RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche

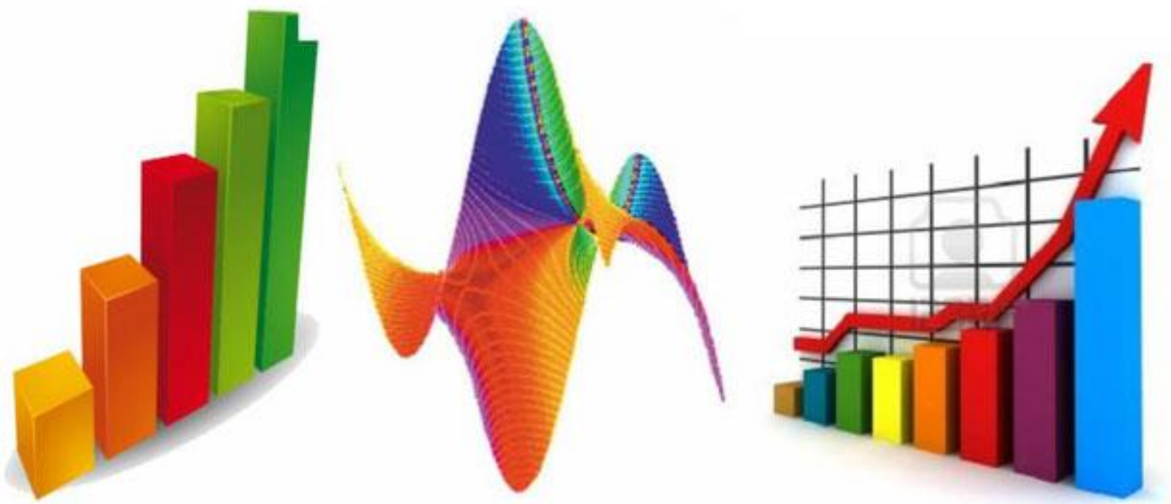
Scientifique Université Abou Bekr Belkaid Tlemcen

Faculté des sciences

Département de Génie Civil

UNIVERSITÉ DE TLEMCEN

Probabilité et statistique appliquées en Génie Civil



Mme BENAHCHILIF SOUAD (

Année universitaire: 2017/2018

Sommaire

La liste des figures

La liste des tableaux

Symboles et Notations

Symbole	Signification
[ ]	La partie entière.
Card(Ω)	Le cardinal : nombre d’éléments de l’ensemble Ω.
N	Ensemble des nombres entiers naturels.
Z	Ensemble des nombres entiers relatifs.
R	Ensemble des nombres réels.
R^2	Ensemble des couples de nombres réels.
$V.S$	La variable statistique
$^{M}0$	Le mode.
$\overline{x}$	La moyenne d’une série statistique X.
σ_X	L’écart-type de X.
Var(X)	La variance de X.
Cov(X,Y)	La covariance entre les variables X et Y .
ρXY	Le coefficient de corrélation entre les variables X et Y .
^{F}X	La fonction s’appelle la fonction de répartition du caractère X
$\sum_i^n X$	La somme de i à n valeur
P	La probabilité

Série statistique à une seule variable

Module : probabilité et statistique

Ce cours ''Série statistique à une seule variable'' vise à définir une série statistique à une seule variable, Distinguer les différents types d'une série statistique à une seule variable, Transférer un tableau statistique en graphique, calculer les différents paramètres de position et de dispersion, et assembler toute les informations obtenues pour tirer des conclusions. À la fin de ce cours, l'apprenant sera capable de faire une expérience statistique de son choix même en dehors de sa spécialité.

Série statistique à une seule variable

Module : probabilité et statistique

Introduction

« Étude méthodique des faits sociaux par des procédés numériques (classements, dénombrements, inventaires chiffrés, recensements) destinée à renseigner les gouvernements » : ceci est la définition du mot « statistique » dans le dictionnaire Petit Robert.

Dès l'Antiquité (à Sumer, en Mésopotamie, en Égypte...), des gouvernements ont effectivement utilisé des « séries statistiques » pour être mieux renseignés sur leurs États et les gérer en conséquence.

Peu après 1750, on commence à faire des représentations graphiques, la moyenne et la médiane sont de plus en plus utilisées pour *résumer et décrire une série statistique.*

Les physiciens, et depuis longtemps les astronomes, doivent tenir compte de séries de mesures pour un même phénomène, ces variations étant en partie aléatoires. A partir des observations statistiques, les économistes tentent de faire des prévisions *en essayant de maîtriser l'incertitude.*

Un chapitre des mathématiques va répondre à ces besoins car les mathématiciens ont commencé (1650) à créer des outils pour étudier les phénomènes aléatoires: *les probabilités.*

Dans notre environnement quotidien (météo, sondages...), professionnel (cabinets d'assurance, de gestion, laboratoires d'analyses médicales, contrôles qualité dans l'industrie), universitaire (physique, chimie, biologie, psychologie, économie, archéologie...), dans tous ces domaines, les statistiques et les probabilités interviennent.

Il est indispensable au citoyen d'aujourd'hui de comprendre ce que sont les statistiques pour savoir ce que veulent réellement dire les informations qu'il reçoit.

Et il est souhaitable qu'un *apprenant* connaisse et sache utiliser les notions de base des statistiques et de calcul des probabilités.

Dans ce module, il s'agit de *statistiques descriptives*. On va s'attacher à résumer des séries statistiques à une seule variable par des nombres significatifs (de position et de dispersion) pour permettre l'utilisation et la comparaison de ces séries.

Pour les explications, les exemples et les exercices qui ont été choisis comportent peu de données. Dans la réalité du travail des statisticiens, il s'agit d'étudier des séries statistiques pour lesquelles les données sont beaucoup plus nombreuses et les outils informatiques permettent de le faire.

TRAITEMENT STATISTIQUE DE L'INFORMATION

I) COLLECTE DES DONNEES

I.1 L'épreuve statistique

C'est une expérience que l'on provoque

- ***Exemple 1*** *: On va étudier l'ensemble des employés des employés d'une usine selon leur salaire (ou leur poids, ou leur niveau d'étude).*

I.2 La population statistique (Notée P ou Ω)

C'est l'ensemble de départ de notre expérience

- ***Exemple 2*** *: Tous les employés de l'usine*

I.3 L'échantillon

C'est une partie de la population (sous ensemble)

- ***Exemple 3*** *: On prend seulement les employés du service gestion.*

I.4 L'unité statistique ou l'individu (ω)

C'est les éléments de notre population ou échantillon.

- ***Exemple 3 :*** *l'employé*

I.5 Caractère statistique

C'est le résultat de notre expérience. , on distingue deux types de caractère :

I.5.1 Le caractère qualitatif : ce caractère ne peut se traduire par des chiffres

- ***Exemple 4*** : *Etat civil de l'employé (Modalité : marié, célibataire, veuf,…)*
- *Niveau d'étude de l'employé (Modalité : Bac, Licence, ..)*
- *Lieu de naissance (Modalité : Tlemcen, Oran, …*

I.5.2 Le caractère quantitatif : ce qui se traduit par des chiffres

- ***Exemple 5*** : *le poids, la taille, nombre d'enfants. nombre de pièce*

On distingue deux cas dans le caractère quantitatif :

- **Le quantitatif discret** : *Le quantitatif discret*: Dont les résultats possibles (modalités) sont des entiers naturels dénombrables, c.-à-d. sont des valeurs isolés.

- ***Exemple 6*** : *nombre de pièce d'une maison (Modalité : 1, 2,3……), le salaire, nombre d'enfant (Modalité : 0, 1, 2,3)*

- **Le quantitatif continue**: Dont les résultats possibles (modalités) peuvent prendre toutes les valeurs comprises entre deux nombres, par suite, les modalités sont données par des intervalles fermé d'un côté et ouvert de l'autre côté pour éviter la répétition et/ou la confusion entre les données.

- ***Exemple 7*** *:* surface de la maison (S € [60, 120] m^2*),* Poids, Taille, Durée de vie, …..

REMARQUE :

- ✓ Un même individu peut avoir plusieurs caractère (étude de la croissance : poids et taille)
- ✓ Un même individu (unité statistique) ne peut pas avoir plusieurs modalités (plusieurs valeurs)

Exercice 1

Déterminer à chaque expérience statistique, la population statistique Ω, la variable statistique et son type.

1) Répartition des étudiants de votre promotion selon la mention obtenue sur le diplôme du bac.
2) Etude du temps de validité des lampes électriques.
3) Etude des notes des étudiants au contrôle continu.
4) Recensement des familles selon leurs nombres d'enfant.
5) Recensement des villes algériennes selon leur température en hiver.
6) Déterminer le poids des marchandises chargées sur un camion.

Solution

N°	Population Ω	Unité statistique ω	La variable statistique	Type	Modalité
1	Les étudiants	L'étudiant	mention	qualitatif	TB, B, AB
2	Les lampes	La lampe	temps	Quantitatif continu	$[0, \alpha]$
3	Les étudiants	L'étudiant	notes	Quantitatif discret	4, 7,10, …..
4	Les familles	La famille	Nombre d'enfant	Quantitatif discret	0, 1, 5, …
5	Les villes	La ville	Température	Quantitatif continu	$[0, \alpha]$
6	Les marchandises	La marchandise	Le poids	Quantitatif continu	$[0, \alpha]$

Exercice 2

a) La variable statistique "couleur de maisons d'un quartier" est-elle :

○ *qualitative* ○ *quantitative*

○ *Discrète* ○ *continue*

b) La variable statistique "revenu brut" est-elle :

○ *qualitative* ○ *quantitative*

○ *discrète* ○ *continue*

c) La variable statistique "nombre de maisons vendues par ville" est-elle :

○ *qualitative* ○ *quantitative*

○ *discrète* ○ *continue*

Solution

a) Qualitatif
b) Quantitatif continu
c) Quantitatif discret

II Distribution statistique

II .1 Effectif ou Fréquence Absolue (n_i)

n_i représente le nombre d'individus ou d'unité statistique qui ont la même modalité c_i. Le couple (c_i, n_i) est dit une série statistique. L'effectif total est notée N ou n :

$$n = N = \sum_{i=1}^{k} n_i$$

II .2 Fréquence ou Fréquence relative (fi)

La fréquence, notée par 'fi', est le quotient (la division) de l'effectif de la valeur 'ni' par l'effectif total 'N'. Intuitivement, elle indique la proportion de la présence de la valeur dans la liste.

$$f_i = \frac{n_i}{N} = \frac{n_i}{n}$$

Ou : N est l'effectif total, et la $\sum_{i=1}^{k} f_i = 1$

II.3 Effectifs cumules (n_{ic}):

C'est la somme de toute les fréquences absolues jusqu'à la valeur n_i.

$$n_{ic} = \sum_{j=1}^{i} n_j$$

$n_{ic}= n_1+n_2+n_3+\ldots\ldots n_i$

II.4 Les fréquences cumulées (fic)

C'est la somme de toute les fréquences relatives jusqu'à la valeur de f_i.

$$f_{ic} = \sum_{j=1}^{i} f_j$$

Ou

$$f_{ic} = f_1 + f_2 + \cdots \ldots \ldots \ldots \ldots + f_i = \frac{n_{ic}}{N} = \frac{n_{ic}}{n}$$

REMARQUE :

- ✓ La distribution statistique est formée par le couple (c_i, n_i) i=1…….k, qui sera par série statistique.
- ✓ La variable statistique et redéfinit par le caractère statistique.

Exemple 1

Une étude statistique a été effectuée pour déterminer le climat de la wilaya de Tlemcen pendant l'année 2016, les résultats sont montrés dans le tableau suivant :

Individus	1	2	3	4	5	6	7	8	9	10	11	12
Modalité	E	E	P	O	P	O	E	P	O	O	P	P

O : Orageux, E : Ensoleillé, P : Pluvieux

1) Déterminer la population, l'unité statistique (individus), la variable statistique (le caractère), les modalités.
2) Déterminer le tableau statistique.

Solution

1) Population : mois de l'année
   Caractère : Climat
   Modalité : c_1=Ensoleillé, c_2= Pluvieux, c_3= Orageux

N=n =12 : effectif total.

2) Tableau statistique

Modalité c_i	n_i	f_i	n_{ic}	f_{ic}
Ensoleillé	3	3/12=0,25	3	3/12=0,25
Orageux	4	4/12= 0,33	7	7/12=0,58
Pluvieux	5	5/12= 0,41	12	1
Total	N=12	1	12	1

Exemple 2

Pour déterminer le type de logement (F2, F3, ….) à construire. On étudie 20 familles selon leur nombre d'enfant, on note les résultats suivants :

1/3/5/5/3/2/4/4/7/0/2/4/3/7/0/5/4/2/3/2

1) Déterminer la population, l'unité statistique (individus), la variable statistique (le caractère), les modalités.
2) Déterminer le tableau statistique.

Solution

1) N=n=20
   La population : les familles
   L'unité statistique : la famille
   La variable statistique : nombre d'enfant
   Les modalités : 0, 1, 2, 3, 4, 5,7

2) Tableau statistique

Modalité c_i	**n_i**	**f_i**	**n_{ic}**	**f_{ic}**
0	2	2/20=0,1	2	0 ,1
1	1	1/20=0,05	3	0,15
2	4	4/20=0,2	7	0,35
3	4	4/20=0,2	11	0,55
4	4	4/20=0,2	15	0,75
5	3	3/20=0,15	18	0,9
7	2	2/20=0,1	20	1
SOMME	20	1	20	1

II.5 Hypothèse fondamentale dans le caractère continu

Dès qu'un caractère est identifier en tant que continu, ces modalités $c_k= [L_k, L_{k+1}]$ sont des intervalles avec :

L_k=borne inférieure, L_{k+1}=borne supérieure

$c_k = x_k = \frac{L_{k+1}+L_k}{2}$: Son centre, amplitude, pas, ou largeur.

Quel pas ou amplitude choisir ?

Le pas $a_i = \frac{Valeur\ maximale - Valeur\ minimale,}{k}, k \in N$

Deux formules sont utilisées pour déterminer nombre de classe pour une série statistique

Formule de Sturge

$k = 1 + 3{,}3 log_{10}(N)$

Formule de Yule

$k = 2{,}5\sqrt[4]{N}$

De ce fait, on peut avoir plusieurs tableaux statistiques selon de nombre de classe

Exemple 1

Dans une étude statistique, on vérifie le poids de 10 zèbres enfermé dans un zoo après leur capture pour voir leur adaptation au nouveau milieu, l'étude nous a donné les valeurs suivantes :

80,5/79/86,6 /72/101,5/115,5/95/120 /122

1) Déterminer, la population, l'individu, la variable statistique, les modalités
2) Déterminer le tableau statistique avec c_i, f_i, n_{ic}, f_{ic} avec 3 classes.
3) Dresser un autre tableau statistique en utilisant la formule de Sturge.

Solution

1) N=n 20
La population : les zèbres du zoo
L'unité statistique : le zèbre
Le caractère : le poids
Type de la variable : quantitatif continu

2) Tableau statistique à 3 classes :

Le pas $a_i = \frac{120-72}{3} = 16.6 \approx 17$

C_i modalités	n_i	f_i	n_{ic}	f_{ic}
[72, 89[	4	0,4	4	0,4
[89, 106[	2	0,2	6	0,6
[106, 123[	4	0,4	10	1
Total	10	1	10	1

3) Tableau statistique avec la méthode de Sturge $k = 1 + 3{,}3log_{10}(N)$

k= nombre de classe, N=10 $\longrightarrow$ k=4,3

On prend k=5 casses

Le pas $a_i = \frac{122-72}{5} = 10$

C_i modalités	n_i	f_i	n_{ic}	f_{ic}

[72,82[	3	0,3	3	0,3
[82, 92[	1	0,1	4	0,4
[92, 102[	2	0,2	6	0,6
[102, 112[	1	0,1	7	0,7
[112, 122[	3	0,3	10	1
Total	10	1	10	1

REMARQUE :

- ✓ Un même caractère statistique (variable), peut avoir plusieurs distributions (tableau) statistique.

III Distribution statistique a un Caractère

III.1 Représentation graphique des séries statistiques

Le graphique est un support visuel qui permet :

- ***La synthèse*** : visualiser d'un seul coup d'œil les principales caractéristiques.
- ***La découverte*** : met en évidence les tendances.
- ***Le contrôle*** : on perçoit mieux les anomalies sur un graphique que sur un tableau.
- ***Recherche des régularités*** : régularité dans le mouvement, répétition du phénomène.

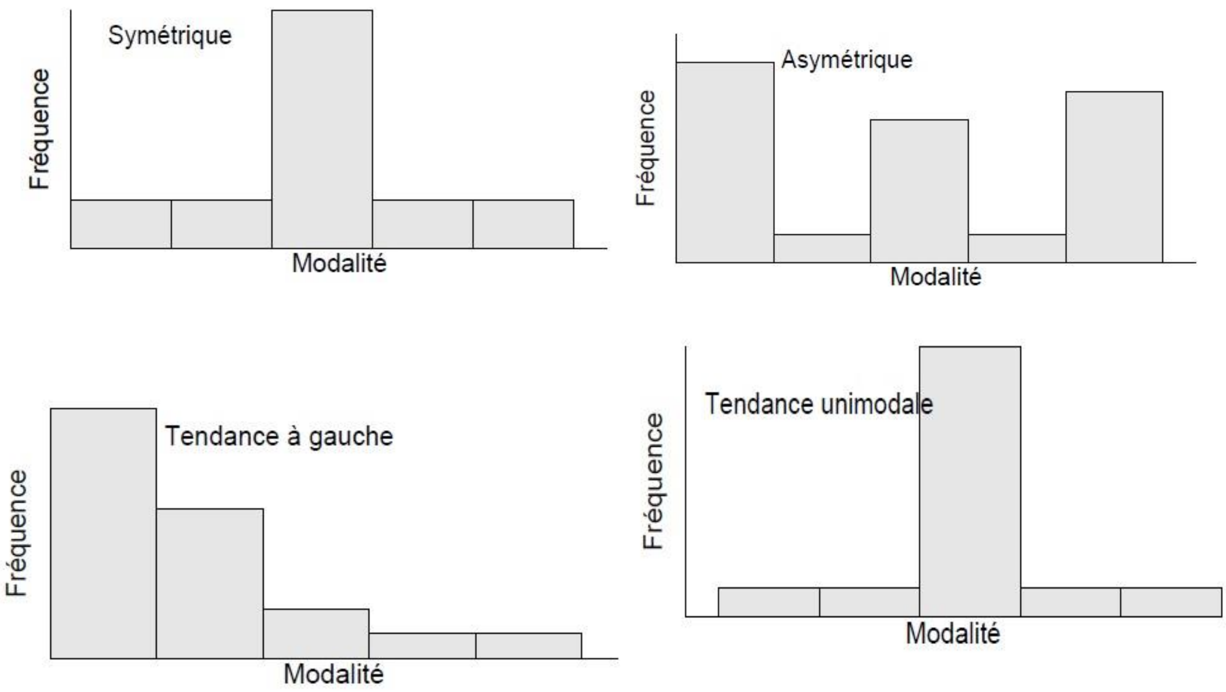


Figure : exemple de représentation graphique

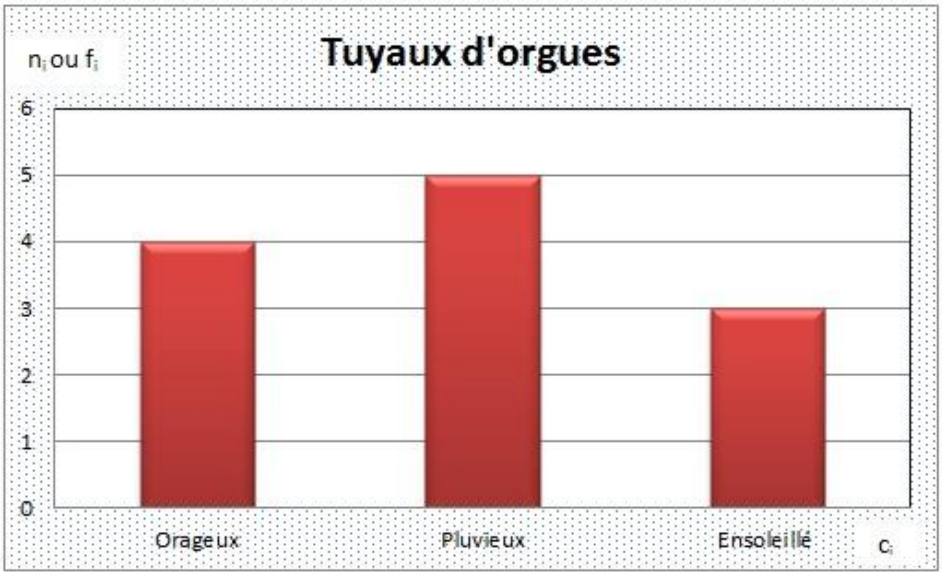
III.2 Distribution à caractère qualitatif

Pour un caractère qualitatif, les effectifs (ou bien les fréquences) peuvent être représentés à l'aide d'un *diagramme à tuyaux d'orgues* ou d'un *diagramme circulaire*.

Les diagrammes circulaires consistent à partager un disque en tranches correspondantes aux modalités observées et dont la surface est proportionnelle à l'effectif (ou la fréquence) de la modalité.

Dans la représentation en diagramme à bandes, les différentes modalités du caractère sont représentées par des segments sur l'axe des ordonnées. Pour chaque abscisse on porte un rectangle dont la longueur est proportionnelle au montant correspondant de l'effectif (ou la fréquence).

Reprenant l'exemple 1 du climat de la région de Tlemcen, les deux représentations possibles de ce caractère qualitatif sont montrées dans la figure :



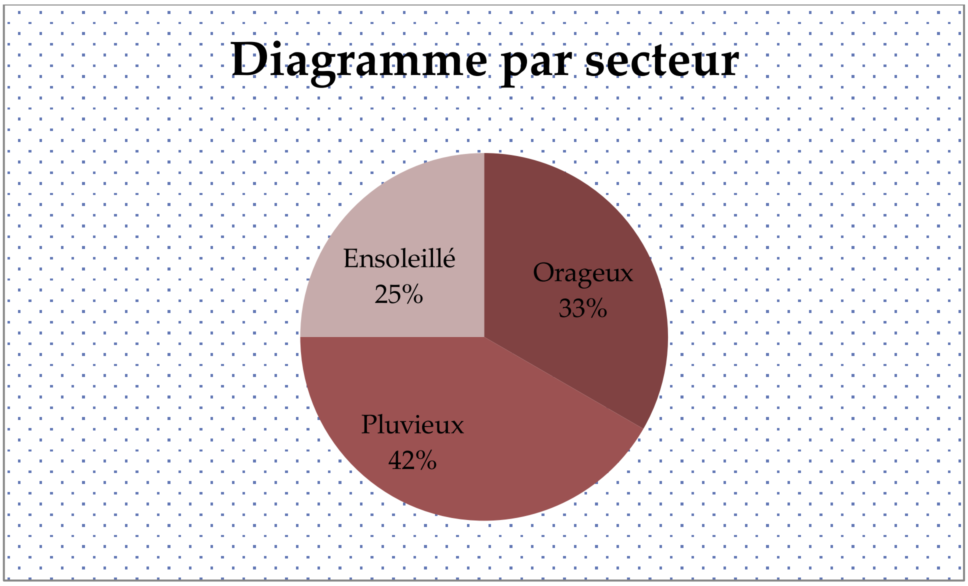


Figure : Distribution à caractère qualitatif

III.3 **Distribution à caractère quantitatif discret**

Dans le cas d'une série discrète, la représentation graphique associée aux effectifs (ou fréquences) est un *diagramme en bâtons*. Il se présente dans un repère orthogonal où figurent sur l'axe des abscisses les valeurs du caractère étudié (les modalités) et sur l'axe des ordonnées les effectifs (ou les fréquences).

On prend l'exemple 2 (nombre d'enfants), deux représentations graphiques sont utilisées pour le caractère discret :

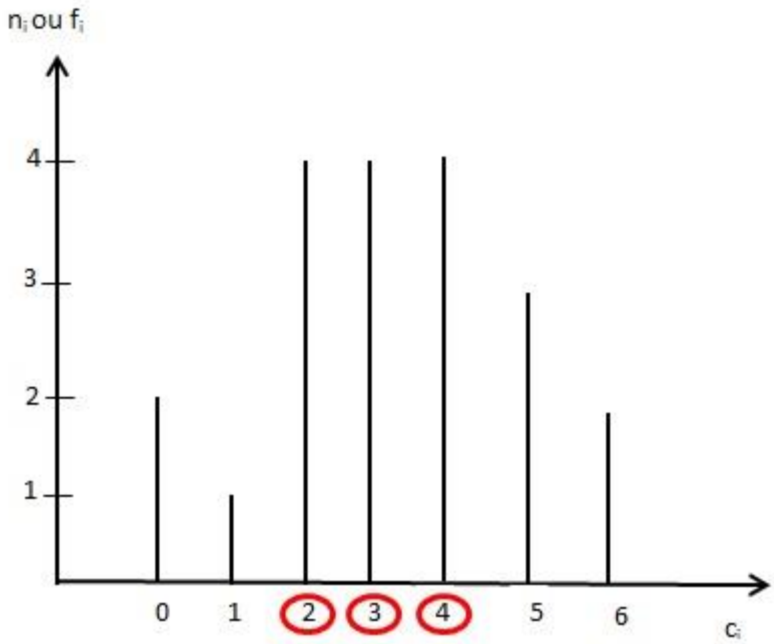


Figure : Diagramme des bâtonnets

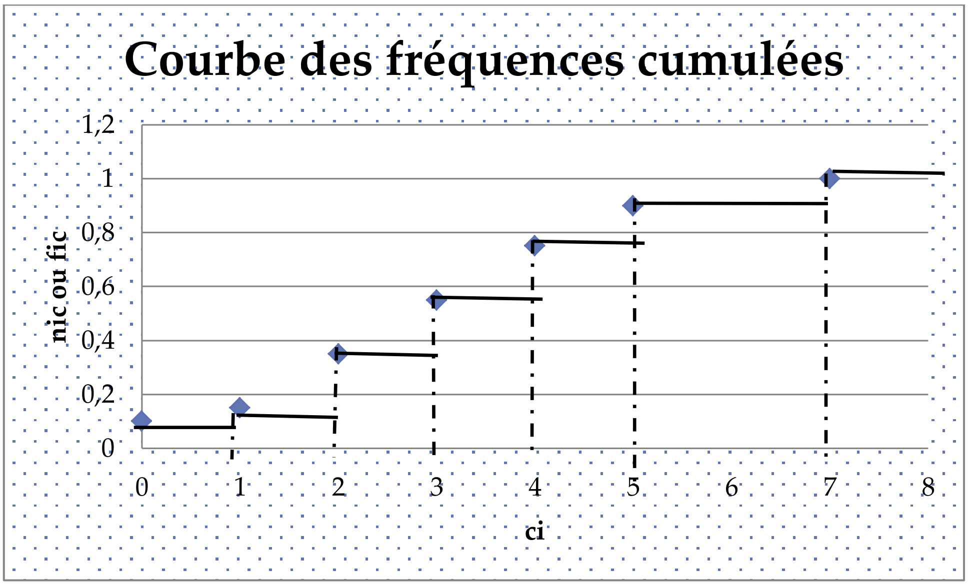


Figure : Courbe des fréquences cumulées

III.4 Distribution à caractère quantitatif continu

Pour représenter une série statistique continue, on utilise *un histogramme* qui est un graphique composé de barres rectangulaires dont la largeur est égale à l'amplitude (le pas 'h') de la classe (c.-à-d. l'intervalle qui représente la modalité) et dont l'aire est proportionnelle à l'effectif (ou la fréquence) de la classe.

Rappelons qu'on fait toujours l'hypothèse que les valeurs observées sont réparties uniformément. L'intérieur de chaque classe et que les classes sont d'amplitudes (c.-à-d. pas 'h') égales. Dans ce cas, il suffit de prendre pour hauteur de chaque rectangle l'effectif (ou la fréquence) de la classe qu'il représente. On prend l'exemple précédent (poids des zèbres) :

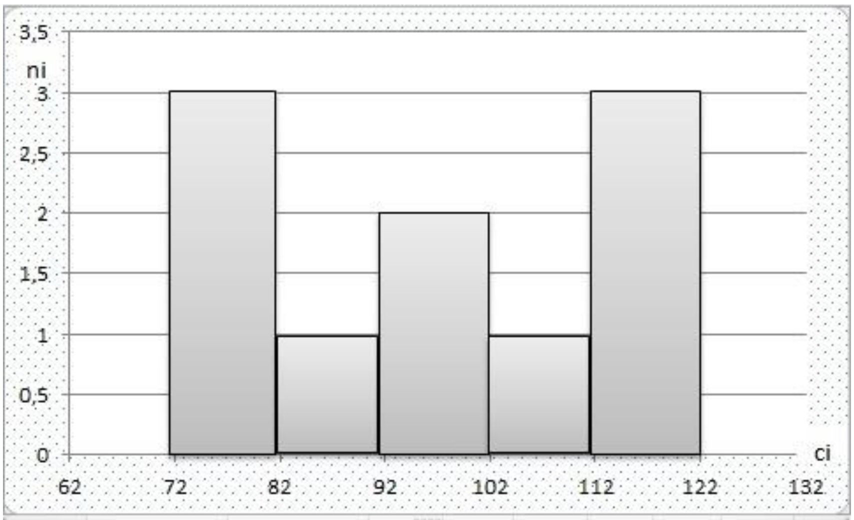


Figure : histogramme des fréquences

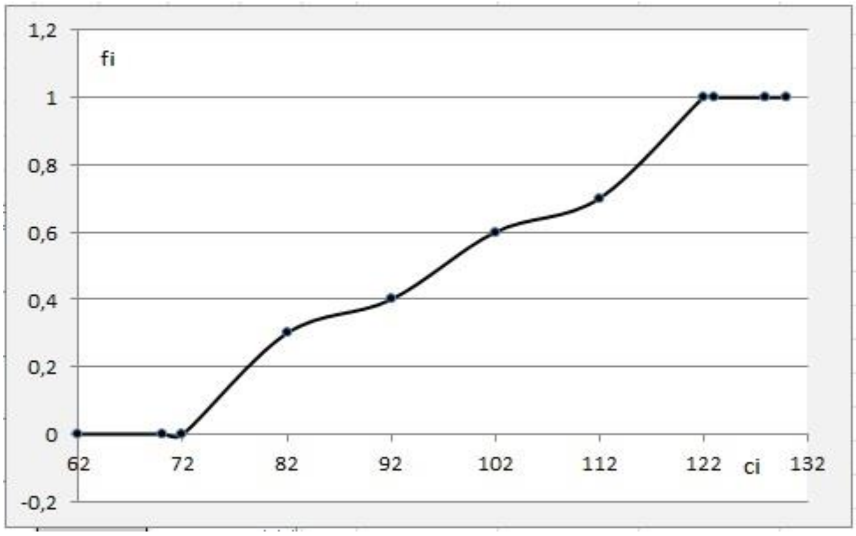


Figure : Courbe des fréquences cumulées

III. 5 Fonction cumulative

Soit X une variable statistique, on définit la fonction de répartition F par :

$$F : R \longrightarrow [0, 1]$$
$$X \longrightarrow F(x)=P(X<x)$$

P : est dite proportion ou fréquence

F(x)=P(X<x) : la proportion des observations qui ne dépasse pas la valeur x.

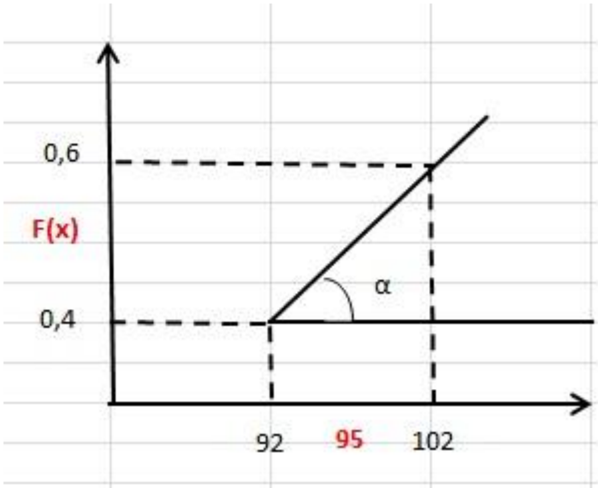
Exemple : pour l'exemple précèdent (cas continu), déterminer la proportion des zèbres ayant un poids inférieure à 102 kg, puis ayant un poids inférieure à 95 kg.

F(x)=F(102)=0,6=60%

- ✓ Donc le pourcentage des zèbres ayant un poids inférieur à 102kg est de 60%, si on veut calculer le nombre des zèbres ayant un poids inférieur à 102 kg :

$$\left.\begin{array}{l} 10 \longrightarrow 100 \\ n \longrightarrow 60 \end{array}\right\} \quad n = \frac{60\times10}{100} = 6\ z\acute{e}bres$$

- ✓ F(x)=F(95)



$$tg = \frac{0{,}6 - 0{,}4}{10} = \frac{F(x) - 0{,}4}{3}$$

F(x)= 46%

Si on veut calculer le nombre des zèbres dont le poids et inférieure à 95kg :

$$\left.\begin{array}{cc} 10 & 100 \\ n & 46 \end{array}\right\} \quad n = \frac{46 \times 10}{100} \approx 5$$

IV Description numérique d'une série statistique

IV.1 Caractéristique de tendance centrale (paramètre de position ou de concentration)

IV.1.1 Définition

Les caractéristiques de tendance centrale décrivent l'ordre de grandeur des valeurs, mais aussi la valeur centrale autour de laquelle se regroupent les observations.

Le but de ses caractéristiques :

- ✓ Réduire l'ensemble de données.
- ✓ Conserver une partie de l'information
- ✓ Résumer l'ensemble des valeurs par une seule valeur
.

IV.1.2 Le Mode

Le mode (m_0) est la valeur de la variable statistique pour laquelle la fréquence ou l'effectif sont les plus élevés (ou plus grands).

Le mode est la valeur la plus fréquente

- ✓ Dans le cas d'une variable statistique continu, on parle de classe modale.
- ✓ Une série statistique peut admettre plus d'un mode ou n'admettre aucun mode.

Exemple :

1) **Exemple 1** : variable qualitative, n_{imax}=5, le mode =Pluvieux.
2) **Exemple 2** : variable discret, n_{imax}=4, le mode m_0=2, m_0=3, m_0=4 (cette série admet 3 classes)
3) **Exemple 3** : variable continu :

Le tableau suivant donne la taille en cm de 30 étudiants :

C_i (cm)	[150, 160[	[160, 170[	[170, 180[	[180, 190[
n_i	6	15	4	5
Fi	0,2	0,5	0,133	0,167
n_{ic}	6	21	25	30
f_{ic}	0,2	0,7	0,833	1

La classe modale : (c'est la classe qui correspond au plus grand effectif ou plus grand fréquence)

N_{imax}= 15 $\longrightarrow$ CM_0= [160, 170[

Dans le cas continu le mode est calculé par la formule suivante :

$$m_0 = L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times a_i$$

L_i : la borne inférieure de la classe modale

a_i : le pas ou amplitude de la classe modale

$\Delta_1 = n_0 - n_1$ ou bien $\Delta_1 = f_0 - f_1$

$\Delta_2 = n_0 - n_2$ ou bien $\Delta_2 = f_0 - f_2$

n_0 ou f_0 : l'effectif ou fréquence de la classe modale

n_1 ou f_1 : l'effectif ou fréquence de la classe modale qui précède CM_0

n_2 ou f_2 : l'effectif ou fréquence de la classe modale qui suit CM_0

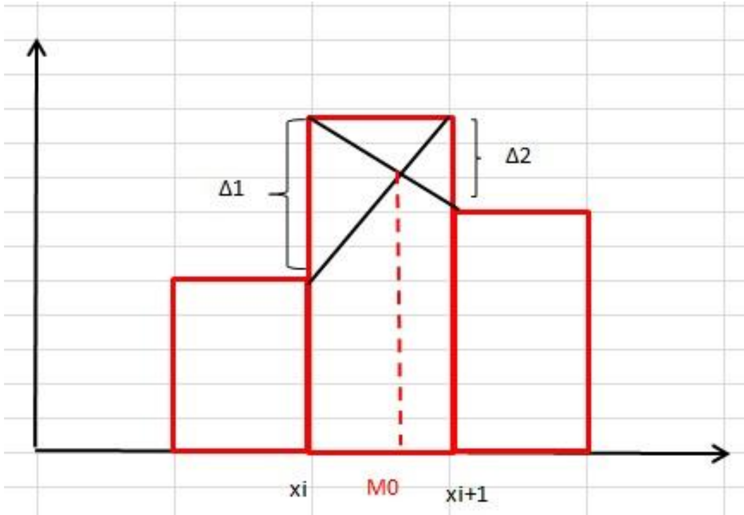
Donc :

$L_i=160$, $a_i=10$, $n_1=6$, $n_2=4$

$\Delta_1 = n_0 - n_1 = 15-6 = 9$ et $\Delta_2 = n_0 - n_2 = 5-4 = 11$

$$m_0 = 160 + \frac{9}{11+9} \times 10 = 164{,}5cm$$

Graphiquement :



IV.1.3 La médiane

La médiane est la valeur de la variable qui partage la série statistique en deux effectifs égaux. Les observations sont préalablement rangées par ordre croissant ou décroissant du caractère.

La médiane est la valeur pour laquelle il y a autant d'individus à gauche qu'à droite dans l'échantillon

La médiane est aussi appelée le deuxième quartile, d'où le premier quartile c'est la valeur de la variable statistique pour laquelle la fonction de répartition vaut 0.25 (F(Q1)=0.25) et le troisième quartile c'est la valeur de la variable statistique pour laquelle la fonction de répartition vaut 0.75 (F(Q3)=0.75).

IV.1.4 La moyenne

Elles se définissent pour les variables statistiques quantitatives. Soit la série x_1, x_2, x_3,…….x_n

La moyenne est la somme des grandeurs mesurées divisée par le nombre d'individus

La moyenne est la somme des grandeurs mesurées divisée par le nombre d'individus

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

En utilisant la lettre grecque Σ pour représenter une somme, on obtient la notation compacte suivante :

$$\overline{X} = \frac{1}{\text{n}} \sum X$$

IV.2 **Les paramètres de dispersion**

IV.2.1 **L'étendue**

L'étendue est la différence entre la plus grande valeur et la plus petite valeur.

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

Dans l'exemple de cas continu, le calcul de l'étendu donne :

122-72=50

IV.2.2 **La variance**

La variance d'une variable statistique caractérise sa capacité à prendre des valeurs plus ou moins éloignées de sa moyenne arithmétique. Elle se calcule comme suit :

$$V(x) = \overline{X^2 - (\bar{X})^2 = \frac{1}{n}\sum_{\iota=1}^{n} X_\iota^2 - (\bar{X})^2}$$

IV.2.3 **L'écart type**

Est une mesure de dispersion de données. Il est défini comme la racine carrée de la variance. Plus l'écart type est faible, plus les valeurs sont regroupées autour de la moyenne

$$\sigma = \sqrt{V(X)} = \frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{X})^2$$

Exercices avec solution

Exercice 1

Les données suivantes représentent le nombre de livres empruntés chaque jour dans une bibliothèque pendant un mois :

60 61 70 68 65 69 67 68 68 67 63 61 64 61 63

70 66 69 60 62 65 66 65 64 60 61 66 68 60 62

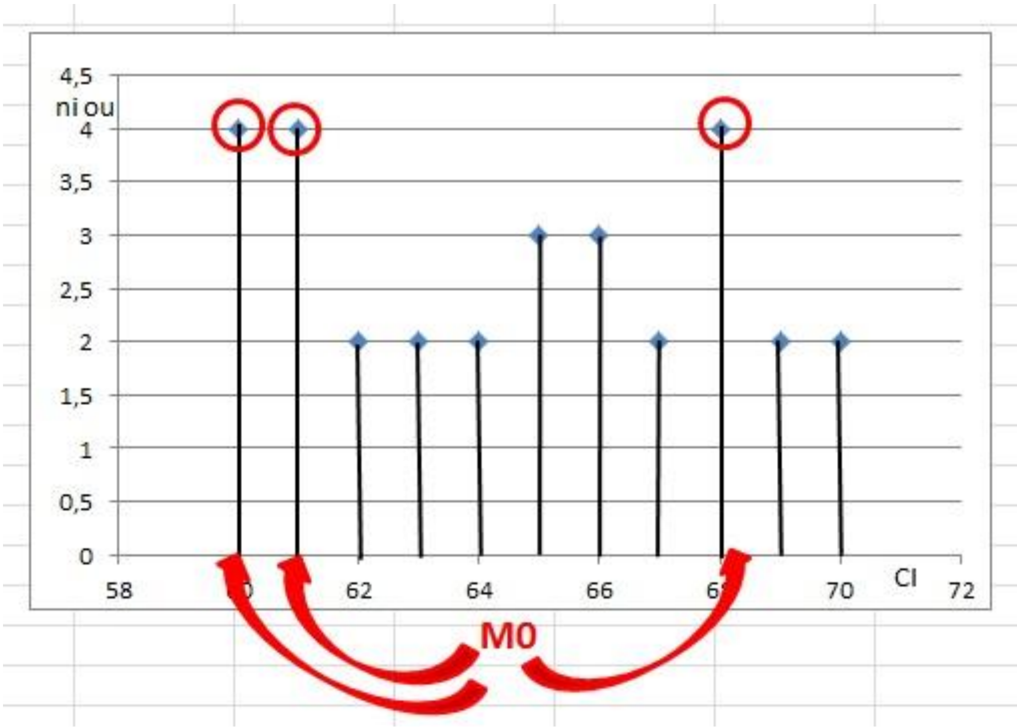
1) Définir la population, l'individu, la variable statistique et son type.
2) Établir le tableau statistique.
3) Déterminer graphiquement le mode. Que représente cette valeur.
4) Déterminer graphiquement la médiane.

Solution

1) Population : 30jours
   Unité statistique : le jour
   Variable statistique : nombre de livre
   Type de la variable : quantitatif discret
2) Tableau statistique

C_i	n_i	f_i	n_{ic}	f_{ic}
60	4	0,133	4	0,133
61	4	0,133	8	0,267
62	2	0,067	10	0,333
63	2	0,067	12	0,4
64	2	0,067	14	0,467
65	3	0,1	17	0,567
66	3	0,1	20	0,667
67	2	0,067	22	0,734
68	4	0,133	26	0,867
69	2	0,067	28	0,934
70	2	0,067	30	1
Total	30	1	30	1

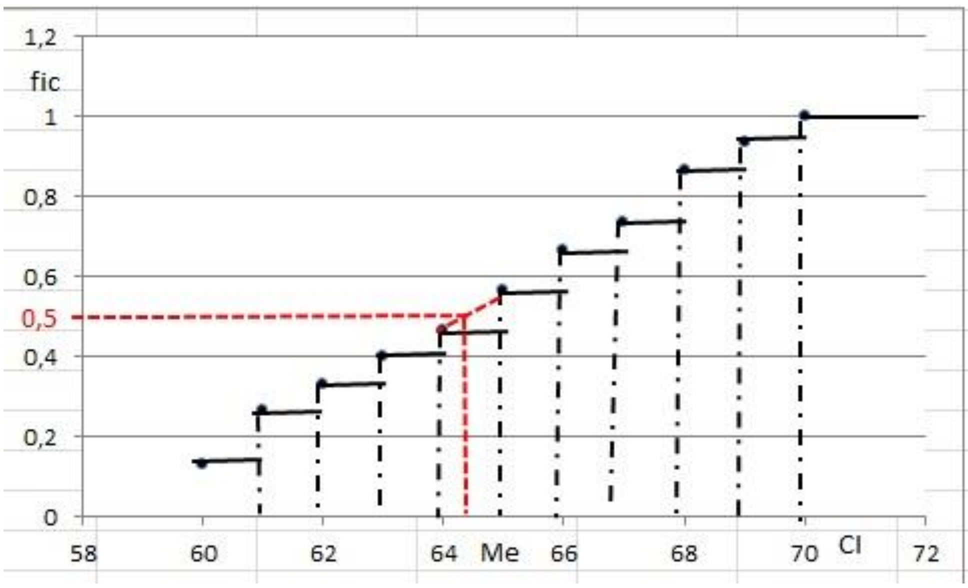
3) Le mode graphiquement



$M_0=60$, $M_0=61$, $M_0=68$

C'est la valeur la plus fréquente.

4) La médiane graphiquement



Si on vous demande de calculer la moyenne, la variance et l'écart type, il vaut mieux pour l'étudiant d'ajouter des colonnes dans le tableau statistique pour faciliter le calcul et gagner du temps pendant l'examen.

c_i	x_i	ni.xi	$ni.x_i^2$
60	60	240	14400
61	61	244	14884
62	62	124	7688
63	63	126	7938
64	64	128	8192
65	65	195	12675
66	66	198	8712
67	67	134	8978
68	68	544	18496
69	69	138	9522
70	70	140	9800
Somme		2211	121285

La moyenne : $\overline{X = \frac{1}{N}\sum n_\iota \times x_\iota =} \frac{1}{30} \times (2211)$=73,7

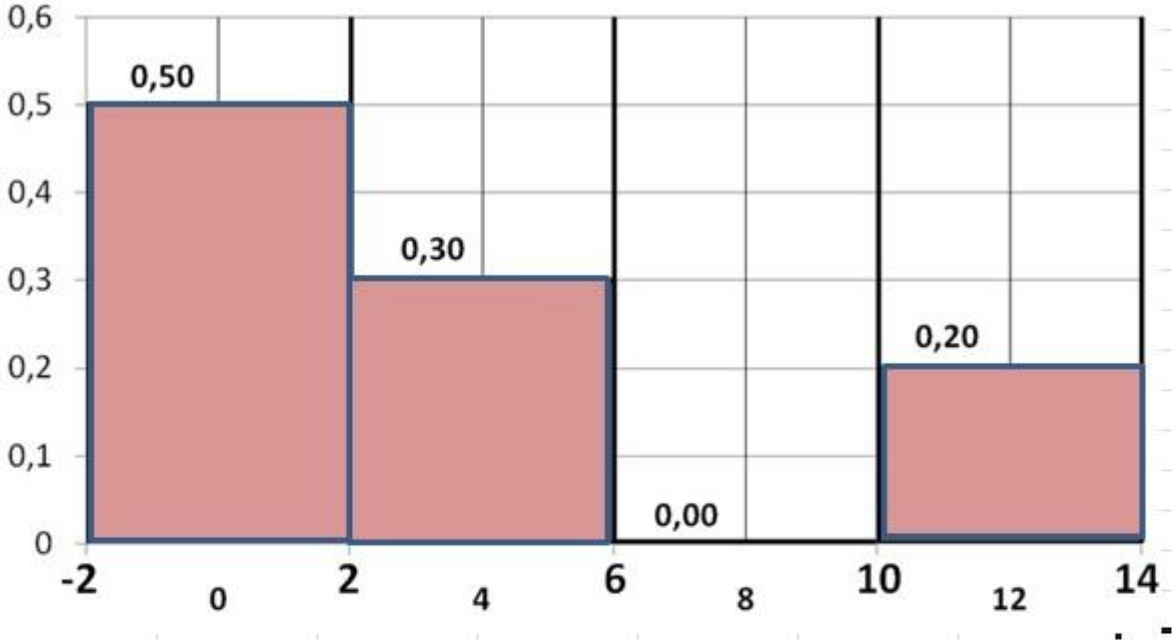
La variance : $V(X) = \frac{1}{N}\sum n_i \times x_i^2 - (\bar{X})^2 == \frac{1}{30} \times 121285 - (37{,}7)^2$=2621,54

L'écart type : $\sigma = \sqrt{V(X)}$=51,2

Exercice 2:

Le traitement de l'information sur un caractère X a permis de dresser son histogramme (voir Figure 1).

1) Déterminer la moyenne de la variable X
2) Déterminer l'écart-type de la variable X
3) Tracer la fonction cumulative
4) Déduire graphiquement la médiane



Solution

1) Tableau statistique

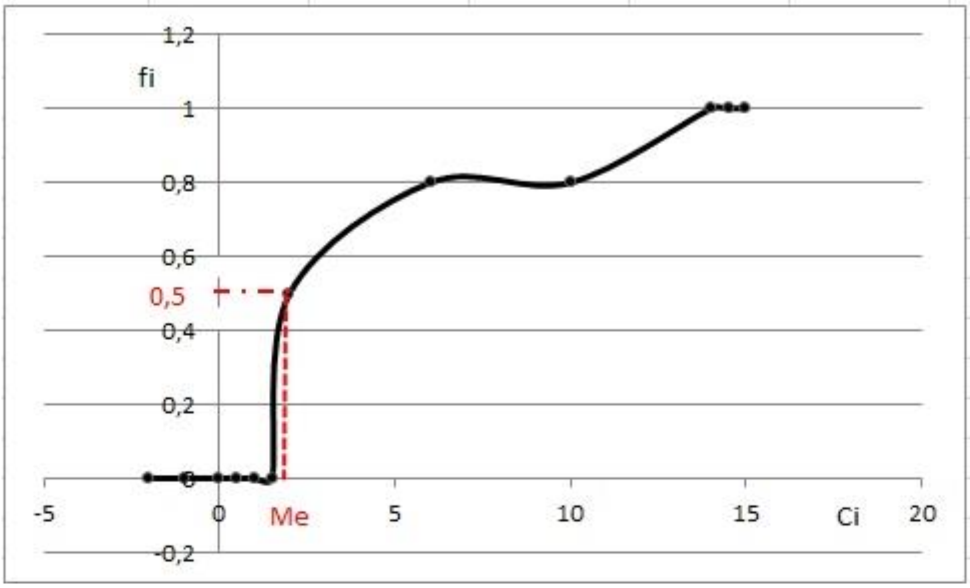
c_i	f_i	f_{ic}	x_i	fi.xi	fi.x_i^2
[-2, 2[	0,5	0,5	0	0	0
[2, 6[	0,3	0,8	4	1,2	4,8
[6, 10[	0	0,8	8	0	0
[10, 14[	0,2	1	12	2,4	28,8

La moyenne : $\overline{X = \sum f_\iota \times x_\iota}$ =3,6

La variance : $V(X) = \sum f_i \times x_i^2 - (\bar{X})^2$ ==33,6$-(3{,}6)^2$=20,64

L'écart type : $\sigma = \sqrt{V(X)}$=4,54

2) La courbe des fréquences cumulées



Exercice 3

On veut déterminer le type de logement (F2, F3,...) à construire pour un ensemble de 50 familles selon leurs salaires. Après l'analyse des données, on note les résultats suivants :

Le type de logement	F1	F2	F3	F4
L'effectif	15	n_2	5	9

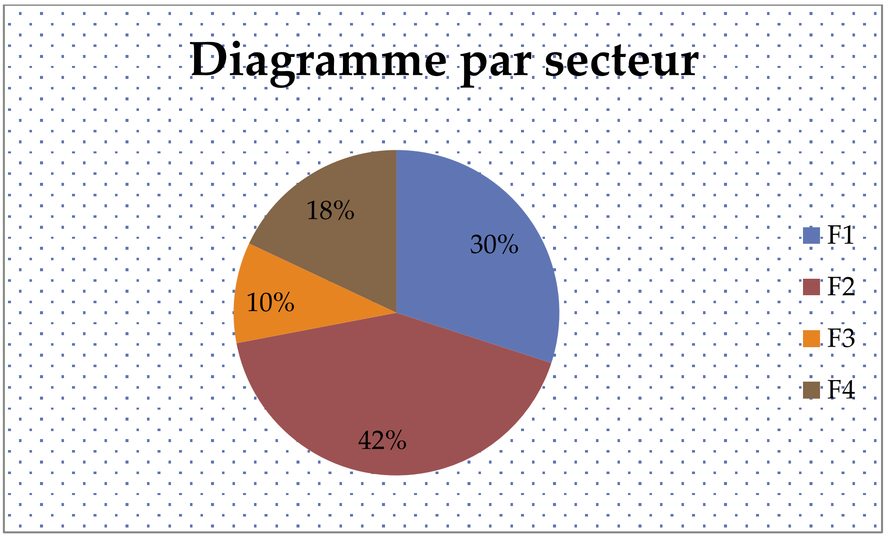
1. Déterminer la variable statistique et son type.
2. Déterminer l'effectif des familles pour lesquelles on construit des logements de type F2.
3. Donner le pourcentage des F4 à construire.
4. Donner toutes les représentations graphiques possibles de cette distribution.

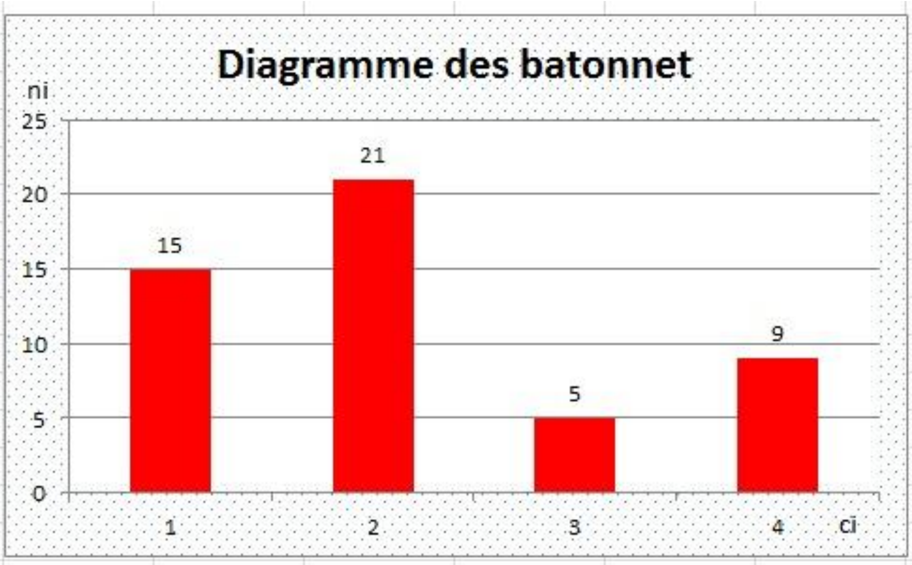
Solution

1) variable statistique : le type de logement
   Type de la variable : qualitatif
   Population : 50 familles
2) $N = \sum n_i$
   $50=15+n_2+5+9$
   Donc $n_2=21$
3) 50 logs $\longrightarrow$ 100%
   9 F4 $\longrightarrow$ x
   $x = \frac{9\times100}{50} = 18\%$*
   On peut aussi calculer le pourcentage des F4 à construire par :

$$f_4 = \frac{N4}{N} = \frac{9}{50} = 18\%$$

4) Les représentations graphiques





Exercice 4

On cherche à modéliser le nombre de collision impliquant deux voitures sur un ensemble de 100 intersections routières choisies au hasard dans la ville de Tlemcen. Des données sont collectées sur une période de un an, et le nombre d'accidents pour chaque intersection est mesuré. On résume ces résultats dans e tableau suivant :

Nombre d'accidents	0	1	2	3	4	5	6	7	8
Nombre d'intersections	7	10	20	25	16	2	7	3	10

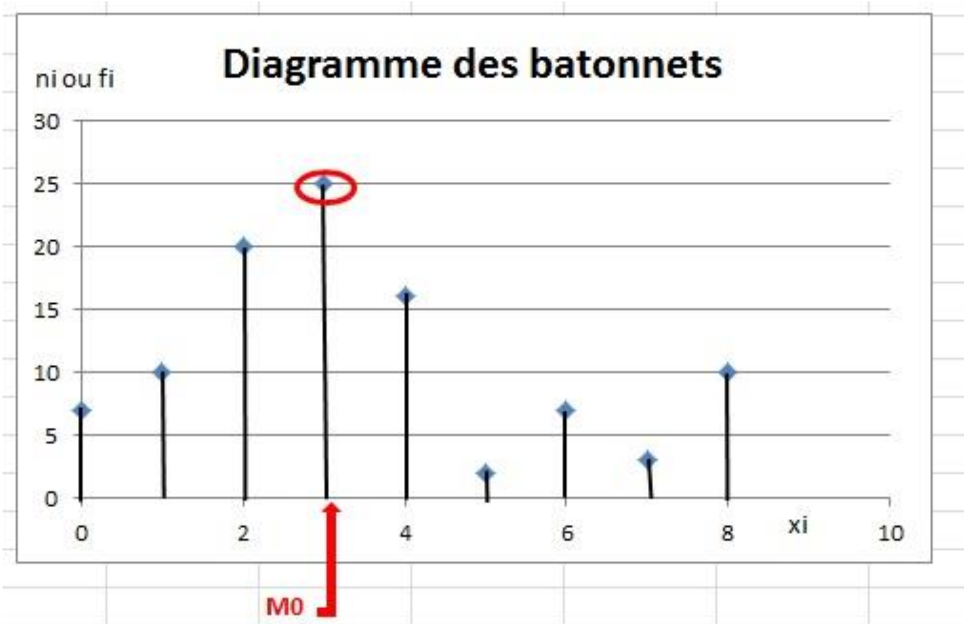
1) Quel type est la variable statistique X étudié ?
2) Déterminer le tableau statistique
3) Tracer le diagramme des bâtonnets et la courbes des fréquences cumulées associé à la distribution statistique X.
4) Soit F la fonction de répartition ou fonction cumulative. Déterminer F. Tracer son graphe.
5) Calculer le mode et la moyenne arithmétique.
6) Déterminer à partir du graphe la valeur de la médiane.
7) Calculer la variance, l'écart type.

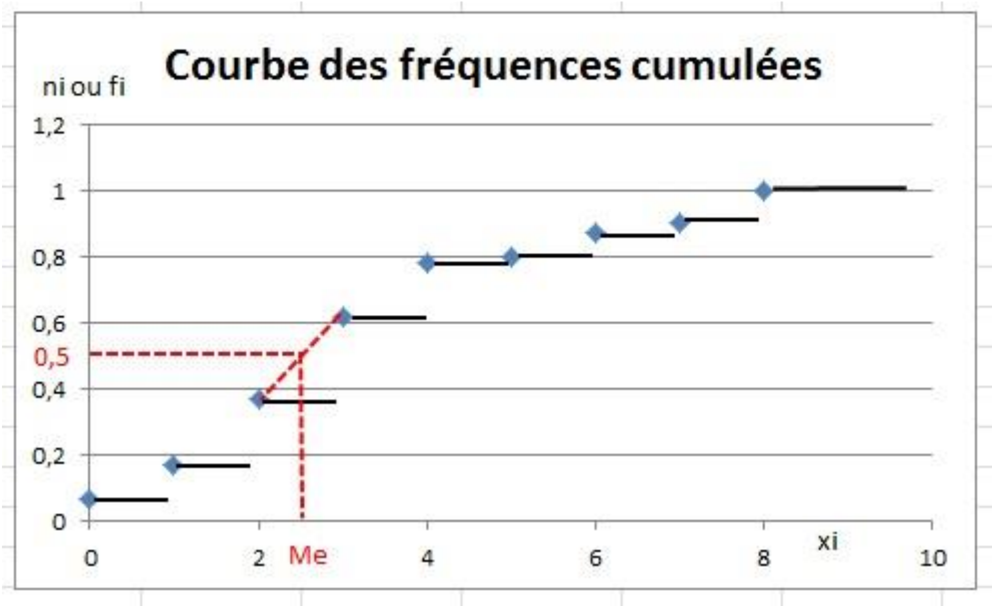
Solution

1) Variable : nombre de collision
   Type : quantitatif discret
2) Tableau statistique

x_i	0	1	2	3	4	5	6	7	8	SOMME
n_i	7	10	20	25	16	2	7	3	10	100
fi	0,07	0,1	0,2	0,25	0,16	0,02	0,07	0,03	0,1	1
n_{ic}	7	17	37	62	78	80	87	90	100	
fic	0,07	0,17	0,37	0,62	0,78	0,8	0,87	0,9	1	
$x_i.n_i$	0	10	40	75	64	10	42	21	80	342
x_i^2	0	1	4	9	16	25	36	49	64	
$n_i.x_i^2$	0	10	80	225	256	50	252	98	640	1611

3)





4) $F : \mathbb{R} \longrightarrow [0, 1]$

$X \longrightarrow F(x)=P(X<x)$

$$F(x) = \begin{cases} 0 & si \quad x < 0 \\ 0{,}07 & si \quad x \leq x < 1 \\ 0{,}17 & si \quad 1 \leq x < 2 \\ 0{,}37 & si \quad 2 \leq x < 3 \\ 0{,}62 & si \quad 3 \leq x < 4 \\ 0{,}78 & si \quad 4 \leq x < 5 \\ 0{,}8 & si \quad 5 \leq x < 6 \\ 0{,}87 & si \quad 6 \leq x < 7 \\ 0{,}9 & si \quad 7 \leq x < 8 \\ 1 & si \quad x \geq 8 \end{cases}$$

5) Le mode : n_{imax}=25 donc M_0=3
6) $\bar{X} = \frac{1}{N}\sum_i n_i \times x_i = \sum_i f_i \times x_i = \frac{1}{100}(342) = 3{,}42$
7) La médiane : voir le graphe
8) Variance et écart type :

$$V(X) = \frac{1}{N}\sum n_i \times x_i^2 - (\bar{X})^2 = \frac{1}{100} \times 1611 - (3{,}42)^2 = 4{,}41$$

$$\sigma(x) = \sqrt{V(x)} = 2{,}1$$

Examen final 2017 /2018

Exercice 5 / (6 pts) / On mesure les diamètres de troncs d'arbres d'une même espèce. On étudie 400 échantillons. On obtient les résultats suivants:

Diamètre en cm	25	26	27	28	29	30
Pourcentage	10%	15%	30%	35%	5%	5%

On donne:

$$(25 \times 0.1) + (26 \times 0.15) + (27 \times 0.3) + (28 \times 0.35) + (29 \times 0.05) + (30 \times 0.05) = 27.25.$$
$$(25^2 \times 0.1) + (26^2 \times 0.15) + (27^2 \times 0.3) + (28^2 \times 0.35) + (29^2 \times 0.05) + (30^2 \times 0.05) = 744.05.$$

1. Établir le tableau statistique en fonction des effectifs et des fréquences relatives.
2. Quel est le diamètre moyen de ces troncs d'arbres?
3. Déterminer la variance puis l'écart-type de la série statistique résumée dans le tableau ci-dessus.
4. Représenter graphiquement ces résultats (juste pour les effectifs).
5. Déterminer le mode et donner son interprétation.

Solution:

1. **Tableau statistique: (1.5 pts en total)**

Modalités x_i	Effectifs n_i	Fréquences f_i
25	40	0.1
26	60	0.15
27	120	0.3
28	140	0.35
29	20	0.05
30	20	0.05
Total	400	1

2. **Diamètre moyen: (0.5 pts en total)**

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{6} n_i x_i = \sum_{i=1}^{6} f_i x_i \quad \textbf{(0.25 pts)}$$
$$= 27.25. \quad \textbf{(0.25 pts)}$$

Donc le diamètre moyen de ces troncs d'arbres est de 27.25 cm.

3. **Variance et écart-type: (1.5 pts en total)**

$$\overline{x^2} = \frac{1}{N}\sum_{i=1}^{6} n_i x_i^2 = \sum_{i=1}^{6} f_i x_i^2 \quad \textbf{(0.25 pts)}$$
$$= 744.05 \quad \textbf{(0.25 pts)}$$
$$Var(x) = \overline{x^2} - \overline{x}^2 \quad \textbf{(0.25 pts)}$$
$$\cong 1.49 \quad \textbf{(0.25 pts)}$$
$$\sigma(x) = \sqrt{Var(x)} \quad \textbf{(0.25 pts)}$$
$$\cong 1.22. \quad \textbf{(0.25 pts)}$$

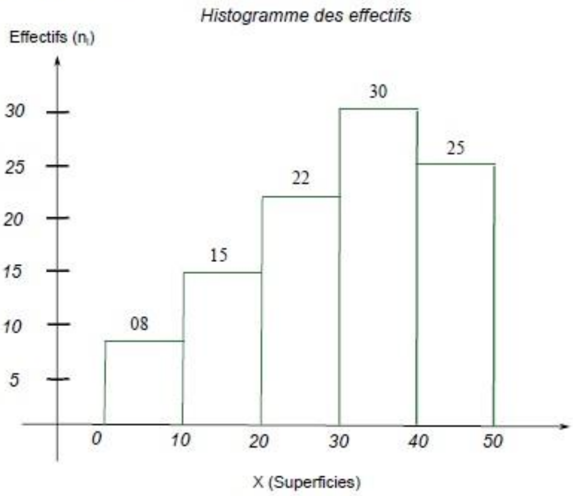
4. **Représentation graphique: (1.5 pts en total)**

5. **Mode et interprétation: (1 pt en total)**

Le mode est 28 cm **(0.5 pts)** car c'est la modalité qui correspond au plus grand effectif (ou fréquence). On peut dire alors que la plus part des troncs d'arbres ont un diamètre de 28 cm. **(0.5 pts)**

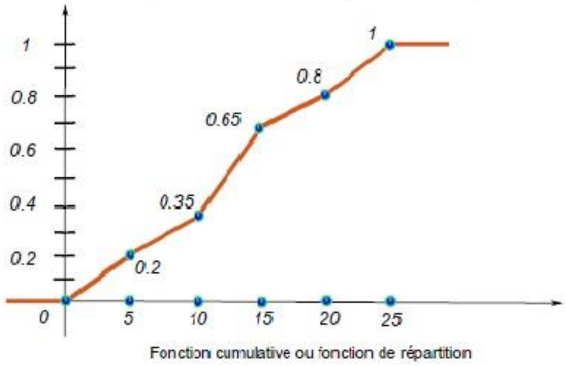
Exercice supplémentaire

Exercice 1 - La répartition d'un échantillon de N exploitations agricoles selon leurs superficies en Hectare se présente par le graphe suivant :



1. Déterminer l'effectif total N, la variable statistique et son type.
2. Quelle est la superficie la plus fréquente ? Justifier votre réponse.
3. Tracer la courbe de la fonction cumulative.
4. Trouver le premier, le deuxième et le troisième quartiles graphiquement et par le calcul. Que représentent ses valeurs ?
5. Calculer l'étendue, la moyenne arithmétique, la variance et l'écart-type de cette série statistique.

Exercice 2 - Le traitement de l'information sur un caractère statistique X a permit de dresser sa fonction de répartition (courbe des fréquences cumulées) dans la figure ci-dessous.



1. De quel type est le caractère statistique X ?
2. Déduire graphiquement la médiane puis trouver sa valeur par le calcul.
3. Dresser le tableau statistique.
4. Tracer l'histogramme du caractère X et calculer le mode.
5. Calculer la moyenne arithmétique et l'écart type.

Série statistique à deux variable

Module : probabilité et statistique

Ce cours "Série statistique à deux variable" vise à définir une série statistique à deux variables, Présenter les différents types de tableau statistique, Transférer un tableau statistique en un nuage de point, calculer les différents paramètres de la droite de corrélation, et assembler toute les informations obtenues pour tirer des conclusions. À la fin de ce cours, l'apprenant sera capable de faire une expérience statistique de son choix même en dehors de sa spécialité.

Série statistique à deux variable

Module : probabilité et statistique

1) Introduction

Un même individu peut être étudié à l'aide de plusieurs caractères, par exemple les salariés (ancienneté et niveau d'étude), une maison (surface et nombre de pièce), croissance d'enfant (taille et poids). Cette analyse est une introduction à l'étude globale des relations entre deux ou trois variables

Les séries statistiques à deux variables peuvent être présentées de deux façons :

a) Données brute à chaque unité on associée les données (x_i, y_i)

individu	1	2	....	x_i	....	n
X	x_1	x_2		x_i		x_n
Y	y_1	y_2		y_i		y_n

b) Tableau à double entrée (tableau de contingence)

	$C_1' = [L_1', L_2'[/y_1$	$C_2' = [L_2', L_3'[/y_2$	....	$C_m' = [L_{m-1}', L_m'[/y_m$	$\sum$
$C_1' = [L_1', L_2'[/x_1$	n_{11} ou f_{11}	n_{12} ou f_{12}		n_{1m} ou f_{1m}	$f_{1.}$ ou $n_{1.}$
$C_1' = [L_1', L_2'[/x_2$				n_{2m} ou f_{2m}	$f_{2.}$ ou $n_{2.}$
. . . . .					
$C_n' = [L_{n-1}', L_n'[/x_k$	n_{k1} ou f_{k1}	n_{k2} ou f_{k2}		n_{km} ou f_{km}	$f_{k.}$ ou $n_{k.}$
$\sum$	$n_{.1}$ ou $f_{.1}$	$n_{.2}$ ou $f_{.2}$		$n_{.m}$ ou $f_{.m}$	N

A chaque couple (x_i, y_i) correspond n_{ij} effectif, regroupant les observations qui prennent la valeur de xi de la variable x et de la valeur y_i de la variable y.

f_{ij} est la fréquence du couple (x_i, y_i) : $f_{ij} = \frac{n_{ij}}{N}$ ou N : effectif total

Avec : $N = \sum_{i=1}^{k} \sum_{j=1}^{m} n_{ij}$ et $\sum_{i=1}^{k} \sum_{j=1}^{m} f_{ij} = 1 = \sum_{j=1}^{m} \sum_{i=1}^{k} n_{ij}$

Remarque

- ✓ La première présentation découle de la collecte des informations brutes alors que la deuxième c'est un classement de ces informations.

En marge du tableau (colonne et ligne) on a :

1.1 Effectifs et fréquences marginale X

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

$n_{i.} = n_{i1} + n_{i2} + \cdots \dots \dots \dots \dots \dots .. + n_{im}$ (Somme par colonne)

1.2 Effectifs et fréquences marginale Y

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

$n_{.j} = n_{1j} + n_{2j} + \cdots \dots \dots \dots \dots \dots .. + n_{kj}$ (Somme par ligne)

2 Description numérique

2.1 Caractéristiques des séries marginales

Les formules de définitions de ces paramètres sont identiques à celle qui ont été définies pour les séries à une variables, seule la notion ses effectifs diffère.

$$\bar{X} = \frac{1}{N} \sum_i n_{i.} \times x_i = \sum_i f_{i.} \times x_i$$

$$\bar{Y} = \frac{1}{N} \sum_j n_{.j} \times y_j = \sum_j f_{.j} \times x_j$$

$$V(X) = \sum_i f_{i.}(x_i - \bar{X})^2 = \overline{X^2} - (\bar{X})^2 \qquad avec \qquad \overline{X^2} = \frac{1}{N} \sum_i n_{i.} \times x_i^2 = \sum_i f_{i.} \times x_i$$

$$V(Y) = \sum_j f_{.j}(y_j - \bar{Y})^2 = \overline{Y^2} - (\bar{Y})^2 \qquad avec \qquad \overline{Y^2} = \frac{1}{N} \sum_i n_{.j} \times y_j^2 = \sum_j f_{.j} \times y_j$$

Donc : $\sigma_X = \sqrt{V(X)},\ \sigma_Y = \sqrt{V(Y)}$

2.2 Séries conditionnelles

La notion de série conditionne est essentielle pour comprendre l'analyse de la régression. Un tableau de contingence se décompose en autant de séries conditionnelles qu'il contient de lignes et de colonnes.

2.2.1 séries conditionnelles de la variance X : (X/y_j) ou X_j

(X/y_j) est la notation de la série conditionnelle de X sachant que $Y=y_j$

La fréquence conditionnelle $f_{i/j}$ (f_i sachant j) est définie par :

$$f_{i/j} = \frac{n_{ij}}{n_{.j}} = \frac{f_{ij}}{f_{.j}} \qquad avec \qquad \sum_i f_{i.j} = 1$$

La moyenne arithmétique ou moyenne conditionnelle $\overline{X_J}$ (la moyenne des valeurs de X sous la condition y_j) est définie par :

$$\overline{X_J} = \sum_i f_{i/j} \times x_i = \frac{1}{n_{.j}} \sum_i n_{ij} \times x_i$$

Pour la variance conditionnelle

$$V_{xj} = \sum_i f_{i/j}(x_i - \overline{X_J})^2 = \overline{X_J^2} - (\overline{X_J})^2$$

2.2.2 séries conditionnelles de la variance Y : (Y/x_j) ou Y_j

(Y/x_j) est la notation de la série conditionnelle de Y sachant que $X=x_i$

La fréquence conditionnelle $f_{j/i}$ (f_j sachant i) est définie par :

$$f_{j/i} = \frac{n_{ij}}{n_{i.}} = \frac{f_{ij}}{f_{i.}} \qquad avec \qquad \sum_i f_{j/i} = 1$$

La moyenne arithmétique ou moyenne conditionnelle $\overline{Y_\iota}$ (la moyenne des valeurs de Y sous la condition x_i) est définie par :$\overline{Y_\iota} = \sum_j f_{j/i} \times y_j = \frac{1}{n_{i.}} \sum_i n_{ij} \times y_j$

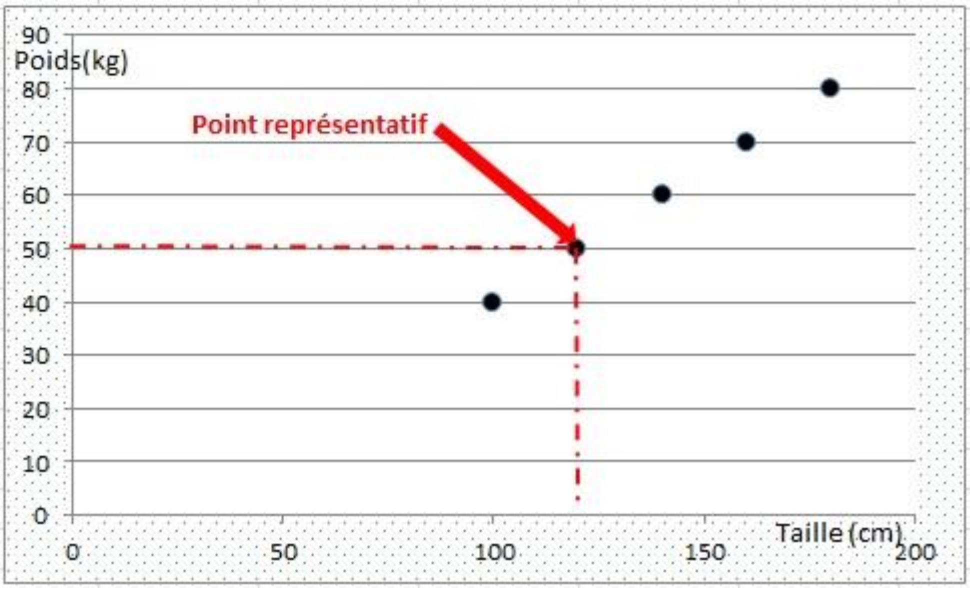
Pour la variance conditionnelle : $V_{yi} = \sum_i f_{j/i}(y_i - \overline{Y_\iota})^2 = \overline{Y_\iota^2} - (\overline{Y_\iota})^2$

3 Corrélation (Méthode des moindres carrées)

Dans le chapitre 1, nous nous sommes intéressés à des relations entre une seule variable, par exemple la note des étudiants, le poids ou la taille des individus. Il existe d'autres problèmes statistiques ou l'on s'intéresse à la relation entre plusieurs variables. Par exemple la relation être le poids et la taille, le revenu annuel et le niveau d'étude…etc.

Pour étudier les relations ou les corrélations entre deux variables, il est plus judicieux de les porter sur un graphique.

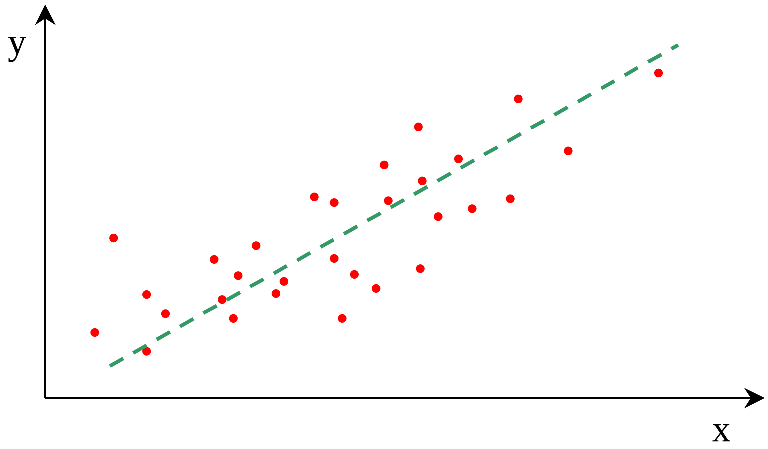
Exemple : si on étudie la relation entre la taille et le poids des étudiants. Donc, pour chaque individu on note sur le graphe : la taille en abscisse et le poids en ordonnée, il sera représenter par un point (point représentatif).



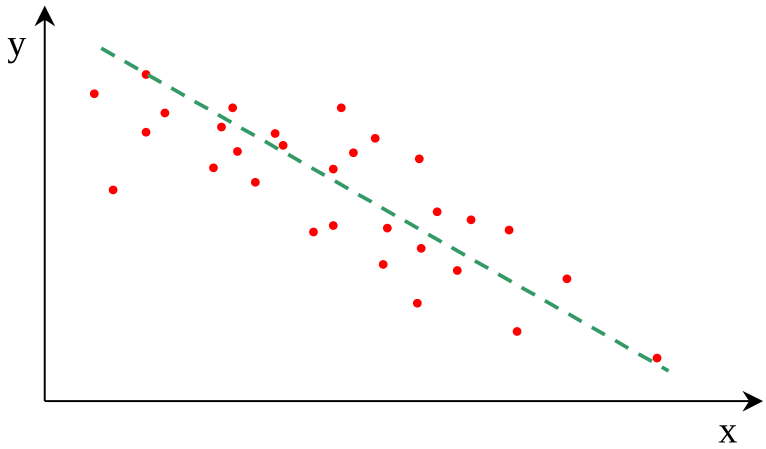
Dans le graphe, il existe plusieurs points en fonctions du nombre d'individus dans l'échantillon.

On peut (par la pensée ou réellement) tracer une droite qui passe au mieux par ces points (au milieu du "nuage" de points).

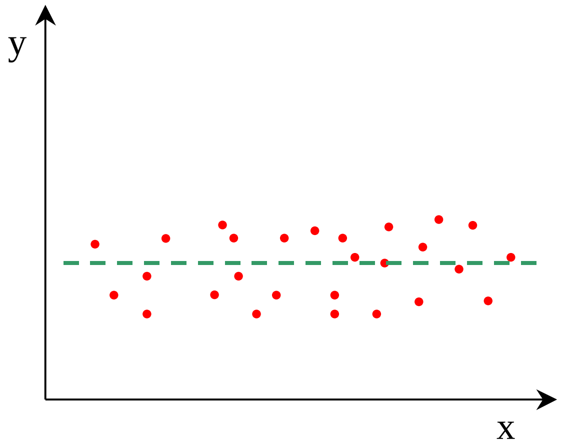
<u>Corrélation positive:</u>

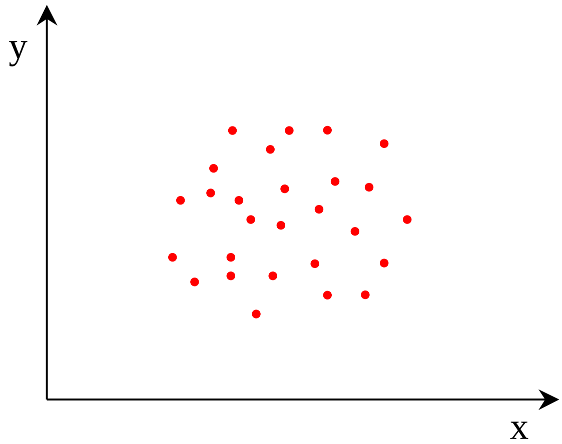


Corrélation négative:

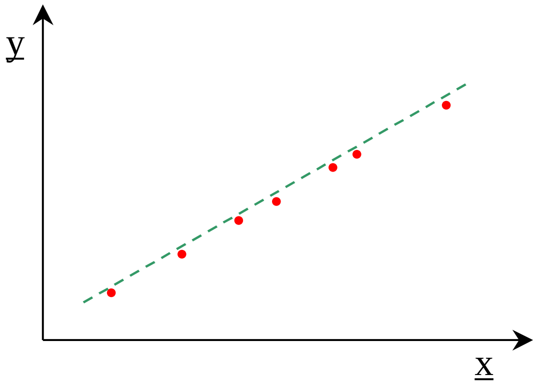


Absence de corrélation:

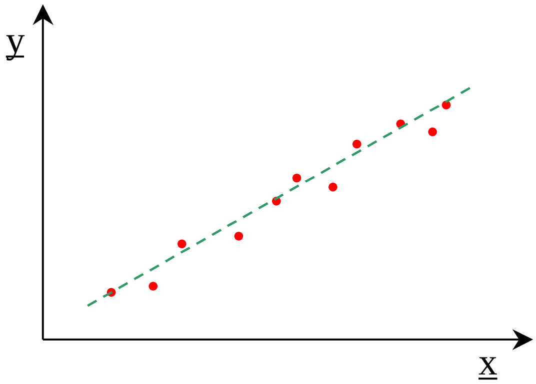




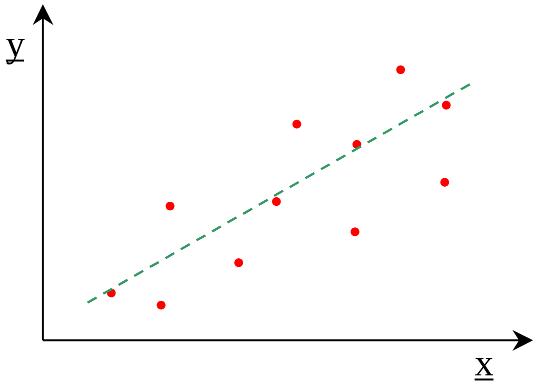
Corrélation parfaite:



Bonne corrélation (corrélation forte):



Mauvaise corrélation (corrélation faible):



3.1 Méthode des moindres carrées

La méthode des moindres carrées est utilisées pour déterminer la meilleur droite qui minimise les carrées des écarts des points représentatifs à cette droite. L'équation de la droite de régression cherchée est de forme :

$$Y=a\ X+b$$

Les coefficients a et b peuvent être calculés à partir des formules suivantes :

$$a = \frac{\sum (X - \overline{X}).(Y - \overline{Y})}{\sum (X - \overline{X})^2}$$

Ordonnée à l'origine: $b = \bar{Y} - a\,\bar{X}$

3.1.1 Coefficient de corrélation

Le signe de la pente a, donne le *sens* de corrélation, mais pas sa *qualité*.

$a > 0$ corrélation positive

$a < 0$ corrélation négative

$a = 0$ pas de corrélation

La qualité de la corrélation peut être mesurée par un *coefficient de corrélation r*

$$r = \frac{\sum (X - \overline{X}).(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2} \times \sqrt{\sum (Y - \overline{Y})^2}}$$

Le coefficient de corrélation est compris entre –1 et +1.

Plus il s'éloigne de zéro, meilleure est la corrélation

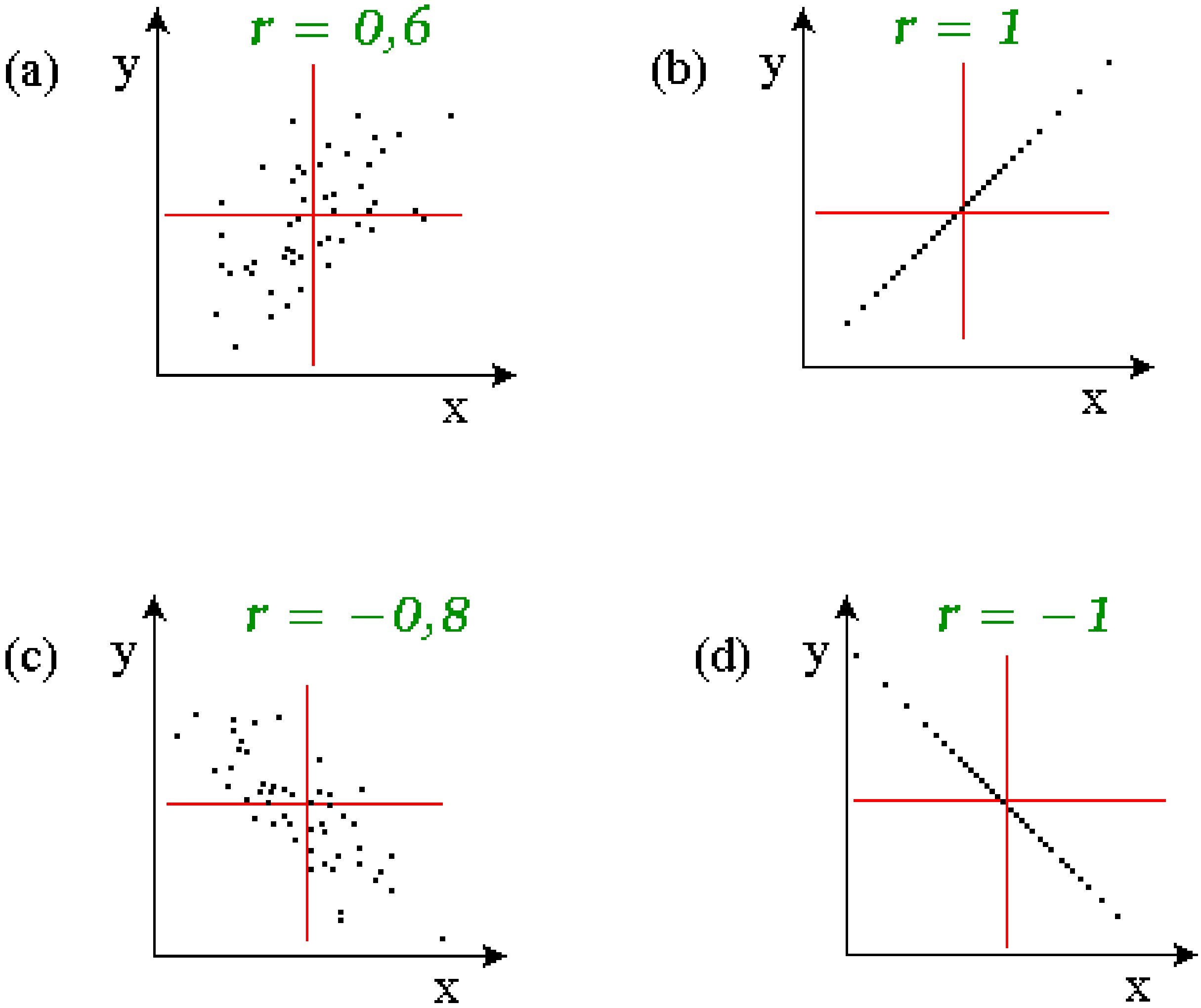
$r = +1$ corrélation positive parfaite

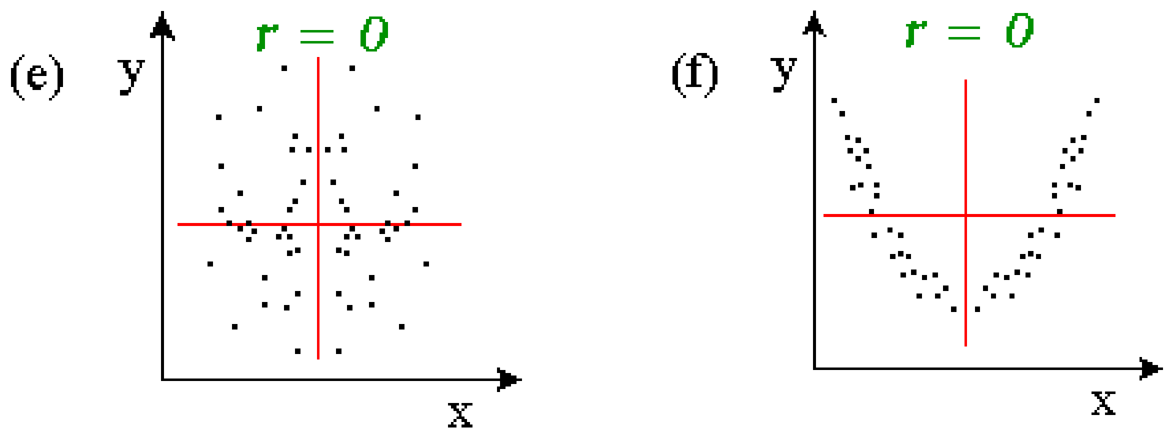
$r = -1$ corrélation négative parfaite

$r = 0$ absence totale de corrélation

Quelques exemples de corrélation

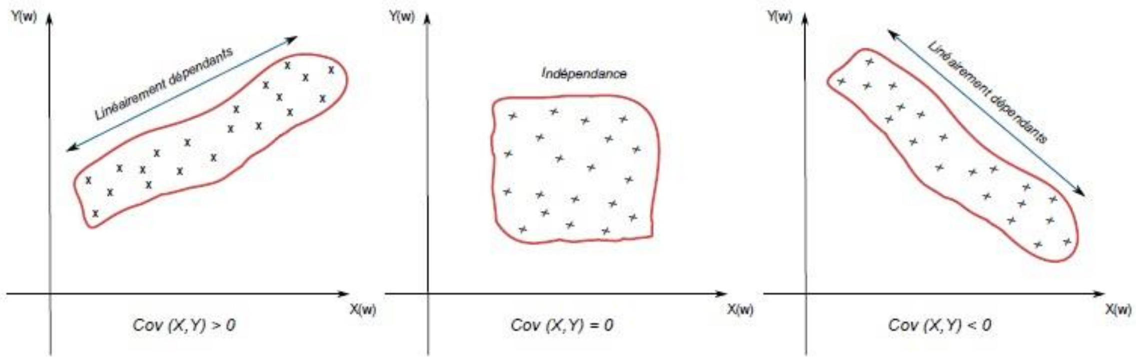
(Le coefficient de corrélation *r* est indiqué dans chaque cas)


(a)
y
r = 0,6
x
(b)
y
r = 1
x
(c)
y
r = −0,8
x
(d)
y
r = −1
x



4 Notion ce covariance

Nous notons par $Cov(X,Y)$ la covariance entre les variables X et Y. La covariance est un paramètre qui donne la variabilité de X par rapport à Y



La covariance se calcule par l'expression suivante

$$Cov(X,Y) = \overline{xy} - \overline{x}\ \overline{y} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - \overline{x}\ \overline{y}.$$

Nous avons aussi cette formule :

$$COV(X,Y) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \bar{x})(y_j - \bar{y})$$

Remarque :

On dit que deux variables statistiques X et Y sont indépendantes si et seulement si, pour tout i et j :

$$f_{ij} = f_{i\bullet} \times f_{\bullet j}$$

Il suffit que cette égalité ne soit pas vérifiée dans une seule cellule pour que les deux variables ne soient pas indépendantes. De manière équivalente, pour tout i et j,

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

Dans ce cas, si X et Y sont indépendantes alors (réciproque est fausse) Cov(X, Y) = 0.

Exercice 1: examen (2014/2015)

Voici les tailles en centimètres et poids en kilogramme (y) de 1à enfants de 6 ans :

Enfant	1	2	3	4	5	6	7	8	9	10
Taille x	121	123	108	118	111	109	114	103	110	115
Poids y	25	22	19	24	19	18	20	15	20	21

1) Calculer la covariance et le coefficient de corrélation.
2) Déterminer la droite de régression linéaire de y par rapport à x.

Rappel

Droite de corrélation linéaire : $Y = \bar{Y} - \frac{COV(X,Y)}{\sigma_X^2}\bar{X} + \frac{COV(X,Y)}{\sigma_X^2}X$

Coefficient de corrélation : $R = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$

Covariance: $COV(X,Y) = \sum_{i=1}^{n}\frac{1}{n}(x_i - \bar{X})(y_i - \bar{Y})$

Solution

Pour répondre aux questions il est pratique de remplir d'abord le tableau suivant :

individu	1	2	3	4	5	6	7	8	9	10	∑
X	121	123	108	118	111	109	114	103	110	115	1132
Y	25	22	19	24	19	18	20	15	20	21	203
X^2	14641	15129	11664	13924	12321	11881	12996	10609	12100	13225	128490
Y^2	625	484	361	576	361	324	400	225	400	441	4197
X.Y	3025	2706	2052	2832	2109	1962	2280	1545	2200	2415	23126

1) La moyenne arithmétique de la variable statistique X est donnée par la formule :

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1132}{10} = 113{,}2$$

2) La moyenne arithmétique de la variable statistique Y est donnée par la formule :

$$\bar{Y} = \frac{1}{10}\sum_{i=1}^{10} y_i = \frac{203}{10} = 20{,}3$$

3) L'écart type de la variable X est de :

$$\sigma_X = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(X_i - \bar{X})^2} = \sqrt{\overline{X^2} - (\bar{X})^2} = \sqrt{\frac{128490}{10} - (113{,}2)^2} = 5{,}9$$

4) L'écart type de la variable Y est de :

$$\sigma_Y = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(Y_i - \bar{Y})^2} = \sqrt{\overline{Y^2} - (Y)^2} = \sqrt{\frac{4197}{10} - (20{,}3)^2} = 2{,}76$$

5) La covariance de a variable X et Y :

$$COV(X,Y) = \frac{1}{10}\sum_{i=1}^{10}(X_i - \bar{X})(Y_i - \bar{Y}) = \frac{1}{10}\sum_{i=1}^{10}(X_i Y_i) - \bar{X}\bar{Y} = 14{,}64$$

6) Les paramètres a et b de la droite de corrélation linéaire y=aX+b sont donnée par :

$$a = \frac{COV(X,Y)}{\sigma_X^2} = \frac{14{,}64}{34{,}81} = 0{,}42$$

$$b = \bar{Y} - a\bar{X} = -27{,}24$$

Donc la droite de régression s'écrit : Y=0,42 X -27,24

Le coefficient de sécurité : $R = \frac{COV(X,Y)}{\sigma_X \sigma_Y} = 0{,}89$

Le coefficient de corrélation est proche de 1, on peut conclure que les deux variables sont bien corrélées.

Exercice2

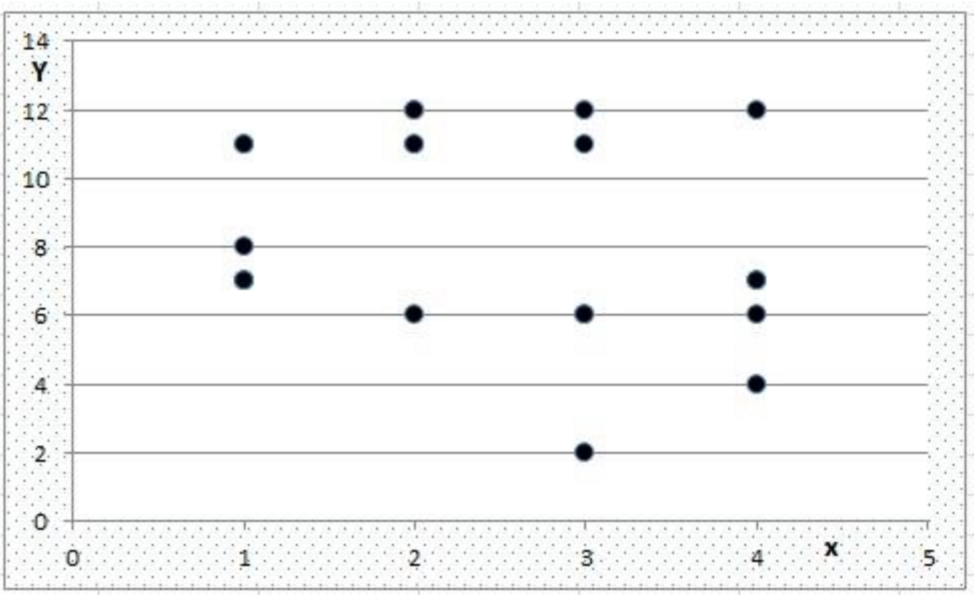
Pour évaluer ses étudiants, un enseignant a préparé quatre tests différents dont chaque étudiant doit choisir un seul test parmi les quatre et le résoudre. Supposant qu'un étudiant a choisi le test numéro 2 et a eu la note 8/20, alors cette information se traduit par le couple de points $(x_i, y_i) = (2, 8)$. A la fin du test, l'enseignant a regroupé tous les résultats comme suit :

(1, 11)	(4, 7)	(3, 6)	(2, 12)	(4,4)	(3, 2)	(4, 6)	(1, 8)	(2, 11)
(2, 12)	(1, 7)	(3, 12)	(4, 7)	(2, 6)	(3, 2)	(1, 7)	(3, 8)	(3, 12)
(3, 8)	(2, 11)	(4, 12)	(1, 11)	(4,12)	(1, 7)	(2, 11)		

1) Etablir le tableau de contingence correspondant à ces couples de points (on note le numéro de test par la variable X et la note obtenue par la variable Y).
2) Quel est le nombre total des étudiants ? le nombre des étudiants qui ont choisi le test numéro 1 ? le nombre des étudiants qui ont eu la note 12/20 ? le nombre des étudiants qui ont choisi le test numéro 3 et ont eu une note supérieure à 10/20 ?
3) Parmi les étudiants qui ont choisi soit le test 3, soit le test numéro 4, quel est le pourcentage des étudiants qui ont eu la note 12/20 ?
4) Donner les effectifs des deux distributions marginales.
5) Calculer la moyenne conditionnelle et l'écart type des deux séries conditionnelles Y/X=3 et X/Y=12.
6) L'enseignant a utilisé les données précédentes pour calculer le coefficient de corrélation linéaire et il a trouvé la valeur -0.17. Commenter.

Solution

1) le nuage de points :



Le tableau de contingence :

Y \ X	2	4	6	7	8	11	12	∑
1	0	0	0	3	1	2	0	6
2	0	0	1	0	0	3	2	6
3	2	0	1	0	2	0	2	7
4	0	1	1	2	0	0	2	6
∑	2	1	3	5	3	5	6	25

2) Le nombre total des étudiants est de 25
   Le nombre total des étudiants qui ont choisi le test numéro 1 est de 6
   Le nombre total des étudiants qui ont eu la note 12/20 est de 6
   Le nombre total des étudiants qui ont choisi le test numéro 3 et ont eu une note supérieure à 10/20 est de 2
3) Parmi les étudiants qui ont choisi le test numéro 3 ou le test numéro 4 : le pourcentage des étudiants qui ont eu la note 12/20 est de 37%

   En effet $\left|\begin{array}{l} n37+n47=4 \\ n3.+n4.=7+6=13 \end{array}\right. \longrightarrow$ le pourcentage=4/13=0,37

4) la distribution marginale (voir le tableau de contingence)
5) **Distribution marginale X :**

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{4} n_{i.} \times x_i = \frac{1}{25}(6.1 + 6.2 + 7.3 + 6.4) = 2{,}52$$

$$(\bar{X})^2 = 6{,}35$$

$$\overline{X^2} = \frac{1}{N}\sum_{i=1}^{4} n_{i.} \times x_i^2 = \frac{1}{25}(6.1^2 + 6.2^2 + 7.3^2 + 6.4^2) = 7{,}56$$

$$Var(X) = \overline{X^2} - (\bar{X})^2 = 1{,}21$$

$\sigma(X) = \sqrt{\mathrm{Var}(X)}$=1,1

Distribution marginale Y :

$$\bar{Y} = \frac{1}{N}\sum_{j=1}^{7} n_{.j} \times y_j = \frac{1}{25}(2.2 + 1.4 + 3.6 + 5.7 + 3.8 + 5.11 + 6.12) = 8{,}48$$

$$(\bar{Y})^2 = 71{,}91$$

$$\overline{Y^2} = \frac{1}{N}\sum_{j=1}^{7} n_{.j} \times y_j^2 = \frac{1}{25}(2.2^2 + 1.4^2 + 3.6^2 + 5.7^2 + 5.11^2 + 8.12^2) = 81{,}52$$

$$Var(Y) = \overline{Y^2} - (\bar{Y})^2 = 9{,}62$$

$\sigma(Y) = \sqrt{\mathrm{Var}(Y)}$=3,10

6) **a) Série conditionnelle Y/$_{X=3}$**

$$Y/_{x=3} = \bar{Y_3} = \frac{1}{n_{3.}}\sum_{j=1}^{7} n_{3j} \times y_j = \frac{1}{7}(2.2 + 4.0 + 6.1 + 7.0 + 8.2 + 11.0 + 12.2) = 7{,}14$$

$$\overline{Y_3^2} = \frac{1}{n_{3.}}\sum_{j=1}^{7} n_{3j} \times y_j^2 = \frac{1}{7}(2.2^2 + 6.1^2 + 8.2^2 + 12.2^2) = 65{,}74$$

$$var(Y_3) = \overline{Y_3^2} - (\bar{Y_3})^2 = 14{,}74$$

$$\sigma(Y_3) = \sqrt{var(Y_3)} = 3{,}83$$

b) Série conditionnelle X/$_{Y=12}$

$$X/_{Y=12} = \bar{X_7} = \frac{1}{n_{.7}}\sum_{i=1}^{4} n_{i7} \times x_i = \frac{1}{6}(1.0 + 2.2 + 3.2 + 4.2) = 9$$

$$\overline{X_7^2} = \frac{1}{n_{.7}} \sum_{i=1}^{4} n_{i7} \times x_i^2 = \frac{1}{6}(2.2^2 + 3.2^2 + 4.2^2) = 9{,}66$$

$var(X_7) = \overline{X_7^2} - (\overline{X_7})^2$=0,66

$$\sigma(X_7) = \sqrt{var(X_7)} = 0{,}8$$

7) le coefficient de corrélation

$$R = \frac{COV(X,Y)}{\sigma(X)\sigma(Y)}$$

$$\overline{XY} = \frac{1}{N} \sum_{i=1}^{4} \sum_{j=1}^{7} n_{ij} x_i y_j$$
$$= \frac{1}{25}(3.1.7 + 1.1.8 + 2.1.11 + 1.2.6 + 3.2.11 + 2.2.12 + +2.3.2 + 1.3.6 + 2.3.8 + 2.3.8 + 2.3.12 + 1.4.4 + 1.4.6 + 2.4.7 + 2.4.12) = 20{,}76$$

Donc :

$COV(X,Y) = \overline{XY} - \bar{X}\bar{Y} = 20{,}76 - (2{,}52)(8{,}48)$=-0,6

Le coefficient de corrélation $R = \frac{-0{,}6}{(1{,}1)(3{,}10)} = -0{,}17$

Le coefficient de corrélation linéaire est très proche de zéro, par suite il n'y a pas de corrélation entre les variables X et Y. On refuse l'ajustement linéaire.

Exercice 3 (examen final 2017/2018)

Une étude a été menée auprès de 12 étudiants afin d'expliquer le score à un examen de mathématique à partir du temps consacré à la préparation de cet examen. Pour chaque étudiant, on dispose du temps de révision en heures (variable X) et du score obtenu sur 800 points (variable Y). Les résultats sont :

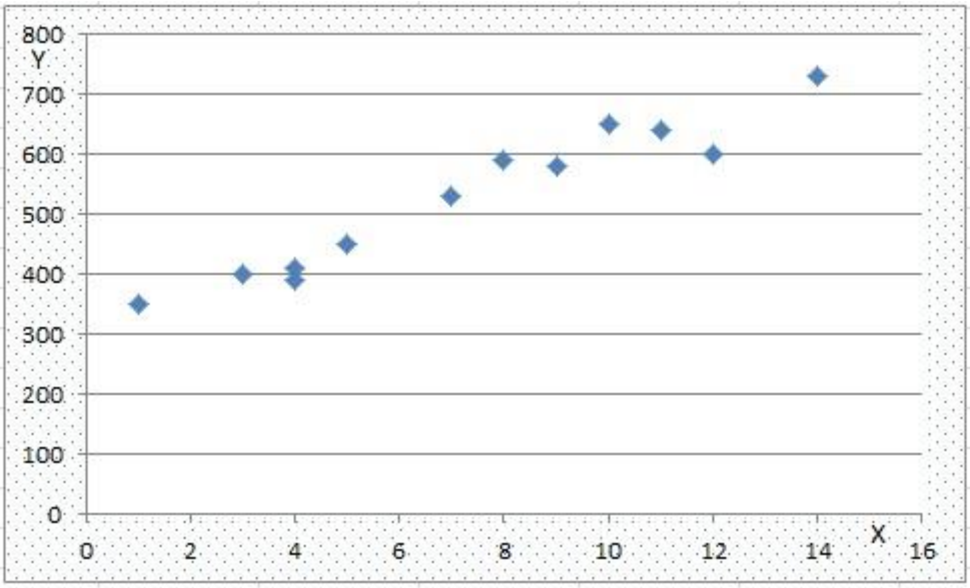
x_i	4	9	10	14	4	7	12	1	3	8	11	5
y_i	390	580	650	730	410	530	600	350	400	590	640	450

On donne : $\sum_{i=1}^{12} x_i^2 = 822$ $\sum_{i=1}^{12} y_i^2 = 3495600$ $\sum_{i=1}^{12} x_i y_i = 51610$

1) tracer le nuage de points.
2) Calculer le coefficient de corrélation linéaire entre ces deux séries. Commenter
3) Déterminer la droite de corrélation linéaire D (Y/X) : Y=aX+b.

Solution

1) Le nuage de points



2) Coefficient de corrélation linéaire

Pour la variable statistique X on a :

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{12} x_i = \frac{88}{12} = 7{,}33$$

$$\overline{x^2} = \frac{1}{N}\sum_{i=1}^{12} x_i^2 = \frac{822}{12} = 68{,}5$$

$$Var(x) = \overline{x^2} - (\bar{x})^2 = 14{,}73$$

$$\sigma(x) = \sqrt{Var(x)} = 3{,}84$$

D'autre part, pour la variable y, on a :

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{12} y_i = \frac{6320}{12} = 525{,}66$$

$$\overline{y^2} = \frac{1}{N}\sum_{i=1}^{12} y_i^2 = \frac{3495600}{12} = 291300$$

$$Var(y) = \overline{y^2 - (\bar{y})^2} = 13922.23$$

$$\sigma(y) = \sqrt{Var(y)} = 117{,}99$$

On peut aussi calculer la covariance entre X et Y comme suit :

$$\overline{xy} = \frac{1}{N}\sum_{i=1}^{12} x_i y_i = \frac{51610}{12} = 4300.83$$

$$Cov(x, y) = \overline{xy} - \bar{x}\bar{y} = 440.41$$

Par suite, le coefficient de corrélation linéaire $R(x, y)$ est donné par:

$$R(x, y) = \frac{Cov(x, y)}{\sigma(x)\sigma(y)}$$
$$\cong 0.97.$$

Puisque le coefficient de corrélation linéaire $R(x, y)$ est très proche de 1, on peut dire qu'il y'a une forte corrélation linéaire entre X et Y.

3. **Droite de corrélation linéaire $Y = aX + b$:**

On commence par calculer les coefficients a et b comme suit:

$$a = \frac{Cov(x, y)}{Var(x)}$$
$$\cong 29.89$$
$$b = \overline{y} - a\overline{x}$$
$$\cong 307.57$$

Par suite la droite de corrélation linéaire $Y = aX + b$ est donnée par:

$$Y = 29.89X + 307.57.$$

La théorie de la probabilité

Module : probabilité et statistique

Ce cours "La théorie de la probabilité" vise à définir la théorie de la probabilité, Présenter des terminologies et des définitions des événements, Les objets et résultats probabilistes sont un support nécessaire à la statistique, c'est le cas par exemple du théorème de Bayes qui sera présenter dans ce cours.

La théorie de la probabilité

Module : probabilité et statistique

1) **Espace de probabilité :**

1.1) **Terminologie et algèbre des événements:**

1.1.1) **l'expérience aléatoire :**

- On appelle expérience aléatoire, toute expérience dont le résultat ne peut être prédit avec certitude, l'ensemble des résultats est par contre complètement déterminer d'avance.

Exemple :

- ❖ Jet d'un dé (on observe le nombre sur la face supérieure du dé).
- ❖ On compte le nombre de voiture qui traverse un pont dans une période donnée.
- ❖ La durée de vie d'une lampe.
- ❖ On mesure a une date fixe de l'année le niveau du cours d'eau d'une rivière ou le taux de précipitation dans une région donnée.

2) **Ensemble fondamentale :**

On appelle ensemble fondamentale ou espace échantillon ou univers associée a une expérience aléatoire, l'ensemble Ω de tous les résultats possible de l'expérience aléatoire en question.

Exemple :

- ❖ Jet de dé : $\Omega = \{1,2, \dots ,6\}$.
- ❖ Jet d'une pièce : $\Omega = \{P, F\}$.
- ❖ Jet de 2 pièces : $\Omega = \{(P, P)\ (P, F)\ (F, P)\ (F, F)\}$.
- ❖ Nombre de voiture : $\Omega = N = \{0,1,2 \dots \}$.
- ❖ Temps de vie d'une lampe : $\Omega = [0, t]\ t > 0$
- ❖ Mesure de cours d'eau : $\Omega = [0, \alpha]$.

3) **Algèbre des événements :**

3.1) **L'événement :**

Soit Ω l'espace échantillon, on appelle «événement» A tous sous ensemble de Ω.

Exemple : Soit A l'événement « avoir un nombre paire » lors de l'expérience du jet de dé : $A = \{2,4,6\}$

Remarque :

- Ω s'appelle l'événement certain.
- φ s'appelle l'événement impossible.
- $\{\omega\}$, $\omega \in \Omega$ s'appelle l'événement élémentaire (singleton)
- L'événement A est la réunion d'événement élémentaire :

$$A = \{\omega_1 , \omega_2 , \omega_3 \} = \bigcup_{i=1}^{3} \{\omega_i\}$$

- On dit que l'événement A est réalisé (à travers ω) $\Leftrightarrow$ le résultat de l'expérience appartient à A :

$$A \text{ se réalise} \Leftrightarrow \omega \in A$$

$$A \text{ ne se réalise pas} \Leftrightarrow \omega \notin A$$

3.2) **L'événement complémentaire** :Il est difinie par le complémentaire de A dans Ω

$$\overline{A} = C_{\Omega}^{A} = \Omega - A$$

$$\omega \notin A \rightarrow \omega \in \overline{A}$$

Remarque :

- $A \cup B$ se réalise si et seulement si A se réalise ou B se réalise

$$\omega \in A \cup B \leftrightarrow \omega \in A \ \vee \ \omega \in B$$

- $A \cap B$ se réalise $\Leftrightarrow$ A se réalise et B se réalise

$$\omega \in A \cap B \leftrightarrow \omega \in A \ \wedge \ \omega \in B$$

- $A - B$ se réalise $\Leftrightarrow$ A se réalise sans B

$$\omega \in A - B \leftrightarrow \omega \in A \ \wedge \ \omega \notin B$$

$$\leftrightarrow \omega \in A \ \wedge \ \omega \in \overline{B}$$

$$\leftrightarrow \omega \in A \cap \overline{B}$$

- $A \Delta B$ se réalise $\Leftrightarrow$ A se réalise toute seule ou B se réalise toute seule

$$\omega \in A \Delta B \leftrightarrow (\omega \in A - B) \vee (\omega \in B - A)$$

- Si $A \cap B = \emptyset \Leftrightarrow$ A et B sont disjoints ou incompatible.
- On appelle système complet $\Leftrightarrow$ Les événements $A_1, A_2, \ldots . A_n$ forment un système complet de $\Omega \Leftrightarrow$

$$\begin{cases} A_i \neq \emptyset \quad \forall \quad i = 1,2,..n \\ A_i \quad A_j = \emptyset \quad \forall \quad i \neq j \\ \bigcup_{i=1}^{n} A_i = \Omega \end{cases}$$

Exercice 1 :

- On jette un dé et on définie les événements A,B,C :

A « Avoir un nombre paire »

B « Avoir un nombre impaire »

C « Avoir un nombre premier »

1- Déterminier Ω , A, B, C puis déterminer $A \cap B$, $A \cap C$, $B \cap C$, $\overline{A}$, $\overline{B}$, $\overline{C}$, $C - A$, $A - C$, $A - (C \cap B)$, A Δ C

2- Déterminer les événements qui forment un système complet.

Solution :

1- $\Omega = \{1,2,3,4,5,6\}$ $\quad A = \{2,4,6\}$ $\quad B = \{1,3,5\}$ $\quad C = \{2,3,5\}$

$A \cap B = \emptyset$ $\quad A \cap C = \{1\}$ $\quad B \cap C = \{3,5\}$ $\quad \overline{A} = \{1,3,5\} = B$

$\overline{B} = \{2,4,6\} = A$ $\quad \overline{C} = \{1,4,6\}$ $\quad C - A = \{3,5\} = B \cap C$ $\quad A - C = \{4,6\}$

$A - (C \cap B) = A - \{3,5\} = A$

$A \,\Delta\, C = (A - C) \cup (C - A) = \{3,5\} \cup \{4,6\} = \{3,4,5,6\}$

2- $\{A, B\}$ forment un système complet.

4) **Espace de probabilité :**

4.1) **Modele uniforme :** « cas d'événement équiprobables »

On désigne par modele uniforme tout espace de probabilité (Ω, F, f), tel que :

- ✓ Ω est un ensemble fini ie $\quad card\, \Omega = N$

- ✓ Et tout les evenements elementaire $\{\omega_i\}$ ont la meme chance de se réalisé, il sont équiprobable et leurs $probailité \ = \ 1/N$.
- ✓ Soit A un evenement t.q $A = \bigcup_{i=1}^{k} \{\omega_i\}$:

$$p(A) = \sum_{i=1}^{k} p(\{\omega_i\})$$

$$= \left(\frac{1}{N} + \frac{1}{N} + \ldots + \frac{1}{N}\right) \text{ k fois}$$

$$= \frac{k}{N} = \frac{cardA}{card\Omega} = \frac{nombre\ de\ cas\ favorable}{nombre\ de\ cas\ possibles}$$

$$p(A) = \frac{cardA}{card\Omega}$$

4.2) **Définition d'une mesure de probabilité :**

On dit que P définie une probabilité sur l'éspace Ω, si seulement si la fonction P vérfie les conditions suivantes :

$$p : P(\Omega) \to [0,1]$$

$$A \to p(A)$$

Tel que : 1) $p(\Omega) = 1$, $p(\emptyset) = 0$

2) $\forall \ \in P(\Omega) \qquad p(A) \geq 0 \qquad 0 \ll P(A) \ll 1$

3) $P(A \cup B) = p(A) + p(B) \Leftrightarrow A \cap B = \emptyset$, ie A et B incompatible.

4) $\forall \ (A_i)_{i=1,n}$ evenements 2 a 2 disjoints ie $A_i \cap A_J = \emptyset \ \forall \ i \neq j$

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} p(A_i)$$

5) **constitution d'une probabilité :**

Propriété 1 : cas continue $\Omega = \mathbb{R}$

Soit f : $\mathbb{R} \to \mathbb{R}$ t.q : - $f(x) > 0 \ \forall x \in \mathbb{R}$

- f intégrable tel que $\int_{\mathbb{R}} f(x)dx = 1$

Alors $p(A)$ est définie par :

$$p(A) = \int_A f(x)dx$$

Propriété 2 : cas discret

p est une probabilité $\Leftrightarrow * \, p(\{\omega_i\}) > 0 \quad \forall \, \omega_i \in \Omega$

$$* \sum_{\omega_i \in \Omega} p(\{\omega_i\}) = 1$$

$$* \, p(\emptyset) = 0$$

Alors p(A) est définie par :

$$p(A) = \sum_{x \in A} p(\{x\})$$

Avec $\mathrm{A} \in \mathrm{P}(\Omega)$, $\mathrm{A} \neq \emptyset$

6) **Propriétés des événements :**

Soit P une loi de probabilité alors :

a) $\mathrm{p}(\overline{\mathrm{A}}) = 1 - \mathrm{p}(\mathrm{A})$

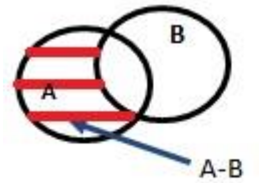
$\mathrm{A} \cup \overline{\mathrm{A}} = \Omega$ et $\mathrm{A} \cap \overline{\mathrm{A}} = \emptyset \rightarrow \mathrm{p}(\Omega) = 1$

$$\rightarrow \mathrm{p}(\mathrm{A} \cup \overline{\mathrm{A}}) = 1$$

$$\rightarrow \mathrm{p}(\mathrm{A}) + \mathrm{p}(\overline{\mathrm{A}}) = 1$$

$$\mathrm{p}(\overline{\mathrm{A}}) = 1 - \mathrm{p}(\mathrm{A})$$

b) $p(A-B) = p(A) - p(A \cap B)$



$(A-B) \cup (A \cap B) = A$ et $(A-B) \cap (A \cap B) = \emptyset$

$p(A) = p(A-B) + p(A \cap B)$

$\rightarrow p(A-B) = p(A) - p(A \cap B)$

Remarque : si $B \subset A$ alors $p(A-B) = p(A) - p(B)$

c) $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$A \cup B = (A-B) \cup B$ et $(A-B) \cap B = \emptyset$

$p(A \cup B) = p(A-B) + p(B)$

$= p(A) - p(A \cap B) + p(B)$

d) $A \subset B \rightarrow p(A) \leq p(B)$

$p(B-A) = p(B) - p(A \cap B)$ or $A \subset B \rightarrow A \cap B = A$

Donc $\left.\begin{array}{c} p(B-A) = p(B) - p(A) \\ p(B-A) \geq 0 \end{array}\right\} \rightarrow p(B) \geq p(A)$

f) $p(A \Delta B) = p(A-B) + p(B-A)$

$= p(A) + p(B) - 2p(A \cap B)$

7 **Probabilités conditionnelles et independance :**

7.1) **Probabilités conditionnelles (événements liés) :**

Soient A et B 2 événements tel.que $P(A) > 0$ et $P(B) > 0$, il est des fois plus facile de calculer $P(A/B)$ « probabilité de A sachant B » avec :

$$p(A/B) = \frac{p(A \cap B)}{p(B)}$$

$$\rightarrow p(A \cap B) = p(A/B).p(B)$$

Ou bien : $p(B/A) = \frac{p(A \cap B)}{p(A)}$

$$\rightarrow p(A \cap B) = p(B/A).p(A)$$

$$\boxed{\begin{array}{l} p(A \cap B) = p(A/B).p(B) \\ p(A \cap B) = p(B/A).p(A) \end{array}}$$

Exemples 2 :

On jette au hasard une piece de monnaie 3 fois , sachat que le jet obtenu est deux cotés différents (pile et face). Quelle est la probabilité qu'on ait 2 apparissions du coté Face.

Solution :

- Soitent les événements :

A « le jet a 2 cotés Face »

B « le jet a 2 cotés différents »

- On …………l'espace des événements Ω :

$\Omega = \{ppp, ppF, pFp, Fpp, FFp, FpF, pFF, FFF,\}$

$card\ \Omega = |\Omega| = 8$

$A \cap B = \{pFF, FpF, FFp\}$, $card\ A \cap B = |A \cap B| = 3$

$p(A \cap B) = \frac{card\ (A \cap B)}{card\ \Omega} = \frac{3}{8}$. , $A = \{FFp, FpF, pFF, FFF\}$

$B = \{pFF, FpF, FFp, Fpp, pFp, ppF\}$, $card\ B = |B| = 6$

$p(B) = \frac{card\ (B)}{card\ \Omega} = \frac{6}{8}$

Donc

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = \frac{3/8}{6/8} = \frac{3}{6} = \frac{1}{2}$$

$$p(A) = \frac{card\ (A)}{card\ \Omega} = \frac{4}{8} = \frac{1}{2}$$

7.2) **Formule de probabilité totales** :

7.2.1 **Pour 2 événements :**

Soient 2 événements E_1, E_2 qui forment un système complet pour

Ω $(ie\ E_1 \cup E_2 = \Omega\ \ , E_1 \cap E_2 = \emptyset)$. Soit A un événement de Ω, alors :

$$p(A) = p(A/E_1).p(E_1) + p(A/E_2).p(E_2)$$

7.2.2 **Pour 3 événements :**

- Soit $\{E_1, E_2, E_3\}$ un système complet de Ω. Soit A un événements de Ω, alors :

$$p(A) = p(A/E_1).p(E_1) + p(A/E_2).p(E_2) + p(A/E_3).p(E_3)$$

7.2.3 **Pour n événements :**

- Soit $\{E_i\}_{i=1,n}$ un système complet de Ω. Soit A un événements de Ω, alors :

$$p(A) = \sum_{i=1}^{n} p(A/E_i)p(E_i)$$

Exemple 3 : (probleme du Tricheur)

On choisit au hasard un individus dans une population possédant $\frac{1}{10}$ de Tricheur.On fait tirer a cette personne une carte d'un jeu de 52 cartes.Si l'individus est un Tricheur, il est certain que la carte soit un As.

∗ Quelle est la probabilité qu'un individu retourne un As ?

Solution :

Soit les événements :

A « l'individus retourne un As »

E « l'individus est un Tricheur »

F « l'individus n'est pas un Tricheur »

Alors $\{E, F\}$ forment un système complet

$$F = \overline{E} \rightarrow p(E) = \frac{1}{10} \quad ; \quad p(F) = \frac{9}{10} \quad ; \quad p(E \cap F) = 0$$

D'après la formule des probabilités totales :

$$P(A) = p(A/E).p(E) + p(A/F).p(F)$$

$$= \frac{1}{10} \times 1 + \frac{9}{10} \times \frac{4}{52} = \frac{11}{65}$$

$p(E) = 1 \rightarrow l'hypoyhése$ $\quad p(F) = \frac{4}{52}$ (4 As dans un jeu de 52 cartes).

7.3) **Formule de Bayes (Probabilité des causes) :**

Soit $\{E_i\}$ un système complet, A un événement t.q $p(A) > 0$ on a :

$$p(E_j/A) = \frac{p(E_j).p(A/E_j)}{\sum_{i=1}^{n} p(E_i).p(A/E_i)}$$

Exemple (suite exemple 3) : (probleme du Tricheur)

Quelle est la probabilité que l'individus soit un Tricheur sachant qu'il a retourné un As ? $p(E/A) = ??$

Solution :

$$p(E/A) = \frac{p(E \cap A)}{p(A)} = \frac{p(A/E).p(E)}{p(A/E).p(E) + p(A/F).p(F)} = \frac{\frac{1}{10} \times 1}{\frac{11}{60}} = \frac{13}{22}$$

7.4) **Evénements indépendants :**

A et B sont 2 événements indépendants $\Leftrightarrow \; p(A/B) = p(A)$

Ou bien $\quad p(A \cap B) = p(A).p(B)$

Remarque : attention à ne pas confondre événement incompatible et événement indépendant.

Exercice 1

Soient A et B deux événement tels que :

$$P(A) = P(B) = \frac{1}{3} \text{ Et } P(A \cap B) = \frac{1}{9}$$

1) Est-ce que les événements A et B sont disjoints (incompatibles) ?
2) Est-ce que les événements A et B sont indépendants ?
3) Calculer les probabilités des événements suivants
   E1 : Aucun des événements A et B se réalise.

   E2 : A se réalise et B ne se réalise pas.

Solution

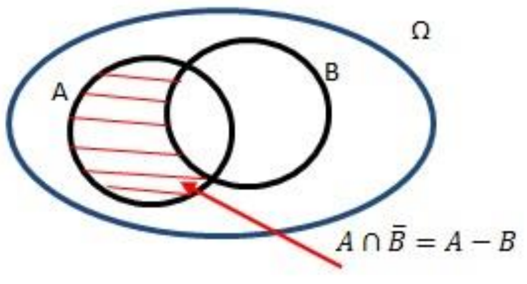
On a : $P(A) = P(B) = \frac{1}{3}$, et $P(A \cap B) = \frac{1}{9}$

1) A et B sont disjoints (incompatible) si seulement si P(A∩B)=0
   $P(A \cap b) \neq 0$ Donc A et B ne sont pas disjoints.
2) A et B sont indépendantes si seulement si P(A∩B)=P(A).P(B)
   $P(A \cap B) = \frac{1}{9}$ Et $P(A).P(B) = \frac{1}{3}.\frac{1}{3} = \frac{1}{9}$ donc A et B sont indépendants
4) E1 : Aucun des événements A et B se réalise.

$$P(E_1) = P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B)$$

$$= 1 - [P(A) + P(B) - P(A \cap B)]$$

$$= 1 - \left(\frac{1}{3} + \frac{1}{3} - \frac{1}{9}\right) = \frac{4}{9}$$



5) E2 : A se réalise et B ne se réalise pas.
   $P(E_2) = P(A \cap \bar{B}) = P(A - B)$

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

$= \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$

Exercice 2

Lors d'une compétition de tir à l'arc, on a constaté qu'un tireur entraîné a 80% de chances d'atteindre sa cible. Parmi les participants, 40% sont des tireurs entraînés. Les autres ont 50% de chances d'atteindre la cible. On choisit un participant au hasard. Sachant que le participant a atteint sa cible, calculer la probabilité qu'il soit un tireur entraîné.

Soient les évènements suivants :

E : Le participant est un tireur entraîné.

C : Le participant atteint sa cible.

Solution

P(C/E)=0,8, P(E)=0,4, $P(C/\bar{E}) = 0,5$

$$\mathrm{P}\left(\frac{\mathrm{E}}{\mathrm{C}}\right) = \frac{P(E \cap C)}{P(C)} = \frac{P(C/E).P(E)}{P(C)}$$

Avec : $P(C) = P(C/E)P(E) + P(C/\bar{E})P(\bar{E}) = 0{,}8.0{,}4 + 0{,}5(1 - 0{,}4) = 0{,}62$

Donc : $P\left(\frac{E}{C}\right) = \frac{0.8.0.4}{0.62} = 0.52$

Exercice 3

Dans une école, on propose deux activités sportives déférentes (athlétisme et gymnas-tique). Le pourcentage d'élèves qui choisissent l'athlétisme est de 55%, celui de ceux qui choisissent la gymnastique est de 40% et celui de ceux qui choisissent les deux activités est de 25%.

1) Calculer la probabilité de choisir au moins une activité sportive.
2) Calculer la probabilité de ne choisir aucune activité sportive.
3) Calculer la probabilité de choisir l'athlétisme seulement.
4) Calculer la probabilité de choisir la gymnastique seulement.
5) En déduire la probabilité de choisir soit l'athlétisme seulement soit la gymnastique seulement.

Si l'élève a choisi l'athlétisme, quelle est la probabilité de choisir la gymnastique aussi ?

Chapitre VI :Variables aléatoires

Module : probabilité et statistique

Dans la plupart des phénomènes aléatoires, le résultat d'une épreuve peut se traduire par une grandeur mathématique, très souvent représentée par un nombre entier ou un nombre réel. La notion mathématique qui représente efficacement ce genre de situation concrète est celle de variable aléatoire. Ce cours donne des définitions de base des variables aléatoires, ainsi que des exemples d'application.

Chapitre VI :Variables aléatoires

Module : probabilité et statistique

1) Introduction

Des fois, après avoir réalisé une expérience aléatoire, on ne s'intéresse pas aux résultats de l'expérience, mais on s'intéresse à une fonction de résultats.

Exemple 1

1) On jette 2 dés distincts et on s'intéresse à la somme des points.
2) On jette 2 dés distincts et on s'intéresse au plus grand chiffre.
3) On observe deux bactéries et on s'intéresse à la durée de vie de la bactérie qui disparaitra la première.

2) Variable aléatoire :

Soit (Ω, P) un espace de probabilité. On appelle variable aléatoire sur cet espace toute application X de Ω dans R, tels que :

$$X: \Omega \rightarrow R$$

$$\omega \rightarrow X(\omega)$$

Exemple 2

On jette 2 dés distincts et on s'intéresse à la somme des points.

$$\mathbf{\Omega} = \{(1.1), (1.2), (1.3), \dots \dots \dots \dots \dots . . (6.5), (6.6) \quad \}$$

Card(Ω)=36

$$X: \Omega \rightarrow R$$

$$(\omega_1, \omega_2) \rightarrow X(\omega_1 + \omega_2)$$

L'ensemble des valeurs possibles de X est $\{2, 3, 4, \dots \dots \dots \dots \dots 12\}$

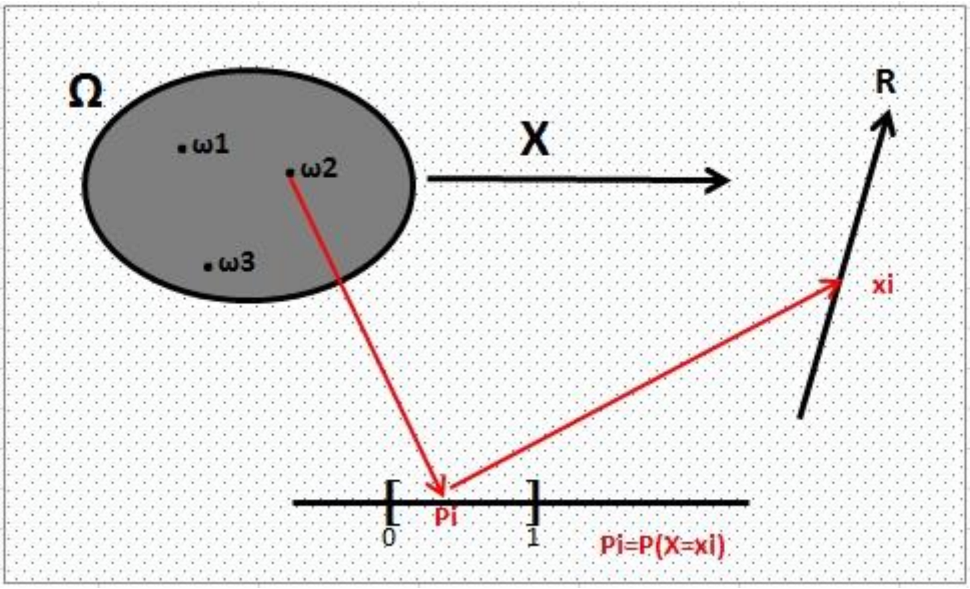
3 Variables aléatoires discrètes :

Lorsque la variable aléatoire X ne prend que des valeurs discrètes, on parle de variable aléatoire discrète. L'exemple 1 et 2 (exemple 1) présentent une variable discrètes, alors que l'exemple 3 présente une variable aléatoire continue.

3) Loi de probabilité :

Une variable aléatoire est caractérisée par l'ensemble des valeurs qu'elle peut prendre et par la chance d'apparition de chaque valeur. On appelle cette chance *« la loi de probabilité »* ou *« la distribution de probabilité »* de la variable aléatoire.

3.1 Cas d'une variable aléatoire discrète :



La loi de probabilité d'une variable aléatoire discrète est entièrement déterminée par les probabilités P_i des événements$\{X = x_i\}$. Donc, la loi de probabilité est donnée par $(x_i, P_i)_i$.

Exemple 3

Pour l'exemple 1, la loi de probabilité est donnée par :

x_i	2	3	4	5	6	7	8	9	10	11	12
$P_i=P(X=x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

$$P_i = \frac{nombre\ de\ cas\ favorables}{nombre\ de\ cas\ possibles}$$

3.2) Cas continue :

La théorie des probabilités est dite **continue** lorsque l'univers Ω n'est plus dénombrable mais quelconque, C'est-à-dire lorsque la théorie des probabilités n'est plus discrète

4. Fonction de répartition :

On appelle fonction de répartition d'une variable aléatoire X, la fonction F_X tel que :

$$F_X : R \;\rightarrow\; R$$

$$t \rightarrow F_X(t) = P(< t)$$

Propriétés :

1) ⊮ $t\ \in R \qquad 0 \leq F_X(t) \leq 1$
2) F_X est croissant sur R.
3) $\lim_{t\to-\infty} F_X(t) = 0$ et $\lim_{t\to+\infty} F_X(t) = 1$
4) Si à ≤ b, $P(a \leq X \leq b) = F_X(b) - F_X(a)$

Exemple 5 : (lance de 3 pièces)

On considère l'événement ω « lancer de 3 pièces ». On introduit une variable aléatoire X définie par :

X(ω) « nombre de piles observés »

1) Déterminer la loi de probabilité de la variable aléatoire X.
2) Déterminer la fonction de répartition.
3) Représenter la loi de probabilité et la fonction de répartition graphiquement.

Solution

La variable aléatoire X est de type discret :

Ω=$\{PPP.PPF.PFP.PFF.FFF.FFP.FPF.FPP\}$

1) La loi de probabilité est donnée par (x_i, P_i)

Nombre de pile (x_i)	$P(X=x_i)$
0	1/8
1	3/8
2	3/8
3	1/8
	$\sum P_i=1$

2) La fonction de répartition est donnée par

Nombre de piles (x_i)	$P(X=x_i)$	F_X
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	1

3) Dans le cas d'une variable aléatoire discrète, on utilise un diagramme en bâton pour visualiser la distribution (la loi) des probabilités et une fonction en escalier pour la fonction de répartition.

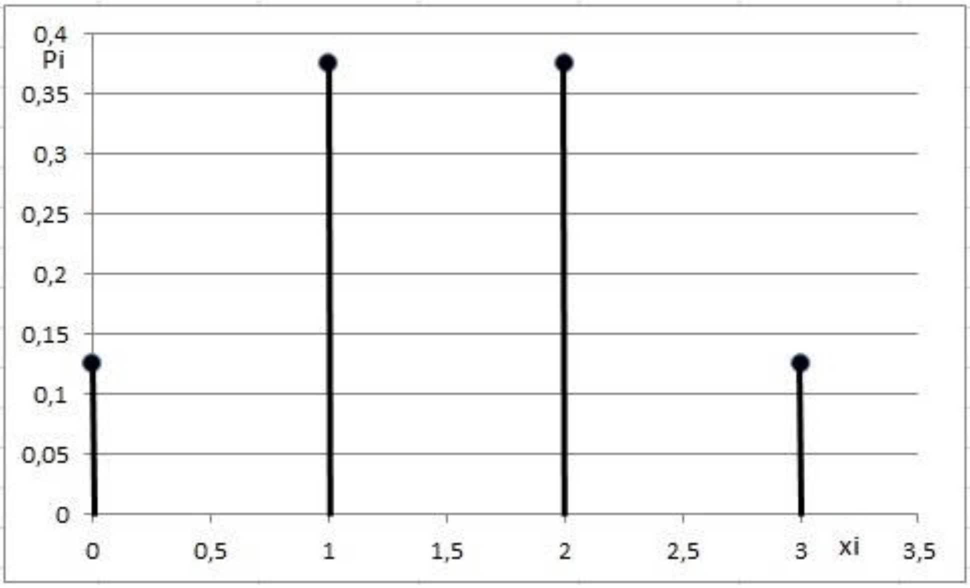
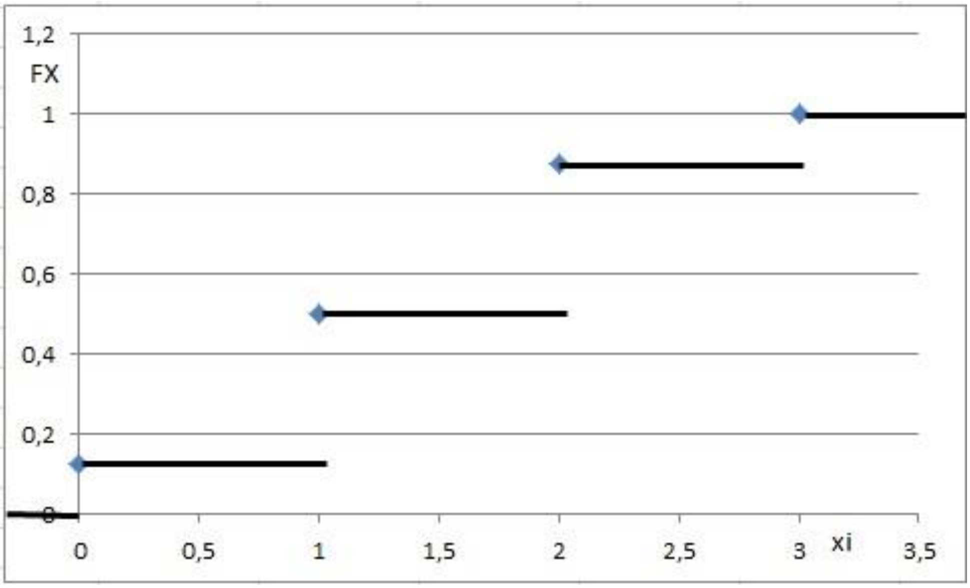


Diagramme des bâtonnets



Fonction de répartition

5 Densité de probabilité :

Elle et définie pour le cas continu : Dans ce cas la fonction de répartition F_X est donnée par :

$$F_X : R \to R$$

$$t \to F_X(t) = P(X < t) = \int_{-\infty}^{t} f(y)dy$$

Ou f est une fonction intégrable sur R satisfaisant :

1) f(x) ≥0 $\forall\ x \in R$
2) $\int_{-\infty}^{+\infty} f(x) = 1$

6. Espérance mathématique :

L'espérance mathématique d'une variable aléatoire X est notée par E[X], elle représente la moyenne prise par la variable X.

Si X est une variable discrète à valeurs dans Δ= $\{x_1, x_2, \ldots\ldots x_n\}$

$$X:\ \Omega \to \Delta$$

$$\omega \to X(\omega)\epsilon\Delta$$

Alors $E[X] = \sum_{i=1}^{n} x_i P(X = x_i) = \sum_{i}^{n} x_i P_i$

Si X est une variable continue à densité f, alors :

$$E[X] = \int_{-\infty}^{+\infty} x.f(x)dx$$

Propriétés :

Soient (a,b) ∈ R^2 et X, Y deux variables aléatoires :

$$E[aX + bY] = aE[X] + bE[Y]$$

- Si x≥0, alors E[X] ≥0.
- Si X≤Y, alors E[X] ≤E [Y]
- Si X est un caractère constant tel que : ∀ ω∈ $\Omega, X(\omega) = k, alors\ E[X] = k$

7. Variance et écart type

La variance d'une variable aléatoire X notée par var(X) est définie par :

$$VAR(X) = E([X - E(X)]^2) = E[X^2] - (E[X])^2$$

Par définition, la variance est toujours positive.

L'écart type d'une variable aléatoire noté par $\sigma(X)$ est défini par :

$$\sigma(X) = \sqrt{Var(X)}$$

L'écart type représente la distance moyenne entre la variable et sa moyenne. Elle mesure la dispersion d'une variable. Plus l'écart type est grand, plus la variable présente des valeurs qui peuvent être éloignées les unes des autres. Plus l'écart type est petit, plus la variable prend des valeurs proches de sa moyenne.

Exercice 1

On jette deux dés différents en même temps et on s'intéresse à la somme des points. On définit la variable aléatoire X par cette somme.

1) Donner les valeurs possibles de X.
2) Déterminer la loi de probabilité de X.
3) Déterminer la fonction de répartition de X.
4) Calculer les probabilités suivantes : P $(1\leq X\leq 3)$, $P(X\leq 6)$, $P(X\geq 7)$ et P $(5\leq X\leq 8)$.
5) Calculer l'espérance et la variance de X.

Solution

Chaque dé contient 6 faces numéroté de 1 à 6, donc :

Ω=$\{(1,1),(1,2),\dots\dots\dots(2,1),(2,2),\dots\dots\dots\dots\dots\dots..(6,1),(6,2),\dots..(6,6)\}$

Card Ω=36

1) La valeur de X est la somme de points, donc elle a les valeurs suivantes :

$$x_i = \{2,3,4,5,6,7,8,9,10,11,12\}$$

2) La loi de probabilité et la fonction de répartition

x_i	2	3	4	5	6	7	8	9	10	11	12
$P(X=x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
$F(x_i)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	1

Les probabilités :

- P (1≤X≤3)= (la somme est entre 1 et 3)

  =P (la somme =1 ou bien =2 ou bien =3)

  =P (la somme =2) ∪ (la somme = 3)

  $=\frac{1}{36}+\frac{2}{36}=\frac{3}{36}=\frac{1}{12}$

- ✓ La somme = 1 cas impossible
- P(X≤6)=P(le somme est ≤6)

=P(X=2) + P(X=3) + P(X=4) + P(X=5) + P(X=6)=$\frac{15}{36}$

- P(X≥7)=P (la somme est ≥7)

  =P (la somme = 8, 9, 10, 11, 12)

  = P(X=8) + P (X=9) + P(X=10) + P(X=11) + P(12)

  $=\frac{5}{36}+\frac{4}{36}+\frac{3}{36}+\frac{2}{36}+\frac{1}{36}=\frac{15}{36}=\frac{5}{12}$
- P (5≤X≤8)=P(X=5) + P(X=6) +P(X=7) + P(X=8)=$\frac{4+5+6+5}{36}=\frac{5}{8}$

6) L'espérance de X

$$E[X]=\sum x_i P(X=x_i)$$

=$2.\frac{1}{36}+3.\frac{2}{36}+4.\frac{3}{36}+\cdots \ldots \ldots \ldots \ldots .12.\frac{1}{36}=\frac{252}{36}=7$

$$E[X^2]=\sum x_i^2 P(X=x_i)=2^2.\frac{1}{36}+3^2.\frac{2}{36}+\cdots \ldots \ldots \ldots \ldots +12^2.\frac{1}{36}=54{,}83$$

La variance : $V[X]=E[X^2]-(E[X])^2=5{,}83$

Exercice 2

Soit la fonction suivante :

$$f_x(x) = \begin{cases} 0,5 & si\ 0 \leq x \leq 2 \\ 0 & sinon \end{cases}$$

1) Vérifier que f_x est une fonction de densité de probabilité.
2) Calculer la probabilité d'avoir X supérieur ou égale à 1.
3) Calculer l'espérance et la variance de X.

Solution

1) On a : $f_x(x) \geq 0$, $\forall\ x\ \in R$

Et : $\int_{-\infty}^{+\infty} f_x(x)dx = \int_0^2 0{,}5dx = 0{,}5[2-1] = 1$

Donc la fonction est une fonction de densité de probabilité.

2) $P(X\geq 1)=\int_{-\infty}^{+\infty} f_x(x)dx = \int_1^2 0{,}5dx = 0{,}5$
3) $E[X] = \int_{-\infty}^{1} xf_x(x)dx = 0{,}5\int_0^2 xdx = 0{,}5\left[\frac{x^2}{2}\right]_0^2 = 1$

$$E[x^2] = \int_{-\infty}^{+\infty} x^2 f_x(x)dx = 0{,}5\int_0^2 x^2dx = 0{,}5\left[\frac{x^2}{3}\right]_0^2 = \frac{4}{3}$$

$$Var[X] = E[X^2] - (E[X])^2 = \frac{1}{3}$$

Exercice 3 : examen 2017/2018 :

On a effectué une étude sur la durée de vie de composants électroniques (mesurée en heure). Soit f_X la fonction de densité de probabilité de la variables aléatoire X qui désigne cette durée, cette fonction est donnée par

$$f_X(t) = \begin{cases} \frac{150}{t^2}, & \text{si } t \geq 150, \\ 0, & \text{si } t < 150. \end{cases}$$

1. Vérifier que cette fonction est bien une fonction de densité de probabilité.
2. Calculer la probabilité pour qu'un composant ait une durée de vie inférieure à 250 heures.

1. **Fonction de densité de probabilité:**

On remarque d'abord que la fonction $f_X(t)$ est toujours positive puis on calcule l'intégrale suivante:

$$\begin{aligned}\int_{-\infty}^{+\infty} f_X(t)dt &= \int_{150}^{+\infty} \frac{150}{t^2}dt \\ &= \left[\frac{-150}{t}\right]_{150}^{+\infty} \\ &= 1.\end{aligned}$$

Donc la fonction $f_X(t)$ est bien une fonction de densité de probabilité.

2. **Probabilité pour qu'un composant ait une durée de vie inférieure à 250 heures:**

$$\begin{aligned}\int_{-\infty}^{250} f_X(t)dt &= \int_{150}^{250} \frac{150}{t^2}dt \ \ \textbf{(0.5 pts)} \\ &= \left[\frac{-150}{t}\right]_{150}^{250} = 0.4. \ \ \textbf{(0.5 pts)}\end{aligned}$$

Références

[1] ZENDAGUI D, Cours et exercices, Polycopié Université Abou Bakr Belkaid, Temcen, 2011.

[2] BENAHCHILIF S, Cours et exercice, Math 4, Université Abou Bakr Belkaid, 2017.

[3] FABRICE MAZEROLLE, *Statistique descriptive: Série statistique à une et deux variables, Séries chronologiques,Indices.* Paris: Gualino, 2005.

[4] ANDRE VESSEREAU, *La Statistique*, Paris, Presses Universitaires de France, coll. « Que sais-je » (no 281), 1976

[5] BLARD-LABORDERIEEe*, L'essentiel des outils de statistique descriptive pour aborder* des études en science *humaines et sociales, 2015.*

[6] CHAUVAT G and REAU J*, Statistiques descriptives, Armand Colin, 2002.*

[7] TENENHAUS G*, statistique : Méthodes pour décrire, expliquer et prévoir, Dunod, 2006.*

[8] DROESBEKE JJ*, Éléments de statistiques, Ellipses, 2001.*

[9] LEBOUCHER L and Voisin G*, Introduction à la statistique descriptive, 2013.*

[10] MAZEROLLE L*, Statistique descriptive, 2009.*

[11] OUKACHA B and BENMESSOUD M*, Statistique descriptive et calcul des probabilités, 2013.*

[12] OUVRARD J Y*, Probabilités : Tome 1, 2001.*

[13] ROGER P*, Probabilités, statistique et processus stochastiques, Pearson Education,, 2004.*

[14] VAILLANT J*, Eléments de Statistique descriptive, 2015.*

[15] THIERRY V*, « cours et exercices » Ecole des Mines de Nancy, 2012*