer les deux groupes.

Exercice 4

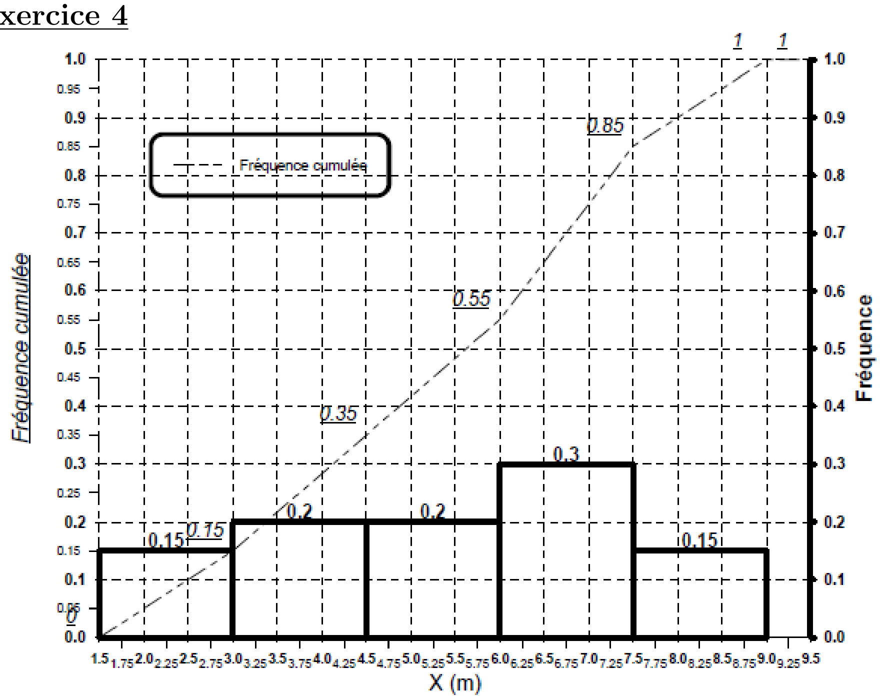


Figure 1

La Figure 1 donne l'histogramme ainsi que la fonction cumulative de la variable statistique X (m).

1- Déterminer graphiquement le mode
2- Déterminer graphiquement la médiane
3- Déterminer la moyenne

<u>**Exercice 5**</u> (**Paramètres de forme**)
Calculer les coefficients d'asymétrie de Fisher et d'aplatissement de Pearson de la série statistique suivante

x_i	1	2	3	4	5	6
n_i	3	4	7	6	3	2

<u>Devoir de maison</u>

Etant donné la série statistique (poids en grammes de 30 pommes) :
42.50- 45.00- 47.75-52.10- 29.50- 31.25- 21.50- 56.30- 55.60- 49.80- 35.55- 42.30- 43.50- 34.60- 65.50- 45.10- 40.25- 58.00- 30.30- 44.80- 36.50- 55.00- 59.20- 36.50- 38.50- 41.10- 46.50- 39.95- 25.35- 49.50.
1/ Effectuer un rangement ascendant de ces données.
2/ Quelle est l'étendue des valeurs.
3/ Construire selon le poids des pommes une distribution de fréquences comportant les classes suivantes : " 20g à moins de 30g ", " 30g à moins de 40g ", " 40g à moins de 50g ", " 50g à moins de 60g " et " 60g à moins de 70g " (faire un tableau des effectifs).
4/ Aurait-il été possible d'utiliser 6 ou 7 classes au lieu de 5 dans la distribution de fréquences précédente ?
5/ Quelle largeur de classe aurait été appropriée pour construire une distribution de fréquence à 8 classes ?
6/ Tracer l'histogramme de la distribution de fréquence établie en 3/.
7/ Tracer le polygone de fréquence de la distribution de fréquence établie en 3/.
8/ Calculer la moyenne arithmétique et la médiane dans le cas 1/ puis dans le cas 3/.
9/ Déterminer les quartiles Q_1 et Q_3 dans le cas 1/ puis dans le cas 3/.
10/ Dans chacun des cas 1/ et 3/ calculer la proportion d'individus dont les valeurs du caractère appartiennent à l'intervalle interquartile.
11/ Dans chacun des cas 1/ et 3/ calculer l'écart type ainsi que la proportion d'individus dont les valeurs du caractère appartiennent à l'intervalle .
12/ Dans chacun des cas 1/ et 3/ calculer le coefficient de variation V.

UNIVERSITE
Abdelhamid Ibn Badis
MOSTAGANEM

Département de Mathématique et de l'Informatique

L1 MI — Stat -Proba — Novembre 2021

TD N1-Statistique descriptive à une variable

Paramètres de position-Paramètres de dispersion

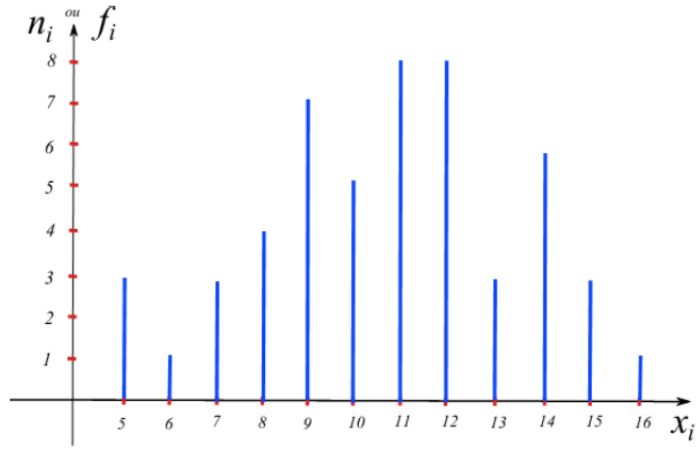
Solution de l'exercice 01 :

- La population est : N = 52 jours
- Le variable statistique : le nombre d'article vendus par jour.
- Le type de V.S : Variable statistique quantitatif discrète.

2/ **Le tableau statistique**

x_i	n_i	f_i	$N_{ic}^{\nearrow}$	$f_{ic}^{\nearrow}$
5	3	$3/52$	3	$3/52$
6	1	$1/52$	4	$4/52$
7	3	$3/52$	7	$7/52$
8	4	$4/52$	11	$11/52$
9	7	$7/52$	18	$18/52$
10	5	$5/52$	23	$23/52$
11	8	$8/52$	31	$31/52$
12	8	$8/52$	39	$39/52$
13	3	$3/52$	42	$42/52$
14	6	$6/52$	48	$48/52$
15	3	$3/52$	51	$51/52$
16	1	$1/52$	52	1
Total Σ	52			1

3. Le diagramme à bâtons :



Le diagramme à bâtons

4. La fonction de répartition :

$$F_x(x) = \begin{cases} 0 & si \quad x < 5 \\ \frac{3}{52} & si \quad 5 \leq x < 6 \\ \frac{4}{52} & si \quad 6 \leq x < 7 \\ \frac{7}{52} & si \quad 7 \leq x < 8 \\ \dots & \dots \quad \dots \\ 1 & si \quad x \geq 7 \end{cases}$$

4- **Le Mode** : est la valeur de la variable qui a le plus grand effectif, c'est à dire, $n_i = 8$ avec $i \in (7, 8)$.
Donc

$$M_o = 11 \text{ et } M_o = 12$$

5- **La moyenne Arithmétique :**

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

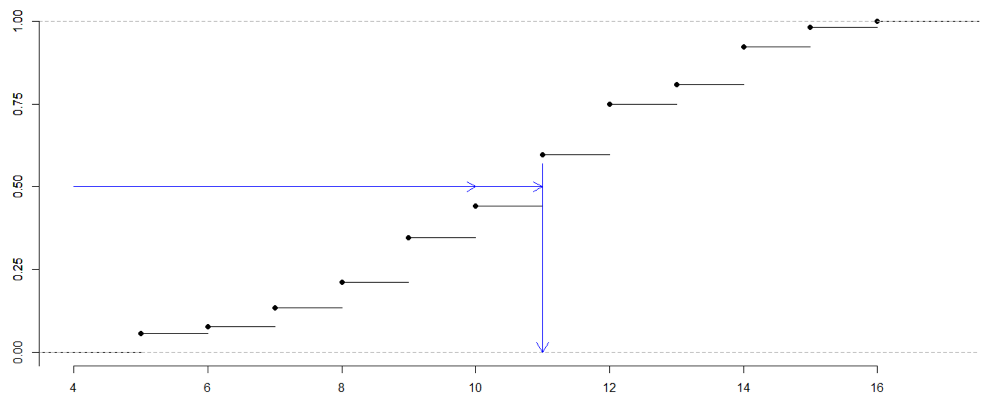
$$\overline{X} = \frac{1}{52} \left((3 \times 5) + (1 \times 6) \times .. \times (1 \times 16)\right) = 10.67$$

6- **La médiane** : est la valeur de la variable qui divise la population de la série statistique en deux parties égales. Nous avons

$$M_e = min \{x_i, i \in \{1, 2, .., 12\}, F(x_i) \geq 0.5\} = min \left\{x_i, i \in \{1, 2, .., 12\}, f_{ic}^{\nearrow} \geq 0.5\right\}$$

d'après le tableau statistique ci-dessus (ou par la formule de la fonction de répartition) on a $M_e = 11$.
2-méthode :
La médiane : Comme $N = 52$ est un nombre pair, donc

$$M_e = (x_{(26)} + x_{(27)})/2 = 11.$$



7- **La variance,**

$$Var(X) = \overline{X^2} - \overline{X}^2 = (1/N)\sum_{i=1}^{k} n_i x_i^2 - \overline{X}^2 = \sum_{i=1}^{k} f_i x_i^2 - \overline{X}^2$$

D'où :

$$Var(X) = 52^{-1}((5^2 \times 3) + (6^2 \times 1) + ... + 1(6^1 \times 1)) - 10.67^2 = 7.64$$

. **L'écart-type est la racine carrée de la variance**

$$\sigma_X = \sqrt{Var(X)} = \sqrt{7.64} \simeq 2.76$$

1. **Les quartiles : ($p = 1/4$)**
   Q_1, le premier quartile : Comme $N \times p = 0.25 \times 52 = 13$ est un nombre entier, on a
   $$Q_1 = x_{(1/4)} = (x_{(13)} + x_{(14)})/2 = 9$$
2. $Q_2 = M_e$.
3. **Q_3, le troisième quartile** : Comme $N \times p = 0.75 \times 52 = 39$ est un nombre entier, on a
   $$Q_3 = x_{(3/4)} = (x_{(93)} + x_{(40)})/2 = 12.5$$

4. **Intervalle interquartiles.** : $IQ = [9 - 12.5]$

Solution de l'exercice 02 :

- Population : les tubes
- Individus : le tube.
- Caractère : le diamètre.
- Type : quantitative continue.
- Modalités : 1.94,..., 3.12.

1 Par la méthode de Yule, nous avons

$$k = 2.5\sqrt[4]{N} = 2.5\sqrt[4]{100} \simeq 7.9 \simeq 8.$$

2 Par la méthode de Sturge, nous avons

$$k = 1 + 3.3\,log_{10}(N) = 1 + 3.3\,log_{10}(100) \simeq 7.6 \simeq 8.$$

Nous avons donc l'amplitude qui égale

$$L_i = \frac{x_{max} - x_{min}}{k} \simeq 0.15$$

Nous obtenons le tableau statistique suivant,

X	n_i	f_i	N_i	F_i
$[1.94, 2.09[$	3	0.03	3	0.03
$[2.09, 2.24[$	9	0.09	12	0.012
$[2.24, 2.39[$	18	0.18	30	0.3
$[2.39, 2.54[$	29	0.29	59	0.59
$[2.54, 2.69[$	25	0.25	84	0.84
$[2.69, 2.84[$	6	0.06	90	0.90
$[2.84, 2.99[$	6	0.06	96	0.96
$[2.99, 3.14[$	4	0.04	100	1
$\sum$	100	1	\	\

4/ On détermine la classe correspond à la fréquence cumulée croissante supérieure ou égale à 50%. D'où d'après le tableau statistique ci-dessus, nous avons $[2.39, 2.54[$.

En utilisant directement la fonction de répartition donnée par

$$F_X(x) = \frac{f_{i+1}}{L}(x - a_i) + F_i$$

Nous mettons

$$F(x) = 0.5 \text{ et } x = \alpha$$

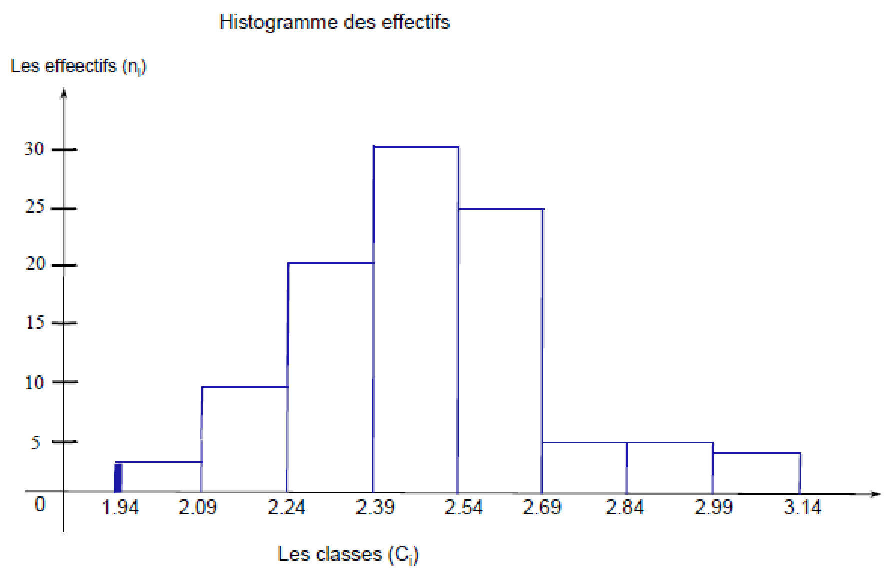
.
Nous retrouvons donc

$$0.5 = \frac{0.29}{0.15}(\alpha - 2.39) + 0.3$$

Alors :

$$\alpha = 2.5.$$

Cette valeur représente la médiane



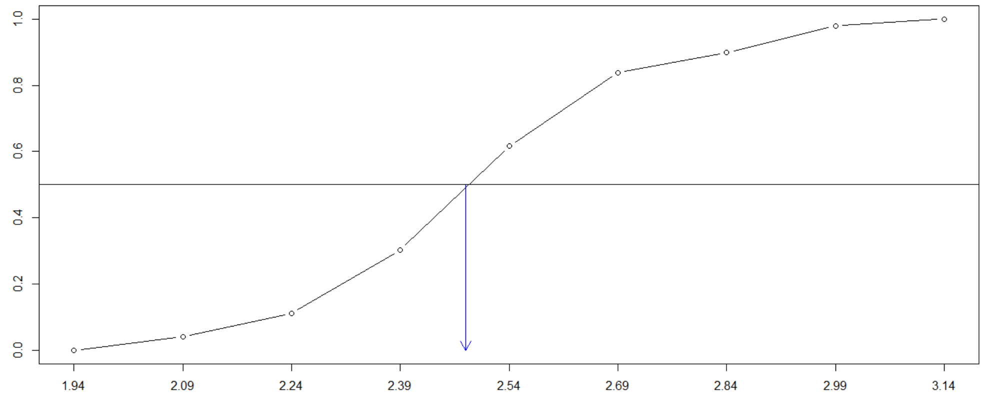


FIGURE 1 – La courbe cumulative des fréquances croissantes

Le calcul du pourcentage de tubes ayant un diamètre inférieur à 2.58

Puisque $2.58 \in [2.54, 2.69[$. Alors par la formule de la fonction de répartition, nous retrouvons donc

$$F_X(2.58) = \frac{0.25}{0.15}(2.58 - 2.39) + 0.59 = 0.66(66\%)$$

.

– **Moyenne arithmétique**

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

– **Moyenne géométrique**

$$\overline{X}_g = \sqrt[N]{\prod_{i=1}^{n} c_i^{n_i}} = exp\left(\frac{1}{N}\sum_{i=1}^{n} n_i ln c_i\right) = ...$$

– **Moyenne Harmonique**

$$\overline{X}_H = N.\left(\sum_{i=1}^{n} \frac{n_i}{c_i}\right)^{-1} = ...$$

– **Moyenne quadratique**

$$\overline{X}_q = \left(\frac{1}{N}\sum_{i=1}^{k} n_i c_i^2\right)^{1/2} = \left(\sum_{i=1}^{k} f_i c_i^2\right)^{1/2} = ...$$

– c_i ; Le centre de la calasse $[a_i, a_{i+1}[$.

Le mode

- La classe modale correspond à l'effectif le plus élevé $n_4 = 29$. La classe modale est donc $[2.39, 2.54[$, alors,

$$M_o = a_i + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right).(a_{i+1} - a_i)$$

  Δ_1 : Différence entre l'effectif de la classe modale et l'effectif de la classe précédente(la classe qui précède la classe modale) .
  Δ_2 : Différence entre l'effectif de la classe modale et l'effectif de la classe qui suit (la classe qui suit la classe modale).
  la valeur du mode est égale à :

$$M_o = 2.39 + \left(\frac{29 - 18}{(29 - 18) + (29 - 25)}\right) \times 0.15 \simeq 2.49.$$

  Les quartiles

- Pour calculer le quantile d'ordre p on doit déterminer la classe de quantile à partir des effectifs (repe. fréquences) cumulés croissants (c'est une classe de rang Np), puis on calcule la valeur ponctuelle du quantile selon la formule suivante : si $q_P \in [a_i, a_{i+1}[$, alors :

$$q_p = a_i + \frac{P - F_{-1}}{f}(a_{i+1} - a_i).$$

  F_{-1} : Somme des fréquences correspondant à toutes les classes qui précède la classe $[a_i, a_{i+1}[$
- **Les quartiles** : Ce sont les quantiles (d'ordre respectivement 0.25, 0.5, et 0.75 et que sont notés : Q_1, Q_2 et Q_3.

- Q_1
  La classe de premier quartile correspondant à les effectifs (repe. fréquences) cumulés croissants égale N/4 (c'est une classe de rang $N \times 1/4$). D'où cette classe égale $[2.24, 2.39[$

$$Q_1 = a_i + \frac{NP - F_{-1}}{f}(a_{i+1} - a_i) = 2.24 + \frac{0.25 - 0.12}{0.18} \times 0.15$$

  . Alors

$$Q_1 \simeq 2.34$$

- Q_2
  La classe médiane est égale $[2.39, 2.54[$

$$Q_2 = 2.39 + \frac{0.50 - 0.3}{0.29} \times 0.15$$

  .
  Alors

$$Q_2 \simeq 2.49$$

– $\underline{Q_3}$
La classe de troisième quartile correspondant à les effectifs (repe. fréquences) cumulés croissants égale 3N/4 (c'est une classe de rang $N \times 3/4$). D'où cette classe égale $[2.54, 2.69[$

$$Q_3 = 2.54 + \frac{0.75 - 0.59}{0.25} \times 0.15$$

.
Alors

$$Q_3 \simeq 2.63$$

– L'écart-type est la racine carrée de la variance . Nous commençons par la variance,

$$Var(X) = \overline{X^2} - \overline{X}^2 = (1/N)\sum_{i=1}^{k} n_i c_i^2 - \overline{X}^2 = \sum_{i=1}^{k} f_i c_i^2 - \overline{X}^2$$

. Alors, $Var(X) =$ .... et $\sigma_X = \sqrt{Var(X)}$............

<u>Solution de l'exercice 03</u> :

Dans un premier temps, nous remarquons que l'effectif total du groupe A est égal à 6 et celui du groupe B est égal à 8.
En utilisant la formule de la moyenne, nous obtenons

$$\overline{X}_A = 9.2 \text{ et } \overline{X}_B = 9.1.$$

On remarque que les moyennes sont très proches. Peut-on pour autant conclure que ces deux groupes ont des niveaux identiques ?
Nous répondons à cette question après le calcule des écarts type et coefficient de variation. Ils sont donnés par

$$\sigma_X^A = 1.11, \qquad \sigma_X^B = 2.8,$$

$$CV_X(A) = \frac{\sigma_X^A}{\overline{X}_A} \simeq 0.12 \qquad CV_X(B) = \frac{\sigma_X^B}{\overline{X}_B} \simeq 0.30$$

Nous remarquons que même si les deux groupes ont des moyennes quasiment identiques, le groupe B est beaucoup plus dispersé que le groupe A car

$$\sigma_X^b > \sigma_X^A \text{ et } CV_X(B) > CV_X(A).$$

Les étudiants de ce groupe ont des notes plus irréguliers. On peut dire donc que le groupe B est moins homogènes que le groupe A. En observant les valeurs du tableau, on voit que c'est cohérent.

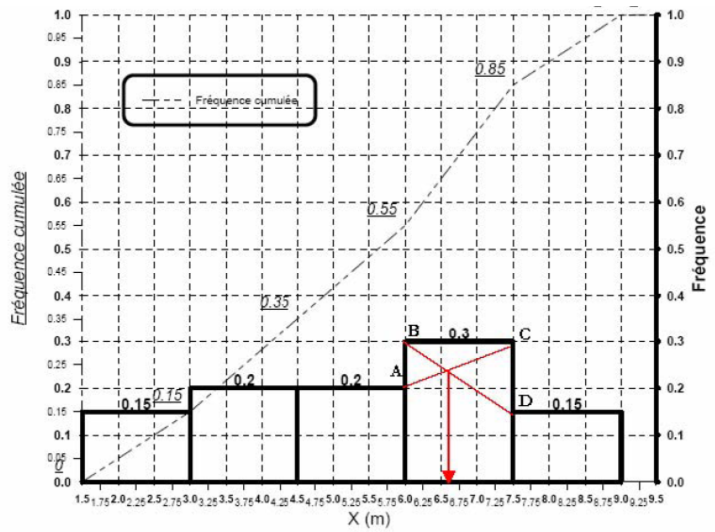
<u>Solution de l'exercice 04</u> :

Détermination graphique le mode et médiane

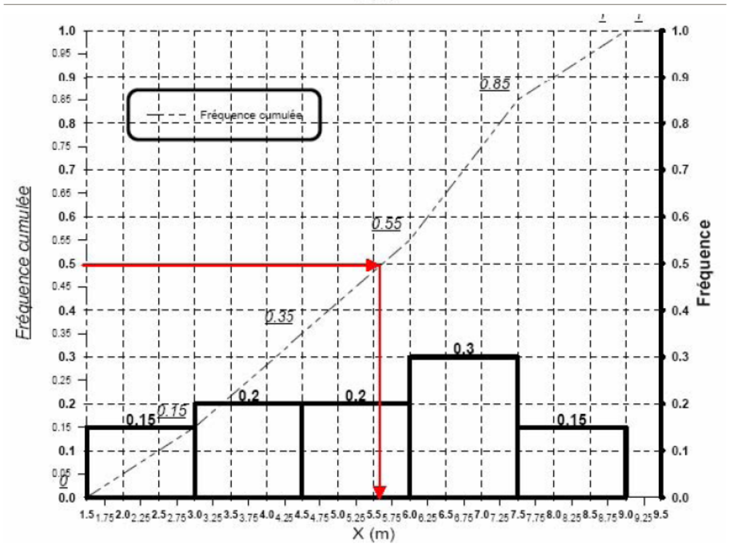
D' après le graphe $Mo = 6.52$. La médiane est l'abscisse de la fonction cumulative pour une fréquence cumulée de 50%. D' après le graphe $Me = 5.6m$.

La moyenne arithmétique :

$$\overline{X} = \sum_{i=1}^{5} f_i c_i = 5,4m$$


Fréquence cumulée
Fréquence
X (m)
0.15
0.2
0.2
0.3
0.15
0.35
0.55
0.85
A
B
C
D

,


Fréquence cumulée
Fréquence
X (m)
0.15
0.2
0.2
0.3
0.15
0.35
0.55
0.85

Chapitre 2

Calcul des probabilités

Sommaire

2.1 Analyse combinatoire

L'analyse combinatoire est une branche des mathématiques qui étudie comment compter les objets. Elle fournit des méthodes de comptage (dénombrements) qui est particulièrement utile dans la théorie des probabilités .

Dénombrer : Savoir comment compter quelque chose (des objets.)

Si on veut construire des objets à partir d'un ensemble finie, formé d'éléments. Nous devons distinguer deux notions importantes dans la façon de choisir les éléments de l'ensemble.

2.1.1 Notion d'ordre

Une disposition [1] est dite ordonnée, si lorsqu' à chaque fois qu'un élément change de position la disposition change.

2.1.2 Notion de répétition

Si un élément peut figurer plus d'une fois dans une disposition, on dit que la disposition est avec répétition. Si un élément peut apparaître 0 ou 1 fois dans une disposition on dit que la disposition est sans répétition

Exemples 2.1.1. *Soit E un ensemble ayant trois éléments $E = \{x, y, z\}$. Le tableau suivant, résumer tous les cas possibles de Choisir deux éléments dans cet ensemble*

Disposition	*avec répétition*	*sans répétition*
avec ordre (ordonnée)	*(x ; x),(x ;y), (x ;z), (y ;x), (y ;y), (y ;z), (z ;x), (z ;y), (z ;z)*	*(x ;y), (x ;z), (y ;x) (y ;z), (z ;x), (z ;y)*
sans ordre (non ordonnée)	$\{x; x\}, \{x; y\}, \{x; z\}, \{y; y\}, \{y; z\}, \{z; z\}$	$\{x; y\}, \{x; z\}, \{y; z\}$

Les techniques de dénombrement se baser selon 3 catégories :
les arrangements - les permutations -les combinaisons

2.1.3 Arrangement

Définition 2.1.1. *Soit E un ensemble fini de n objets. un arrangement de p de ces éléments(objets) est une suite ordonnée de p éléments pris parmi ces n éléments.*

Il existe deux types d'arrangements : avec et sans répétition.

Arrangement sans répétition

On appelle arrangement sans répétition de p objets parmi n objets de E ($p \leq n$), toute suite ordonnée et sans répétition de p objets choisis parmi les n objets. Le nombre d'arrangements sans répétition, noté A_n^p, est :

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

avec $0 \leq p \leq n$.

1. Une disposition est l'ensemble formé d'éléments choisis parmi les n éléments de l'ensemble étudié

Arrangement avec répétition

On appelle arrangement avec répétition de p objets parmi n objets de E , toute suite ordonnée et avec répétition de p objets choisis parmi les n objets. Le nombre d'arrangements avec répétition, noté n^p, est :

$$n^p = \underbrace{n \times n \times .... \times n}_{pfois}.$$

Exemples 2.1.2. *Combien de mots de deux lettres peut-on former à l'aide des quatre lettres a,b,c,d,e ?*
Dans cette expérience l'ordre est important et avec répétition. Il y a donc $4^2 = 16$ *mots.*

Permutation

Définition 2.1.2. *Soit E un ensemble fini de n objets. Une Permutation de n objets distincts est une suite ordonnée de n objets ou est un arrangement de n à n de ces objets.*

Permutation sans répétition

Les permutations de n éléments est un cas particulier des arrangements sans répétition lorsque $p = n$.
Le nombre de permutations de n objets noté p_n, est : $p_n = n! = n(n-1)!$.

Exemples 2.1.3. *Le nombre de permutations des lettres du mot math est* $p_4 = 4!$

Permutation avec répétition

Le nombre de permutations de n objets avec répétitions est donnée par

$$p_n^{(k_1,k_2,..,k_n)} = \frac{n!}{k_1! \times k_2! \times ... \times k_n!},$$

où $k_1, k_2, .., k_n$ désignent le nombre d'objets identiques.

Exemples 2.1.4. *Combien y-a-t-il d'anagrammes du mot STATISTIQUES ? Le nombre d'anagrammes du mot STATISTIQUES est :*

$$p_{12}^{(3,3,2)} = \frac{12!}{3! \times 3! \times 2!} = 6652800$$

2.1.4 Combinaisons

Combinaison sans répétitions (sans remise)

Définition 2.1.3. *Soit E un ensemble fini de n objets. La combinaison de p objets est une collection non ordonnée de p objets parmi de n objets dans l'ensemble E sans répétition. Le nombre de combinaisons de p objets parmi n et sans répétition, noté C_n^P, est :*

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

avec $0 \leq p \leq n$.

Exemples 2.1.5. *Une urne contient 5 boules , indiscernables au toucher. On tire 3 boules simultanément sans remise.*
Combien le nombre de tirages possibles. Le tirage simultané indique que l'ordre n'est pas important et sans remise c'est-à-dire sans répétition, donc le nombre de tirages possible c'est une combinaison de 3 boules prises parmi 5 boules.

$$C_5^3 = \frac{5!}{2! \times 3!} = 20 \textit{ tirages possibles.}$$

Combinaison avec répétitions (avec remise)

Le nombre de combinaisons de p objets parmi n et avec répétitions, noté C_n^{n+P-1}, est :

$$C_n^{n+P-1} = \frac{n!}{(p-1)! \times (n+p-1)!}.$$

Propriétés 2.1.1. *Pour tous n et p, tels que $0 \leq p \leq n$, on a :*

1. $C_n^0 = C_n^n = 1$.
2. $C_n^1 = C_n^{n-1} = 1$.
3. $C_n^p = C_n^{n-p}$, *(formule symétrie).*
4. *Si* $1 \leq p \leq n-1$, $C_{n-1}^{p-1} + C_{n-1}^p = C_n^p$, *(formule de pascal).*
5. *Formule de Binôme de Newton, :*

$$\forall n \geq 0, \quad (x+y)^n = \sum_{k=0}^{n} C_n^k x^{n-k} y^k.$$

2.2 Notions de base des probabilités

- Une expérience aléatoire est une expérience sur lesquelles le résultat n'est pas connu d'avance et elle peut donner des résultats différents lorsqu'on répète l'expérience dans les mêmes conditions.

  Exemples 2.2.1. *Lancer un dé , jet d'une pièce de monnaie, Durée de vie d'une ampoule, l'extraction des boules dans une urne, lancer une fléchette en direction d'une cible,...*

- Univers appelé aussi l'espace des états , noté Ω est l'ensemble des résultats possibles de l'expérience

Exemples 2.2.2. *1- Lancer d'une pièce de monnaie :* $\Omega = \{F, P\}$.
2- Lancer d'une pièce de monnaie deux fois successifs : $\Omega = \{(F, P), (F, F), (P, F), (P, P)\}$.
3- Lancer indéfiniment une pièce de monnaie $\Omega = \{F, P\}^{\mathbb{N}}$.
4- Lancer d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$.
5- La mesure de la durée de vie d'une ampoule. $\Omega = [0, \infty[$.

Remarque 2.2.1. ·
Avec les exemples précédents on observe que l'espace des états peut être :
- *Fini (deux premiers exemples),*
- *Infini dénombrable (troisième exemple),*
- *Infini non dénombrable (cinquième exemple).*

Événements

Définition 2.2.1. *On appelle évènement associé à une expérience aléatoire toute proposition logique relative au résultat de cette expérience.*

Remarque 2.2.2. ·
D'après la définition (2.2.1), un événement correspond un sous-ensemble (une partie) A de l'univers Ω

- A est un événement certain si $A = \Omega$.
- A est un événement impossible si $A = \phi$.
- $\overline{A}$ est l'événement contraire d un événement A si $\overline{A} = \Omega \setminus A = \{w \in \Omega \ et \ w \notin A\}$.
- A_1 et A_2 deux événements sont incompatibles si $A_1 \cap A_2 = \phi$.

Tribu-Algèbre

Définition 2.2.2. *Soit Ω un univers, une algèbre sur Ω est une famille $\mathcal{A}$ de sous-ensembles de de Ω vérifiant :*
i) $\Omega \in \mathcal{A}$.
ii) $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$.
iii) $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

Définition 2.2.3. *Soit Ω un univers, une σ algèbre ou une tribu sur Ω est une famille $\mathcal{A}$ de sous-ensembles de Ω vérifiant :*
i) $\Omega \in \mathcal{A}$.
ii) $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$.
iii) Soit $(A_i)_{i\in\mathbb{N}}$, est une suite de famille d'éléments de $\mathcal{A}$, alors

$$\bigcup_{i\in N} A_i \in \mathcal{A}$$

.

Exemples 2.2.3. $*$
i) $\mathcal{A}_0 = \{\phi, \Omega\}$ *est une tribu sur Ω appelée tribu triviale sur Ω.*
ii) $\mathcal{A} = \mathcal{P}(\Omega)$ *est une tribu sur Ω appelée tribu grossière sur Ω.*
iii) Soit $\Omega = \{a, b, c\}$ et $\mathcal{A} = \{\phi, \Omega, \{a, c\}, \{c\}\}$, alors $\mathcal{A}$ n'est pas une tribu sur Ωa car :
$\overline{\{a, c\}} = \{b\} \notin \mathcal{A}$

Si univers est fini ou dénombrable, on considère en général la tribu grossière $\mathcal{P}(\Omega)$ et tout élément de la tribu A appelée un événement.

Définition 2.2.4. *(Espace probabilisable)*
On appelle espace probabilisable le couple $(\Omega, \mathcal{A})$ où $\mathcal{A}$ constitue une tribu sur Ω.

Définition 2.2.5. *(Une probabilité)*
Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle probabilité sur $(\Omega, \mathcal{A})$ toute application P : $\mathcal{A} \to \mathbb{R}$ satisfaisant les trois axiomes(axiomes de Kolmogorov) suivants :

1. $\forall B \in \mathcal{A}, P(B) \geq 0$.
2. $P(\Omega) = 1$.

3. *Pour toute suite $(A_i)_{i\in\mathbb{N}}$ d'éléments de $\mathcal{A}$ deux à deux incompatibles, on a*

$$p\left(\bigcup_{n\in N} A_n\right) = \sum_{n=0}^{\infty} A_n.$$

Le triplet $(\Omega, \mathcal{A}, P)$ est appelé espace de probabilité.

Propriétés 2.2.1. *Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité, $A, B \in \mathcal{A}$, alors*

1. $\forall A \in \mathcal{A},\ 0 \leq P(A) \leq 1.$
2. $P(\phi) = 0.$
3. $P(\overline{A}) = 1 - P(A).$
4. *P est une application croissante (Croissance)*

$$A \subset B \Rightarrow P(A) \leq P(B).$$

5. $P(A \cup B) = P(A) + P(B) - P(A \cap B).$

2.2.1 Équiprobable

Soit Ω un univers de cardinal fini. On dit qu'il y a équiprobabilité lorsque tous les événements élémentaires ont la même probabilité (la même chance de réalisation). Dans ce cas, on a

$$P(A) = \frac{\text{Nombre de cas favorables}}{\text{Nombre de cas possibles}} = \frac{Card(A)}{Card(B)}.$$

2.2.2 Probabilité conditionnelle

Définition 2.2.6. *Soit A et B deux événements, avec P(A)>0, La probabilité de B sachant A est donnée par :*

$$P_A(B) = \frac{P(A \cap B)}{p(B)}.$$

Propriétés 2.2.2. ·
Soit A et B deux événements avec $P(A) > 0$ et $P(B) > 0$, on a :

$$P(A \cap B) = P(B)P_B(A) = P(A)P_A(B).$$

2.2.3 Formule de BAYES

Définition 2.2.7. *Un système complet d'événement* $(A_i)_{1\leq i\leq n}$ *est une partition de* Ω *:*
- $\forall i \neq j \Rightarrow B_i \cap B_j = \phi$.
- $\cup_{i=1}^{n} B_i = \Omega$.

Théorème 1. *(Formule des probabilités totales)*
Soit $(B_i)_{1\leq i\leq n}$ *un système complet d'événement et soit* $A \in \mathcal{A}$*. Alors :*

$$P(A) = \sum_{i=1}^{n} P(B_i)P_{B_i}(A).$$

Théorème 2. *(Théorème de Bayes)*
Soit $(B_i)_{1\leq i\leq n}$ *un système complet d'événement et soit* $A \in \mathcal{A}$ *de probabilité non nul*

$$P_A(B_i) = \frac{P(B_i)P_{B_i}(A)}{\sum_{i=1}^{n} P(B_i)P_{B_i}(A)}.$$

2.2.4 Indépendance

Définition 2.2.8. *Soit* A *et* B *deux événements. On dit que* A *et* B *sont indépendants si et seulement si :*

$$P(A \cap B) = P(A)P(B).$$

Définition 2.2.9. *Soit* A *et* B *deux événements avec* $P(A) > 0$ *et* $P(B) > 0$*. On dit que* A *et* B *sont indépendants si et seulement si*

$$P_B(A) = P(A).$$

Remarque 2.2.3. *La réalisation de l'événement* B *ne donne aucune influence quant à la réalisation de l'événement* A*.*

Propriétés 2.2.3. *Si les événements* A *et* B *sont indépendants, alors les trois propositions suivantes sont équivalences :* A *et* $\overline{B}$ *sont indépendants,* B *et* $\overline{A}$ *sont indépendants et* $\overline{A}$ *et* $\overline{B}$ *sont indépendants.*

2.3 Exercice corrigés

Exercice 2.3.1. 1. *On considère l'algorithme :*

A un entier naturel,
répéter 3 fois
A prend une valeur aléatoire entière entre 5 et 7.
Fin répéter
Afficher A.

-***Combien de résultats peut-on obtenir ?***

2. *Le client d'une banque se rappelle que 2, 4, 7 et 9 sont les chiffres d'un code d'accés à quatre chiffres différents pour le distributeur automatique de billets. Malheureusement, il a oubliè le code.*

   -***Calculer le plus grand nombre possible d'essais nécessaires pour obtenir le code secret****.*

3. *Une urne contient 10 jetons numrotés de 1 à 10, indiscernables au toucher. On tire 3 jetons simultanément.*

   Combien de tirages différents peut-on faire contenant au moins un jeton à numèro pair ?

4. *Trouver le nombre de façons différentes de permuter les chiffres 1, 1, 1, 3, 3 ?*

Solution

1. L'algorithme affiche le nombre A où le tirage aléatoire d'un numéro entre 5 et 7, lors de 3 tirages. Il aura donc $3^3 = 27$ choix possibles.

2. Un code comportant des chiffres distincts étant vu comme une liste ordonnée des 4 chiffres (2,4,7,9) est une permutation de ces quatre chiffres. Il y en a donc $4! = 24$ choix possibles.

3. Il s'agit de tirages simultanés, le nombre total de tirages de 3 jetons parmi 10 est donc de $C_3^{10} = 120$. Parmi ces 120 tirages, il y a $C_3^5 = 10$ tirages où les trois numéros sont impairs. Le nombre de tirages avec au moins un jeton à numéro pair est donc égal à $120 - 10 = 110$. (deuxième méthode) : $2 \times C_2^5 C_1^5 + C_3^5 = 110$.

4. Le nombre comportant des chiffres distincts étant vu comme une liste ordonnée des 5 chiffres (1,1,1,3,3) est une permutation de ces cinq chiffres.

Il y en a donc, a priori, 5! Mais si au sein de ces nombres, on " permute " les trois chiffres 1, on retombe sur le même nombre. Pour éviter de compter ces permutation trois fois, on doit diviser 3! par le nombre de permutations possibles des trois chiffres, soit $3! = 6$. Le nombre de façons différentes de permuter les 5 chiffres est donc égal à $7!/3! = 840$ choix possibles.

Exercice 2.3.2. *Avec les lettres du mot CONSEIL, combien de mots différents de 7 lettres peut-on former*
– *a) au total ?*
– *b) Si les voyelles doivent être ensemble ?*
– *c) Si les voyelles doivent être ensemble de même que les consonnes ?*
– *d) Si le C doit être à la première ou à la seconde position du mot ?*

<u>**Solution de l'exercice** 3</u>

a) $7! = 5040$
b) $5.3!.4! = 720$
c) $2.3!.4! = 288$
d) $2.6! = 1440$

Exercice 2.3.3. *Dans une classe de 30 étudiants : 18 filles et 12 garçons. On veut former un comité composé de trois membres. Combien de comités peut-on former si :*
– *a) Aucune conditition n'est supposée.*
– *b) Le comité doit contenir les deux sexes.*
– *c) Le comité doit contenir au moins une fille.*

a) $C_{30}^{3} = 109620$
b) $C_1^{12}C_2^{18} + C_1^{18}C_2^{12} = 3024$
c) Le contraire de "au moins une fille" est "aucun fille ". Le nombre comités contenant au moins une fille est égal au nombre total des comités diminué du nombre comités ne contenant aucun fille, on conclut que le nombre comités contenant au moins un célibataire est égal à $C_{30}^3 - C_{12}^3 = 109620 - 220 = 109400$

Exercice 2.3.4. *Un sac contient 5 jetons verts (numérotés de 1 à 5) et 4 jetons rouges (numérotés de 1 à 4). On tire successivement et au hasard 3 jetons du sac, sans remettre le jeton tiré. Déterminer le nombre de tirages permettant d'obtenir*
– *a) 3 jetons verts,*
– *b) aucun jeton vert,*
– *c) au plus 2 jetons verts,*

– *d) exactement 1 jeton vert.*
– *2. On tire simultanément et au hasard 3 jetons du sac. Reprendre alors les questions a), b), c) et d).*

Solution
a) $A_5^3 = 5.4.3 = 60.$
b) $A_4^3 = 4.3.2 = 24.$
c) $A_4^3 + 3.A_5^1 A_4^2 + 3A_5^2 A_4^1 = 37.$
d) $3A_5^1.A_4^2 = 180.$
– Tirage simultané :
a) $C_5^3 = 10 = 60.$
b) $C_4^3 = 4.$
c) $C_4^3 + C_5^1 C_4^2 + C_5^2 C_4^1 = 74.$
d) $C_5^1.C_4^2 = 30.$

Exercice 2.3.5. ***(A rendre)***
Dans tout l'exercice, on suppose qu'il n'y a pas de répétition.
– *1. Combien de nombres de 4 chiffres peut-on former à l'aide des 7 chiffres 1, 2, 4, 5, 6, 8, 9 ?*
– *2. Combien de ces nombres sont inférieurs à5000 ?*
– *3. Combien sont paires ?*
– *4. Combien sont impaires ?*
– *5. Combien sont des multiples de 5 ?*

Exercice 2.3.6. ***(A rendre)***
Lors d'un examen, un étudiant doit répondre à 10 questions sur 13.
– *1. Combien de choix possibles a-t-il ?*
– *2. Combien de choix a-t-il s'il doit répondre aux 2 premières questions ?*
– *3. Combien de choix a-t-il s'il doit répondre à la première ou à la deuxième question exclusivement ?*
– *4. Combien de choix a-t-il s'il doit répondre à exactement 3 des 5 premières questions ?*
– *5. Combien de choix a-t-il s'il doit répondre 'a au moins 3 des 5 premières questions ?*

Exercice 2.3.7. *On lance un dé à six faces et on regarde le nombre inscrit sur sa face supérieure.*

1. *Donner l'ensemble fondamental Ω et son cardinal.*
2. *Déterminer la probabilité de chacun des événements :*

$$A : \text{"on obtient un nombre pair"}, \quad B : \text{"on obtient un nombre impair,"}$$
$$C : \text{"on obtient un nombre négatif,"} \quad D : \text{"on obtient un nombre inférieure ou égale à 4.}$$

3. *Calculer : $P(A \cup B), P(B \cap D), P(\overline{A}), P(\overline{D} \cap A)$ et $P(\overline{A} \cap D)$.*

<u>Solution</u>

(1) $\Omega = \{1, 2, 3, 4, 5, 6\}$ et Card(Ω)=6.

(2) L'événement A : " on obtient un nombre pair ", d'où : $A = \{2, 4, 6\}$.

$$p(A) = \frac{card(A)}{card(\Omega)} = \frac{1}{2}.$$

(3) L'événement B est le contraire de l'événement A. Donc $B = \overline{A}$, d'où :

$$p(B) = p(\overline{A}) = 1 - p(A) = 1/2.$$

(4) L'événement C est événement impossible : $C = \phi$. Alors : $P(C) = 0$.

(5) L'événement D : " On obtient un nombre inférieure ou égale à 4 ", d'où : $D = \{1, 2, 3, 4\}$. Alors :

$$p(D) = \frac{card(D)}{card(\Omega)} = \frac{2}{3}.$$

♠

$$\begin{cases} P(A \cup B) = \dfrac{card(A \cup B)}{card(\Omega)} \\ \text{et} \\ A \cup B = A \cup \overline{A} = \Omega \end{cases} \Rightarrow P(A \cup B) = 1 = P(\Omega)$$

.

♠

$$P(B \cap D) = \frac{card(B \cap D)}{card(\Omega)} \ \ et \ \ B \cap D = \{1, 3\} \Rightarrow P(B \cap D) = 1/3$$

.

♠

$$P(\overline{D} \cap A) = \frac{card(\overline{D} \cap A)}{card(\Omega)} \;\; et \;\; \overline{D} \cap A = \{6\} \Rightarrow P(\overline{D} \cap A) = 1/3$$

.

♠

$$On \text{ a } \overline{A} \cap D = B \bigcap D \Rightarrow P(\overline{A} \cap D) = P(B \cap D) = 1/3$$

.

Exercice 2.3.8. *Un outil fabriqué en série peut être présenter deux défauts désignés par A et B. 0.76 des outils ne présentent aucun défaut, 0.11 présentent le défaut A et 0.16 le défaut B. Une personne achèt un outil, déterminer la probabilité que l'outil présente :*

1. *Au moins l'un des défauts.*
2. *Les deux défauts.*
3. *Le défaut B seulement.*
4. *Un seul des deux défauts.*

Solution

On a :

$$\begin{cases} A \text{ un outil fabriqué pré sente le dé faut A} \Rightarrow \text{ P(A) = 0.11} \\ \\ B \text{ un outil fabriqué pré sente le dé faut B} \Rightarrow \text{ P(B) = 0.16} \\ \\ C \text{ un outil fabriqué ne pré sente aucun dé faut } \Rightarrow \text{ P(A) = 0.76} \end{cases}$$

Avec langage ensembliste on a

– Au moins l'un des défauts $= A \cup B$.

  On sait que :

$$P(A \cup B) = p(\overline{\overline{A} \cap \overline{B}}).$$

  D'où :

$$P(A \cup B) = 1 - P(\overline{A} \cap \overline{B}) = 1 - 0.76 = 0.24.$$

– Les deux défaut$=A \bigcap B$, d'où :

$$P(A \cap B) = P(\overline{\overline{A} \cup \overline{B}}) = 1 - P(\overline{A} \cup \overline{B}) = 1 - [p(\overline{A}) + p(\overline{B}) - P(\overline{A} \cap \overline{B})$$

$$= 1 - [1 - p(A) + 1 - P(B) - P(\overline{A} \cap \overline{B})] = 1 - (0.89 + 0.84 - 0.76)$$

$$= 0.03.$$

– Le défaut B seulement$= \overline{A} \cap B$

$$P(\overline{A} \cap B) = P(B/A) = P(B) - P(B \cap A) = 0.16 - 0.03 = 0.13.$$

– Un seul des deux défauts $=(\overline{A} \cap B) \cup (\overline{B} \cap A) = E$.

$$P(E) = P(\overline{A} \cap B) + P(\overline{B} \cap A) = P(B/A) + P(A/B)$$

$$= P(A) + P(B) - 2P(A \cap B) = 0.11 + 0.16 - 0.06 = 0.21.$$

Deuxième méthode

On présenter les résultats dans un tableau à double entrée :

	B avec défaut	**$\overline{B}$ sans défaut**	**Total**
A avec défaut	$p(A \cap B)$=**0.03**	$p(A \cap \overline{B})$=**0.08**	**0.11**
$\overline{A}$ sans défaut	$p(\overline{A} \cap B)$=**0.13**	$p(\overline{A} \cap \overline{B})$=**0.76**	**0.89**
Total	**0.16**	**0.84**	**1**

Exercice 2.3.9. *On dispose de deux urnes et d'un dé cubique bien équilibré dont les faces sont numérotées de 1 à 6. L'urne U_1 contient trois boules rouges et une boule noire. L'urne U_2 contient trois boules rouges et deux boules noires. Le joueur lance le dé ; si le résultat est 1, il tire au hasard une boule dans l'urne U_1, sinon il tire au hasard une boule dans l'urne U_2. On considère les évènements suivants :*

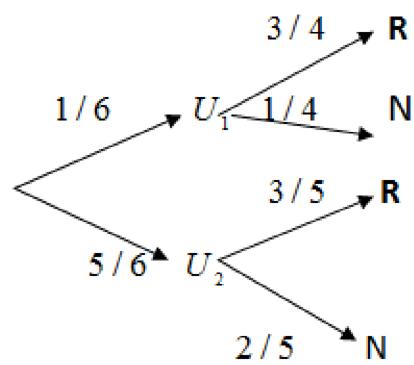
– *A : " obtenir 1 en lançant le dé "*

– *B : " obtenir une boule noire ".*

1. *Construire un arbre pondéré traduisant cette expérience aléatoire.*
2. *Montrer que la probabilité d'obtenir une boule noire est 3/8*
3. *Sachant que l'on a tiré une boule noire, calculer la probabilité d'avoir obtenu 1 en lançant le dé.*

Solution

1. **Arbre pondéré** :



2. L'événement B peut être écrite sous la forme :

$$B = (N \bigcap U_1) \bigcup (N \bigcap U_2) \Rightarrow P(B) = P(N \bigcap U_1) + P(N \bigcap U_2).$$

D'après la formule des probabilités totales on a :

$$P(B) = P(U_1)P(N/U_1) + P(U_2)P(N/U_2) = \frac{1}{6}.\frac{1}{4} + \frac{5}{6}.\frac{2}{5} = \frac{3}{8}.$$

3. La probabilité d'avoir obtenu 1 sachant que l'on a tiré une boule noire est :

$$P(A/B) = P(U_1/N) = \frac{P(U_1 \bigcap N)}{P(N)},$$

et par la formule de Bayes on obtient :

$$\frac{P(U_1 \bigcap N)}{P(N)} = \frac{P(N/U_1).P(U_1)}{P(N)} = \frac{\frac{1}{4}.\frac{1}{6}}{\frac{3}{8}} = 1/9.$$

Exercice 2.3.10. *(A rendre)*
Une urne contient 12 boules : 3 rouges, 4 bleues et 5 jaunes. On tire simultanément sans remise 3 boules. Calculer la probabilité des événements suivants :

1. *A : "les trois boules sont rouges".*
2. *B : "on a tiré une boule de chaque couleur".*
3. *C : "aucune des trois boules n'est rouges".*
4. *D : "au moins une des trois boules est rouge".*