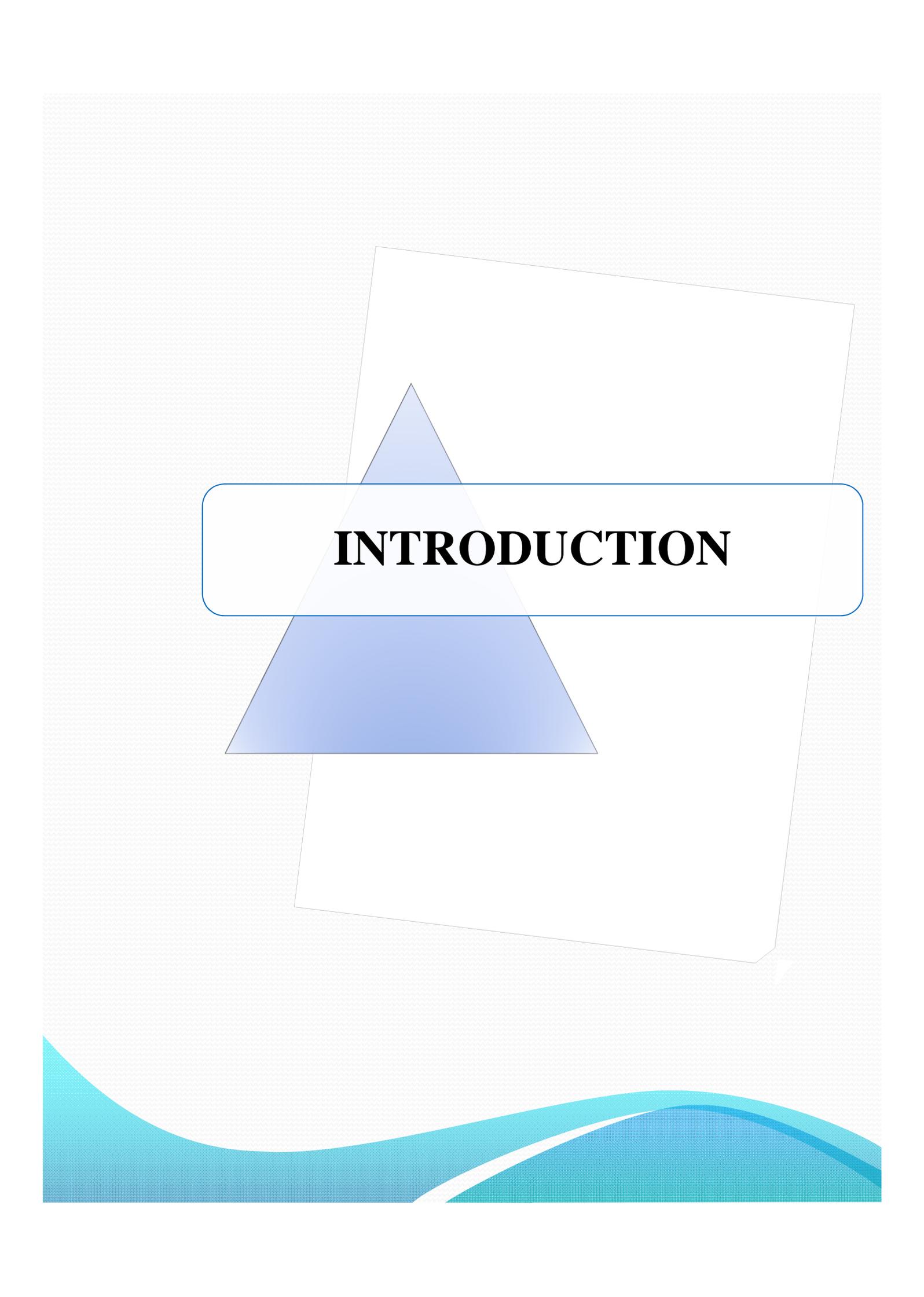


TABLE DE MATIERES

INTRODUCTION.....	03
1. Historique et définitions de la statistique.....	05
2. La statistique descriptive.....	06
3. Les notions de base de la statistique.....	06
3.1. Élément.....	06
3.2. Population	07
3.3. Echantillon.....	07
3.4. Echantillonnage.....	07
3.5. Inférence statistique.....	08
3.6. Caractère	08
3.7. Modalité.....	08
3.8. Effectif	08
3.9. Fréquence d'une modalité ou d'une classe.....	10
4. La nature des caractères	10
4.1. Caractère qualitatif.....	11
4.1.1. Echelle nominale.....	11
4.1.2. Echelle ordinale	11
4.2. Caractère quantitatif.....	12
4.2.1. Repérable ou Mesurable.....	12
4.2.2. Stock ou Intensité.....	12
4.2.3. Discret ou Continu.....	13
4.3. Combinaisons de critères.....	13
4.4. Principe du nombre de classes.....	14
4.4.1. Forme d'une classe ou intervalle.....	14
4.4.2. Nombre de classes	15
4.4.3. Intervalle de classe (amplitude).....	15
4.4.4. Centre d'une classe.....	15
5. Tableaux statistiques.....	15
5.1. Tableau élémentaire (brut).....	15
5.1.1. Cas qualitatif.....	16
5.1.2. Cas quantitatif.....	16
5.2. Tableau de dénombrement	17
5.3. Tableau de distribution des fréquences	17

6. Représentations graphiques.....	19
6.1. Cas qualitatif	19
6.1.1. Diagramme en tuyaux d'orgue.....	19
6.1.2. Le camembert.....	20
6.2.Cas quantitatif.....	20
6.2.1. Variable statistique discrète.....	21
6.2.2. Variable statistique continue.....	22
7. Description numérique	23
7.1. Tendance centrale ou Paramètres de position.....	23
7.1.1. Le mode.....	23
7.1.2. La médiane.....	25
7.1.3. La moyenne.....	28
7.2. Valeurs centrales et forme des distributions.....	29
7.2.1. Distributions bimodales ou multimodales.....	29
7.2.2. Distributions unimodales symétriques.....	30
7.2.3. Distributions unimodales dissymétriques	30
7.3. Paramètres de dispersion absolue.....	31
7.3.1. La fourchette des valeurs ou étendue	31
7.3.2. Ecart absolu moyen et l'écart absolu médian.....	31
7.3.3. Variance et écart type.....	31
7.3.4. L'étendue interquartile	32
7.4. Paramètres de dispersion relative.....	33
7.4.1. Le coefficient de variation.....	33
7.4.2. Le coefficient interquartile relatif.....	34
7.4.3. Ecart relatif moyen.....	34
7.4.4. Ecart relatif médian.....	34
7.5. Paramètres de dispersion absolue et valeurs centrales.....	34
7.6. Les paramètres de forme	35
7.6.1. Le Coefficient d'asymétrie.....	35
7.6.2. Le coefficient d'aplatissement.....	35
8. Notion de loi normale.....	36
QUELQUES EXERCICES D'APPLICATION ET SOLUTIONS.....	38
BIBLIOGRAPHIE.....	66



INTRODUCTION

INTRODUCTION

La « *Biostatistique* » est née à la fin du 19^{ème} siècle avec les travaux de Karl Pearson, le fondateur du premier département de statistique au monde, à l'Université de London, et de Ronald Fisher, un pionnier dans le domaine des plans expérimentaux.

Cette science, qui est aussi connue sous le nom « *Biométrie* », est un champ scientifique constitué par l'application de la « *Science Statistique* » à un large éventail de sujets dans le domaine du vivant, plus particulièrement la « *Biologie* » qui recouvre une partie des sciences de la nature et de l'histoire naturelle des êtres vivants, et la « *Médecine* » qui étudie l'organisation du corps humain, son fonctionnement normal, et cherchant à préserver la santé par le traitement et la prévention des pathologies.

Par ailleurs, les « *Statistiques* » sont la science des données, qui est à la fois une science, une méthode et un ensemble de techniques. Ceci implique la collection, la classification, le résumé, l'organisation, l'analyse et l'interprétation d'une information numérique.

De ce fait, la science de la Biostatistique englobe à la fois : **(1)** La conception des expériences biologiques ; **(2)** La collecte des informations ; **(3)** La compilation et analyse des données chiffrées de ces expériences ; **(4)** L'interprétation des résultats en vue d'avancer une conclusion.

Elle permet donc de décrire un groupe d'individus selon ses attributs et ses qualités, de mesurer la précision d'une estimation ou de définir le degré d'association entre une série de caractères et d'événements.

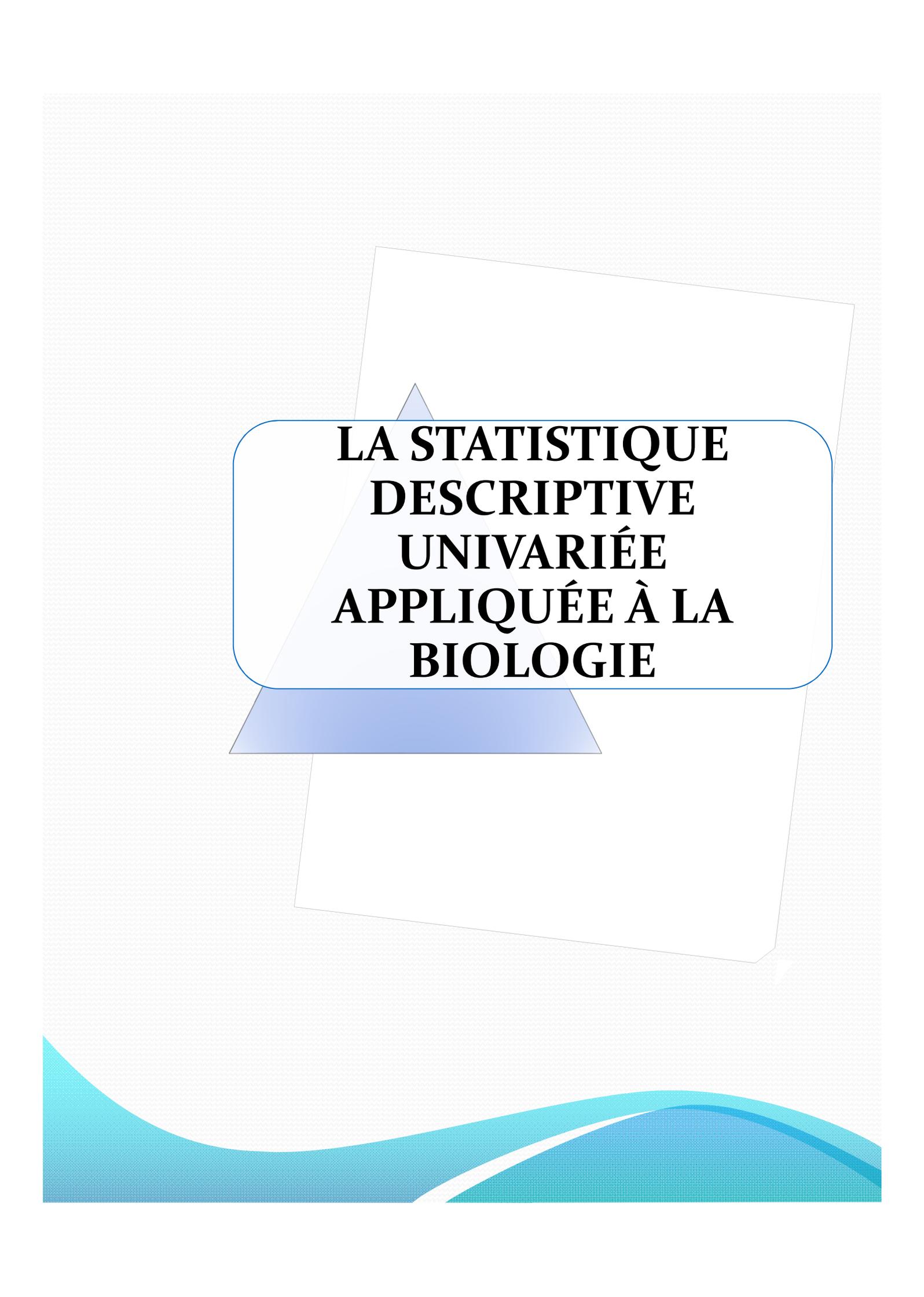
Cette science est exploitée alors dans plusieurs domaines: **(1)** La santé publique, y compris l'épidémiologie, les services de santé, la nutrition et l'environnement ; **(2)** La conception et analyse d'essais cliniques en médecine ; **(3)** La génomique, génétique des populations et la génétique statistique afin de relier la variation dans le génotype avec une variation dans le phénotype ; **(4)** L'agriculture afin d'améliorer les cultures et les animaux d'élevage ; **(5)** L'écologie en vue de mettre en place des prévisions écologiques ; **(6)** L'analyse de séquences biologiques.

La Biostatistique est l'une des matières fondamentales intégrées dans le programme de tous les étudiants de Biologie de l'Université de Béjaïa. Le but recherché par ces programmes d'enseignement est l'initiation de l'étudiant aux traitements de données liées aux thématiques biologiques. L'enseignement de biostatistique aux étudiants de 3^{ème} année « Biochimie Appliquée » que j'ai pris en charge depuis 2006 jusqu'à ce jour a été travaillé et remanié à

plusieurs reprises afin de l'adapter au maximum possible aux diverses applications à l'affiliation des étudiants en question.

Ce document permet à l'étudiant de découvrir tout d'abord les différents vocabulaires de la statistique et leurs synonymies en biologie expérimentale, et par la suite, l'initier à l'organisation des différentes données univariées générées lors d'une expérimentation sur terrain ou au laboratoire, le calcul des principaux paramètres et la représentation graphique des résultats.

Le polycopié est structuré selon un enchaînement logique de plusieurs sections, dont la première est consacrée à la définition de la statistique et sa composante descriptive. Ensuite, il a été expliqué la signification des différents concepts et notions adoptées dans la démarche statistique, suivi d'un détail sur la nature des caractères étudiées en statistique et les différentes techniques d'organisation et de structuration d'une série de données. Ce polycopié expose également les différents paramètres descriptifs d'une série statistique simple à savoir les paramètres de position, les paramètres de dispersion et les paramètres de forme d'une distribution de fréquences, avec enfin une introduction à la loi normale. Chaque section est accompagnée d'un ensemble d'exemples explicatifs. La fin du document a été réservée à quelques exercices d'application avec leurs corrections.

The slide features a light blue background with a fine grid pattern. A white rectangular box with rounded corners and a thin blue border is centered on the page. Inside this box, the title is written in bold, black, uppercase letters. The text is arranged in five lines. Behind the white box, there is a light blue triangle pointing upwards. At the bottom of the slide, there is a decorative graphic consisting of several overlapping, wavy shapes in various shades of blue and cyan, creating a sense of movement and depth.

**LA STATISTIQUE
DESCRIPTIVE
UNIVARIÉE
APPLIQUÉE À LA
BIOLOGIE**

LA STATISTIQUE DESCRIPTIVE UNIVARIEE

1. Historique et définitions de la statistique

L'origine du mot « statistique » remonte au latin classique « status » (état) qui, par une série d'évolutions successives, aboutit au mot français « statistique », attesté pour la première fois en 1771.

C'est vers la même époque que « statistik » apparut en allemand, alors que les anglophones utilisaient l'expression « political arithmetic » jusqu'en 1798, date à laquelle le mot « statistics » fit son entrée dans cette langue.

A l'origine, cette discipline concernait exclusivement les affaires de l'Etat. Actuellement, on distingue généralement les statistiques (au pluriel) de la statistique (au singulier).

Les statistiques concernent l'étude méthodique des faits sociaux qui définissent un Etat, par des procédés numériques (dénombrements, inventaires, recensements,...).

La statistique est un ensemble de techniques d'interprétation mathématique appliquées à des phénomènes pour lesquels une étude exhaustive de tous les facteurs est impossible à cause de leur grand nombre ou de leur complexité

La statistique est pour les uns un domaine des mathématiques, pour les autres (en particulier les anglo-saxons) une discipline à part entière hors des mathématiques. Enfin, de plus en plus, elle fait partie de ce que l'on appelle aujourd'hui la Science des Données.

Fisher a défini la statistique comme la discipline qui étudie les méthodes de réduction de données, la variabilité et les populations.

Les méthodes de réduction des données font partie de la statistique descriptive (ou exploratoire). Elles consistent à essayer de résumer un échantillon de données via des graphiques ou des caractéristiques numériques.

L'étude de la variabilité cherche à l'expliquer. Elle fait partie de la théorie de l'échantillonnage.

L'étude des populations fait partie de la statistique inférentielle qui prend un échantillon et en tire des conclusions pour toute la population. Elle part donc de l'expérience à l'hypothèse (faite au départ).

2. La statistique descriptive

La statistique descriptive est la branche des statistiques qui regroupe les nombreuses techniques utilisées pour décrire un ensemble relativement important de données. Il est assez compliqué de définir la meilleure description possible d'un phénomène. Dans le cadre des statistiques, il s'agira de fournir toute l'information disponible sur le phénomène en moins de chiffres et de mots possibles

Cette démarche a pour but de :

- ✓ Résumer et synthétiser l'information contenue dans la série statistique ;
- ✓ Mettre en évidence ses propriétés ;
- ✓ Suggérer des hypothèses relatives à la population dont est issu l'échantillon.

Les Outils utilisés sont :

- ✓ Les Tableaux ;
- ✓ Les Graphiques ;
- ✓ Les indicateurs.

Le type d'outils utilisés dépend de :

- ✓ La nature de la série (uni ou multidimensionnelle) ;
- ✓ La nature des variables (quantitatives ou qualitatives).

Si les données ne sont relatives qu'à *une seule* variable, on parle de statistique descriptive « **univariée** ». Dans le cas où l'on s'intéresse à *deux* variables simultanément, on met en œuvre la statistique descriptive « **bivariée** ». Si l'ensemble de données provient de l'observation de *plusieurs* variables, on doit faire appel aux méthodes de la statistique descriptive « **multivariée** ».

Ce document est consacré à la présentation des outils et méthodes de la statistique descriptive univariée.

3. Les notions de base de la statistique

3.1. Élément

C'est une unité qui peut être :

- ✓ *Un individu* (êtres vivants) : humain, animal, végétal ... ;
- ✓ *Un sujet* : modules enseignés en biologie, les nationalités, les métiers ou professions... ;

- ✓ *Un objet* : Table, chaise, verrerie de laboratoire ;
- ✓ *Une association* (dans les études écologiques en général) : une parcelle d'herbe, une association d'arbustes...

3.2. Population

C'est un ensemble d'éléments possédant au moins une caractéristique commune et exclusive permettant de l'identifier et la distinguer sans ambiguïté de toutes les autres.

Exemple

- ✓ Une population algérienne ;
- ✓ Une population estudiantine ;
- ✓ Une population de plantes médicinales ;
- ✓ Une population de poissons d'eau douce.

3.3. Echantillon

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

Exemple

- ✓ Etude de 20 étudiants pris à partir d'une population de 57.
- ✓ Etude de 5 régions prises à partir d'une population de 25.
- ✓ Etude de 5 modules pris à partir d'une population de 13.
- ✓ Etude de 200 patients pris à partir d'une population de 660.

3.4. Echantillonnage

C'est l'opération ou la méthode qui consiste à prélever une partie de la population (échantillon). Cette méthode doit assurer la représentativité de cette population c'est-à-dire qu'elle doit refléter fidèlement sa composition et sa structure. Il existe plusieurs méthodes d'échantillonnage qui varient en fonction de la nature de l'étude envisagée. On peut citer l'échantillonnage : Stratifié, par degré, systématique... Le plus utilisé est l'échantillonnage aléatoire et simple qui est basé sur le principe que tous les éléments de la population ont une probabilité égale (non nulle) de faire partie de l'échantillon : c'est une méthode

d'échantillonnage permettant de choisir n unités parmi les N de la population de façon aléatoire.

3.5. Inférence statistique

C'est une opération logique qui consiste à porter un jugement sur un ensemble vaste « population statistique », à partir d'un sous ensemble « échantillon ».

3.6. Caractère

C'est une propriété possédée par les unités statistiques permettant de les décrire et de les distinguer les unes des autres. Toute unité statistique peut être étudiée selon un ou plusieurs caractères –le détail de cette notion est donnée dans la section 4).

Exemple

- ✓ Couleur des yeux ;
- ✓ Poids des souris ;
- ✓ Superficie d'une pièce ;
- ✓ La température de l'air.

3.7. Modalité

Ce sont les diverses situations (cas, état, valeur) susceptibles d'être prises par le caractère. Un caractère peut posséder une ou plusieurs modalités.

Exemple

- ✓ Couleur des yeux : vert, bleu, noir ;
- ✓ Poids des souris (en grammes) : 15, 18, 20, 39 ;
- ✓ Superficie d'une pièce (en mètres) : 3, 5, 6 ;
- ✓ La température de l'air (en °C) : 8, 16, 27, 30, 38.

3.8. Effectif

Plusieurs définitions sur les effectifs.

- ✓ L'effectif de la valeur x_i est le nombre d'individus de la population ayant cette valeur ou appartenant à cette classe : on le note n_i ;

- ✓ L'effectif total N est la somme de tous les effectifs, appelé également en Maths « cardinal ou la taille » ;
- ✓ En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer l'effectif cumulé croissant en faisant la somme des effectifs de cette valeur et de tous ceux qui la précèdent.

Exemple

Dans une promotion de 20 étudiants de Biochimie, voici les notes obtenues au dernier examen de Biostatistique : 10, 14, 12, 15, 7, 8, 10, 11, 12, 18, 2, 4, 12, 13, 14, 15, 19, 11, 9, 0.

On va calculer les effectifs et les effectifs cumulés.

- ✓ Premièrement, les effectifs :
- ✓ Combien d'étudiants ont eu 10 ? c'est 2.
- ✓ Combien d'étudiants ont eu 12 ? c'est 3.

On continue ainsi et on forme le tableau suivant (Tableau I):

Tableau I : Les effectifs n_i relatifs à chaque modalité x_i (notes)

Notes x_i	0	2	4	7	8	9	10	11	12	13	14	15	18	19
Effectifs n_i	1	1	1	1	1	1	2	2	3	1	2	2	1	1

Les effectifs cumulés maintenant. On fait la somme des effectifs de la note + la somme des effectifs de toutes les notes qui la précèdent. Ce qui nous donne (Tableau II):

Tableau II : Les effectifs n_i et $N_i\uparrow$ relatifs à chaque modalité x_i (notes)

Notes x_i	0	2	4	7	8	9	10	11	12	13	14	15	18	19
Effectifs n_i	1	1	1	1	1	1	2	2	3	1	2	2	1	1
Effectifs $N_i\uparrow$	1	2	3	4	5	6	8	10	13	14	16	18	19	20

Remarque

Pour vérifier qu'on ne s'est pas trompé dans le calcul des effectifs cumulés, on vérifie bien que le dernier effectif cumulé correspond bien au nombre d'individus N_i . Ici, on retrouve bien 20, le nombre d'étudiants de cette promotion de Biochimie.

3.9. Fréquence d'une modalité ou d'une classe

- ✓ La fréquence d'une valeur est le quotient de l'effectif de la valeur par l'effectif total ;
- ✓ En rangeant les valeurs du caractère dans l'ordre croissant, on peut calculer les fréquences cumulées croissantes en faisant la somme des fréquences de cette valeur et de tous ceux qui la précèdent ;
- ✓ Pour les fréquences cumulées croissantes, c'est un peu le même principe que pour les effectifs cumulés croissants.

Remarque

Les fréquences sont comprises entre 0 et 1.

Exemple

On reprend l'exemple précédent et on applique tout simplement la formule des fréquences pour les calculer (Tableau III).

Tableau III : Les fréquences f_i et $F_i\uparrow$ relatives à chaque modalité x_i (notes)

Notes x_i	0	2	4	7	8	9	10	11	12	13	14	15	18	19
Fréquences f_i	0,05	0,05	0,05	0,05	0,05	0,05	0,1	0,1	0,15	0,05	0,1	0,1	0,05	0,05
Fréquences $F_i\uparrow$	0,05	0,1	0,15	0,2	0,25	0,3	0,4	0,5	0,65	0,7	0,8	0,9	0,95	1

Remarque

Pareil, pour vérifier qu'on ne s'est pas trompé dans le calcul des fréquences cumulées, on vérifie bien que la dernière fréquence cumulée vaut bien 1.

4. La nature des caractères

Le caractère d'une série statistique est l'objet de notre étude sur une population donnée. Ce caractère prend un certain nombre de valeurs ou de modalités.

On distingue deux grandes familles de caractères :

- ✓ Les caractères qualitatifs ;
- ✓ Les caractères quantitatifs.

Un critère commode est le suivant : la moyenne d'un caractère quantitatif a un sens alors que la moyenne d'un caractère qualitatif est impossible à réaliser ou bien n'a aucune signification.

4.1.Caractère qualitatif

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

4.1.1. Echelle nominale

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

Exemple : cas des noms

- ✓ Etat matrimoniale : marié, célibataire, veuf, divorcé ;
- ✓ Sexe : féminin, masculin ;
- ✓ Profession : enseignant, médecin ;
- ✓ Nationalité : Algérienne, Tunisienne ;
- ✓ Les différentes séquences nucléotidiques ;
- ✓ Les hormones : œstradiol, progestérone.

Exemple : cas des codes

- ✓ Etat matrimoniale : marié (1), célibataire (2), veuf (3), divorcé (4) ;
- ✓ Sexe : féminin (1), masculin (2) ;
- ✓ Profession : enseignant (1), médecin (2) ;
- ✓ Nationalité : Algérienne (1), Tunisienne (2) ;

4.1.2. Echelle ordinale

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

Exemple :

- ✓ Degré d'intelligence : pas intelligent (0), peu intelligent (1), moyennement intelligent (2), très intelligent (3) ;
- ✓ Forme des fruits : petite (1), moyenne (2), grosse (3) ;
- ✓ Abondance/Dominance : peu abondant (1), abondant (2), très abondant (3).

4.2. Caractère quantitatif

Un caractère est quantitatif si ses modalités s'expriment par des nombres. Le nombre de valeurs que peut prendre la variable est illimité. Il existe 3 critères :

4.2.1. Repérable ou Mesurable

On distingue les caractères quantitatifs repérables sur une échelle d'intervalle et les caractères quantitatifs mesurables selon que la valeur 0 est arbitraire ou selon qu'elle a un sens concret.

a. Quantitatif repérable sur une échelle d'intervalle

Ces caractères permettent de repérer la position de chaque individu par rapport à une origine arbitraire. La valeur 0 est donc conventionnelle et ne signifie pas l'absence du phénomène.

Exemple :

- ✓ Latitude, longitude, température, altitude.

b. Quantitatif mesurable sur une échelle numérique

Le 0 signifie bien l'absence du phénomène.

Exemple :

- ✓ Un « 0 cm » de précipitations indique l'absence de la pluie ;
- ✓ Un « 0 » dans les notes indique que l'étudiant ne possède aucune réponse juste ;
- ✓ Un « 0 » dans un compte bancaire indique que le compte est vide ;
- ✓ Un « 0 » dans un dosage d'une molécule dans un organe indique l'absence de cette molécule.

4.2.2. Stock ou Intensité

On peut par ailleurs distinguer les caractères quantitatifs de stock et les caractères quantitatifs d'intensité (taux) selon que la somme des modalités a un sens ou non.

a. Les caractères quantitatifs de stock

Expriment des quantités concrètes : la somme des modalités des éléments a un sens.

Exemple :

- ✓ La population est un caractère de stock car la somme de plusieurs populations a un sens.

b. Les caractères quantitatifs d'intensité

Expriment des caractéristiques des individus mais leur total n'a pas de signification.

Exemple :

- ✓ Les âges des individus car la somme de plusieurs âges de population n'a aucun sens.

4.2.3. Discret ou Continu

Enfin, on peut distinguer les caractères quantitatifs discrets et les caractères quantitatifs continus selon que leurs modalités (valeurs) sont définies sur un intervalle continu de l'ensemble des réels (modalités en nombre infini) ou selon qu'elles correspondent à un ensemble fini et dénombrable de valeurs entières ou réelles.

a. Les caractères quantitatifs discrets

Sont des caractères dont les modalités sont des nombres isolés, pas nécessairement entiers.

Exemple :

- ✓ Nombre de pièces d'un immeuble ;
- ✓ Nombre d'enfants d'une famille ;
- ✓ Nombre des doigts

b. Les caractères quantitatifs continus

Sont des caractères dont les modalités sont définies sur un intervalle (continu) de valeur donné appelé domaine de variation et défini par les valeurs minimales et maximales.

Exemple :

- ✓ Notes des étudiants, la taille, le poids, l'âge.

4.3. Combinaisons de critères

Les typologies précédentes se combinent lorsque l'on veut définir le type d'un caractère quantitatif :

Exemple :

- ✓ Le poids des étudiants : continu, d'intensité, mesurable.
- ✓ Le poids des souris dans le laboratoire : continu, de stock, mesurable.
- ✓ Nombre d'enfant d'une famille : discret, de stock, mesurable.

- ✓ Température : continue, d'intensité, repérable.
- ✓ Altitude : continue, d'intensité, repérable.

Il est très important de savoir définir le type d'un caractère car les outils statistiques ou les représentations graphiques ne sont pas les mêmes selon le type de caractère à étudier.

4.4.Principe du nombre de classes

Dans le cas continu, il est nécessaire de regrouper les résultats en classes à cause de leur grande masse.

Discrétiser une variable quantitative c'est, mathématiquement, transformer un vecteur de nombres réels en un vecteur de nombres entiers nommés « indices de classe ».

En statistiques, discrétiser c'est à la fois réaliser cette transformation mathématique, nommer et justifier les classes.

Un bon découpage correspond à des classes homogènes et séparées, ce qui correspond respectivement aux notions statistiques de faible variance intraclasse et de forte variance interclasse. Mais d'autres critères sont possibles, comme l'équirépartition, le respect d'un nombre minimal de données par classe, etc. Voici ci-après (Tableau IV) le nombre de classe en fonction du nombre d'élément.

Tableau IV : Nombre de classes en fonction du nombre d'éléments d'une série statistique donnée.

Nombre d'éléments	Nombre de classes
10	2 à 3
20	3 à 4
50	4 à 6
100	5 à 8

4.4.1. Forme d'une classe ou intervalle

$[e_{i-1}, e_i[$ avec :

e_{i-1} : Borne ou extrémité inférieure et e_i : Borne ou extrémité supérieure.

4.4.2. Nombre de classes

Il existe plusieurs formules pour le calcul du nombre de classes. Pour les applications de ce cours, nous avons opté pour la *Règle de STURGE - La formule mathématique de HUNTSBERGER-*

$$k = 1 + 3.3 \log N \dots\dots\dots\text{Formule 1}$$

4.4.3. Intervalle de classe (amplitude)

$$a_i = \frac{x_{\max} - x_{\min}}{k} \dots\dots\dots\text{Formule 2}$$

4.4.4. Centre d'une classe

$$c_i = \frac{e_{i-1} + e_i}{2} \dots\dots\dots\text{Formule 3}$$

La fréquence moyenne ou densité d'effectif qui mesure la concentration des observations à l'intérieur d'une classe. Cette valeur est toujours calculée à une constante près de sorte qu'il est indifférent d'utiliser l'une ou l'autre des formules suivantes.

$$f_{im} = f_i / a_i \text{ ou } n_{im} = n_i / a_i \dots\dots\dots\text{Formule 4}$$

5. Tableaux statistiques

Un tableau statistique constitue un résumé ou une synthèse numérique des résultats d'une distribution statistique, on distingue trois formes de tableaux statistiques qui sont fonction de l'objectif envisagé et de la nature du caractère étudié.

5.1. Tableau élémentaire (brut)

Il est de la forme (Tableau V):

Tableau V : Tableau élémentaire

Eléments i	Modalités x_i
1	x_1
2	x_2
3	x_3
.	.
.	.
.	.
k	x_k

Après la collecte des données, celles-ci apparaissent de façon brute. Sous cette forme, elles sont peu informatives. Il nous faut donc des moyens pour en extraire un maximum d'informations.

5.1.1. Cas qualitatif

Dans une étude d'extraction de substances végétales, on fait un sondage pour savoir ce que chacun des 25 étudiants/es a trouvé le plus intéressant. On utilise la règle suivante (Tableau VI) :

Tableau VI : Tableau des codes relatifs aux réponses des étudiants

Réponse	X
Les huiles essentielles	1
Les flavonoïdes	2
Les polyphénols totaux	3
Les tannins	4
Autres	5

Les résultats obtenus sont les suivants (Tableau VII) :

Tableau VII : Tableau des résultats du sondage de 25 étudiants (E : étudiant ; R : réponse)

E	R	E	R	E	R	E	R	E	R
E1	3	E6	2	E11	2	E16	4	E21	2
E2	3	E7	4	E12	2	E17	3	E22	2
E3	5	E8	3	E13	3	E18	4	E23	3
E4	4	E9	4	E14	5	E19	4	E24	5
E5	2	E10	5	E15	5	E20	3	E25	1

5.1.2. Cas quantitatif

Voici les résultats du dosage des polyphénols obtenus par 25 étudiants de Biochimie par des méthodes d'extraction différentes (unité en $\mu\text{g/g} * 100$) (Tableau VIII):

Tableau VIII : Tableau des résultats des doses en polyphénols obtenus par des méthodes d'extraction différente (E : Etudiant, D : Doses en polyphénols)

E	D								
E1	68	E6	65	E11	78	E16	55	E21	88
E2	74	E7	52	E12	66	E17	61	E22	68
E3	42	E8	41	E13	49	E18	72	E23	90
E4	47	E9	57	E14	59	E19	56	E24	63
E5	50	E10	65	E15	60	E20	79	E25	69

5.2. Tableau de dénombrement

Il est de la forme –Tableau IX) :

Tableau IX : Tableau de dénombrement.

Modalités x_i	Effectifs n_i	Fréquences relatives f_i
x_1	n_1	f_1
x_2	n_2	f_2
x_3	n_3	f_3
.	.	.
.	.	.
.	.	.
x_k	n_k	f_k
Total	n	1

5.3. Tableau de distribution des fréquences

Il est de la forme (Tableau X):

Tableau X : Tableau de distribution.

Classes	Centre c_i	Effectifs n_i	Fréquences relatives f_i
$[e_{i-1}, e_{i-1} + a_i[$	c_1	n_1	f_1
$[e_{i-1} + a_i, e_{i-1} + 2a_i[$	c_2	n_2	f_2
$[e_{i-1} + 2a_i, e_{i-1} + 3a_i[$	c_3	n_3	f_3
.	.	.	.
.	.	.	.
.	.	.	.
	c_k	n_k	f_k
Total	-	n	1

Exemples

a. Tableau statistique relatif à un caractère qualitatif

Analyse de sang pour 100 étudiants (Tableau XI) :

Tableau XI : Tableau des résultats d'analyse du sang pour 100 étudiants

Groupe sanguin x_i	Effectifs n_i	Fréquences relatives f_i
A	40	0.40
B	43	0.43
Ab	12	0.12
O	05	0.05
Total	100	1

b. Tableau statistique relatif à un caractère quantitatif

- Cas discret

On observe 20 lots au laboratoire, on a le nombre de lapins dans chacun : 10, 1, 0, 0, 0, 10, 12, 18, 5, 5, 12, 10, 12, 12, 0, 10, 15, 10, 20, 20 (Tableau XII) :

Tableau XII : Nombre de lapins dans 20 lots au laboratoire

x_i	0	1	5	10	12	15	18	20	T
n_i	4	1	2	5	4	1	1	2	20
fi	0.20	0.05	0.10	0.25	0.20	0.05	0.05	0.10	1

- Cas continu

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

Tableau XIII : La taille de 20 étudiants (cm)

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

Application

Dans ce cas, on doit regrouper cette série en classes (Tablea XIV) :

Nombre de classe :

$$k = 1 + 3.3 \log n = 1 + 3.3 * \log 20 = 1 + 3.3 * 1.30 = 5.29 \approx 5$$

Intervalle de classe = amplitude

$$a_i = \frac{v_{\max} - v_{\min}}{k} = \frac{163 - 140}{5} = 4.6 \approx 5$$

Centres des classes

$$x_1 = \frac{e_{i-1} + e_i}{2} = \frac{140 + 145}{2} = 142.5$$

$$x_2 = \frac{e_{i-1} + e_i}{2} = \frac{145 + 150}{2} = 147.5$$

$$x_3 = \frac{e_{i-1} + e_i}{2} = \frac{1150 + 155}{2} = 152.5$$

$$x_4 = \frac{e_{i-1} + e_i}{2} = \frac{155 + 160}{2} = 157.5$$

$$x_5 = \frac{e_{i-1} + e_i}{2} = \frac{160 + 165}{2} = 162.5$$

Tableau XIV : Tableau de la distribution des fréquences des tailles des étudiants

Classes	x_i	n_i	f_i
[140-145[142.5	6	0.30
[145-150[147.5	3	0.15
[150-155[152.5	5	0.25
[155-160[157.5	5	0.25
[160-165[162.5	1	0.05
Total	-	20	1

6. Représentations graphiques

6.1. Cas qualitatif

6.1.1. Diagramme en tuyaux d'orgue

Le diagramme en bâtonnets (ou tuyaux d'orgue) est une représentation graphique de la distribution de fréquences d'une variable qualitative. Les «bâtonnets» sont bien séparés pour indiquer les différentes catégories. La hauteur d'un bâtonnet est proportionnelle à la fréquence de la catégorie correspondante (Figure 1).

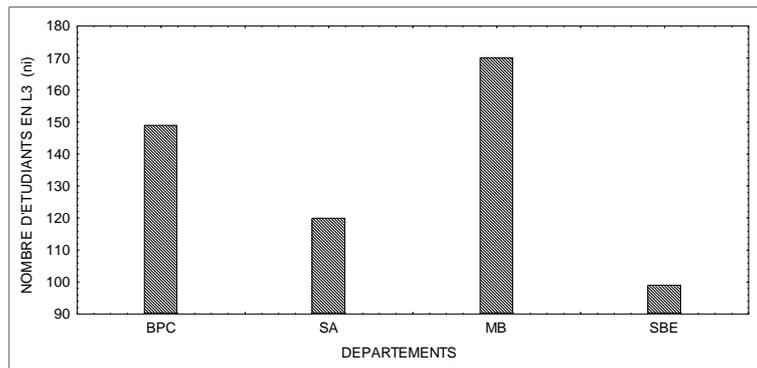


Figure 1 : Tuyaux d'orgue illustrant le nombre d'étudiants en L3 par département (**SA** : Science Alimentaire, **MB** : Microbiologie, **SBE** : Science Biologique de l'Environnement, **BPC** : Biologie Physico-Chimique).

6.1.2. Le camembert

Dans le diagramme circulaire, chaque secteur a une surface proportionnelle à la fréquence de chaque modalité (Figure 2).

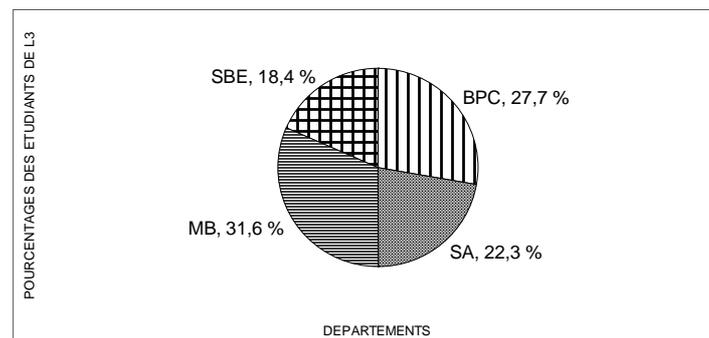


Figure 2 : Le camembert illustrant le nombre d'étudiant en L3 en fonction des départements (**SA** : Science Alimentaire, **MB** : Microbiologie, **SBE** : Science Biologique de l'Environnement, **BPC** : Biologie Physico-Chimique).

6.2.Cas quantitatif

Il existe deux types de représentation graphique d'une distribution statistique à caractère quantitatif :

- ✓ Le diagramme différentiel correspond à une représentation des effectifs ou des fréquences.
- ✓ Le diagramme intégral correspond à une représentation des effectifs cumulés, ou des fréquences cumulées.

6.2.1. Variable statistique discrète.

a. Diagramme différentiel

Diagramme en bâtons : est réalisé en fonction des effectifs ou des fréquences.

La différence avec le cas qualitatif consiste en ce que les abscisses ici sont les valeurs de la variable statistique (Figure 3).

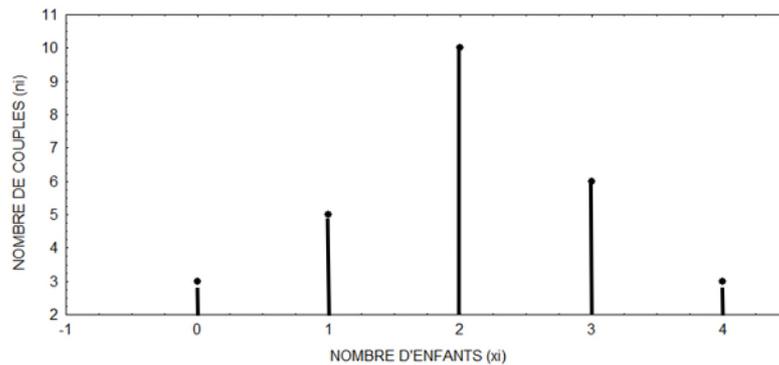


Figure 3 : Diagramme différentiel illustrant le nombre de couples en fonction du nombre d'enfants.

b. Diagramme intégral

Courbe en escalier : est réalisée en fonction des effectifs cumulés ou des fréquences cumulées. Dans cette représentation les effectifs ou les fréquences des diverses valeurs de la variable statistique correspondent aux hauteurs des marches de la courbe (Tableau 4).

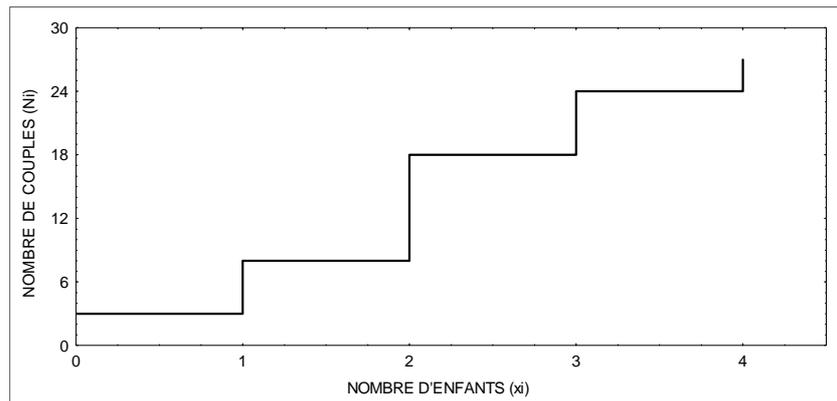


Figure 4 : Diagramme intégral illustrant le nombre de couples en fonction du nombre d'enfants.

6.2.2. Variable statistique continue

a. Diagramme différentiel

Histogramme et polygone des effectifs ou des fréquences

L'*histogramme* est une représentation graphique (en tuyaux d'orgue) de la distribution des effectifs ou des fréquences d'une variable quantitative. Souvent, les «tuyaux» sont accolés pour montrer la continuité de la variable. La hauteur du tuyau est proportionnelle à l'effectif ou la fréquence de la classe correspondante (Figure 5).

Le *polygone des effectifs ou des fréquences* : est une autre représentation graphique (en ligne brisée) de la distribution des effectifs ou des fréquences d'une variable quantitative.

Pour tracer le polygone, on joint les points milieu du sommet des rectangles adjacents par un segment de droite. Le polygone est fermé aux deux bouts en le prolongeant sur l'axe horizontal.

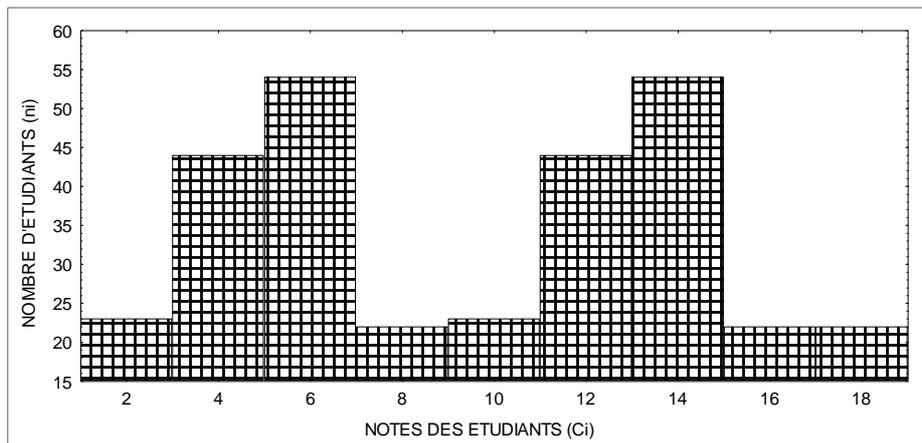


Figure 5 : Histogramme illustrant le nombre d'étudiants en fonction de leurs notes.

b. Diagramme intégral

Représentation graphique intégrale en fonction des effectifs cumulés ou des fréquences cumulées appelée parfois ogive (Figure 6).

Une telle figure fournit des données descriptives intéressantes telles que la valeur médiane, i.e. le point séparant le groupe en deux parties égales.

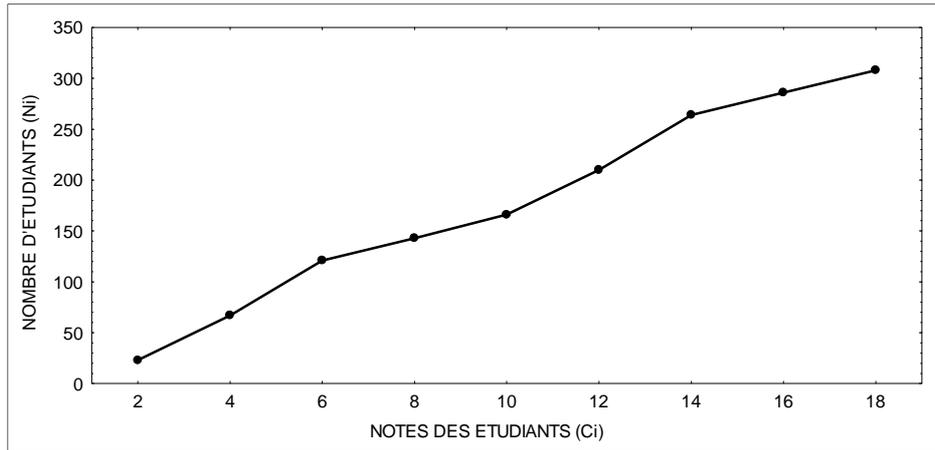


Figure 6 : Diagramme intégral illustrant le nombre d'étudiants en fonction de leurs notes.

7. Description numérique

L'objectif de la description numérique est de résumer en un seul nombre la valeur typique.

En statistiques, les distributions de variables sont caractérisées à travers trois critères, qui suffisent généralement : *Forme de la distribution, tendance centrale et dispersion.*

7.1. Tendance centrale ou Paramètres de position

7.1.1. Le mode

Le mode est la valeur la plus fréquente d'une distribution. Il se calcule toujours à partir d'un dénombrement des modalités du caractère. Comme pour le tableau de dénombrement, il faut distinguer le cas des caractères discrets et des caractères continus.

a. Caractère discret

Pour un caractère quantitatif discret le mode est la modalité qui a la fréquence la plus élevée (ou l'effectif le plus élevé).

Exemple :

- ✓ 10, 11, 12, 10, 10, 10, 9, 14 → Mode : 10 (4 fois) → Distribution unimodale.
- ✓ 10, 11, 12, 10, 10, 12, 12, 9, 14 → Modes : 10 (3 fois) et 12 (3 fois) → Distribution bimodale.

b. Caractère quantitatif continu

Approche 1

Les modalités étant en nombre infini, il est peu probable que deux éléments aient la même valeur. Dans ce cas, le mode ne peut pas être défini directement, il faut au préalable établir une partition en classes. Le mode est alors le centre de la classe modale, c'est à dire la classe qui a l'effectif le plus élevé ou la fréquence la plus élevée (Figure 7).

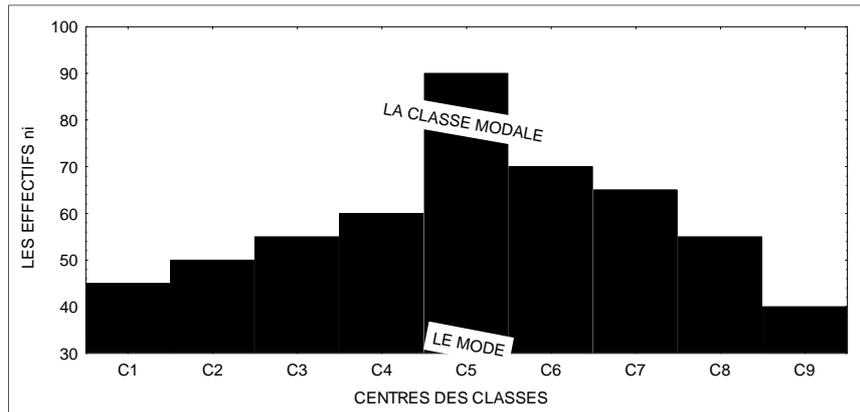


Figure 7 : Histogramme précisant la classe modale et le mode par la 1^{ère} approche.

Approche 2

Dans ce cas on peut calculer le mode par une deuxième méthode appelée « le calcul approché », en utilisant la formule :

$$Mo = e_{i-1} + a_i \frac{\Delta_1}{\Delta_1 + \Delta_2} \dots \dots \dots \text{Formule 5}$$

Δ_1 : C'est l'excès de la classe modale par rapport à la classe précédente.

Δ_2 : C'est l'excès de la classe modale par rapport à la classe suivante.

e_{i-1} : C'est la borne inférieure de la classe modale.

a_i : C'est l'amplitude

La figure 8 explique les différents paramètres de cette formule.

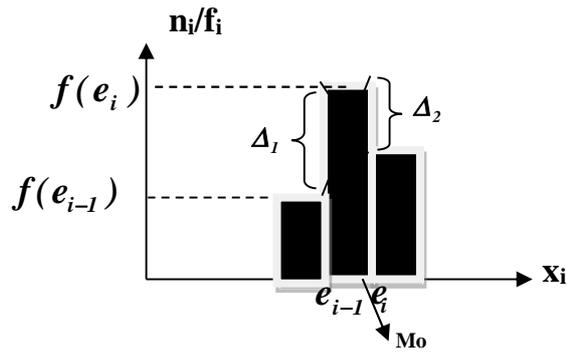


Figure 8 : Histogramme précisant le calcul du mode par la 2^{ème} approche.

7.1.2. La médiane

Les valeurs étant classées par ordre croissant, la médiane est la valeur du caractère qui partage celui-ci en deux ensembles d'effectifs égaux : 50 % des valeurs lui sont supérieures et 50 % lui sont inférieures.

a. Cas discret

n est impair

$$n = 2p + 1 \dots\dots\dots \text{Formule 6}$$

$$Me = x_{(p+1)} \dots\dots\dots \text{Formule 7}$$

p : est l'ordre de la variable dans la série statistique supposée ordonnée dans l'ordre croissant ou décroissant.

n est pair

$$n = 2p \dots\dots\dots \text{Formule 8}$$

$$Me = [x_{(p)} + x_{(p+1)}] / 2 \dots\dots\dots \text{Formule 9}$$

Exemple :

- ✓ 0, 0, 0, 0, 1, 5,5, 10, 10, 10, 10, 10, 12, 12, 12, 12, 15, 15, 18, 20, 20.
- ✓ n=20 (pair) : n=2*p → p=20/2=10.
- ✓ Me=[x₁₀+x₁₁] / 2 → Me= [10+10] / 2=10.

b. Cas continu

Si la variable statistique est continue, on calcule la valeur de la variable correspondante à la fréquence relative cumulée.

Comme la fonction $F(x)$ est continue et monotone entre 0 et 1, alors l'équation $F(x)=1/2$ admet une racine unique.

Approche 1 : Calcul géométrique (Figure 9)

Supposons que la classe médiane est $[e_{i-1}, e_i[$ on a :

$$F(e_{i-1}) < 1/2 < F(e_i)$$

En application du théorème de Thalès on trouve

$$\frac{BC}{D} = \frac{AB}{AD} \Rightarrow BC = \frac{AB * DE}{AD}$$

Par interpolation, on aura :

$$Me = e_{i-1} + BC = e_{i-1} + \frac{AB * DE}{AD}$$

On a : $AB = 1/2 - F(e_{i-1})$ et $DE = a_i$ et $AD = f_i$ (f_i est la fréquence de $[e_{i-1}, e_i[$)

Alors:

$$Me = e_{i-1} + a_i \frac{1/2 - F(e_{i-1})}{f_i} \dots\dots\dots \text{Formule 10}$$

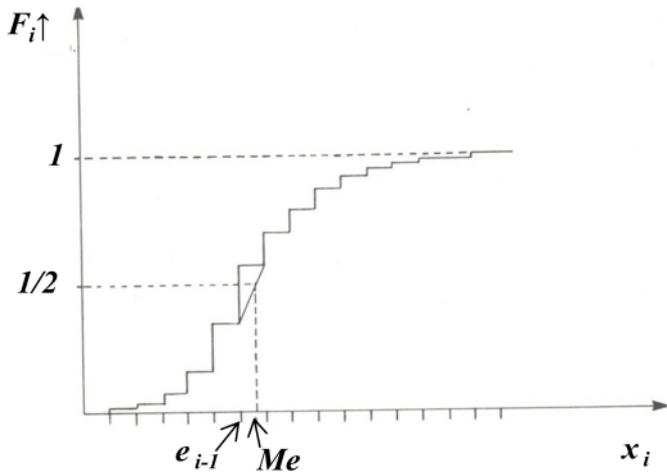
On peut également exprimer f_i ainsi :

$$f_i = F(e_i) - F(e_{i-1})$$

La médiane alors prend la formule :

$$Me = e_{i-1} + a_i \frac{1/2 - F(e_{i-1})}{F(e_i) - F(e_{i-1})} \dots\dots\dots \text{Formule 11}$$

A



B

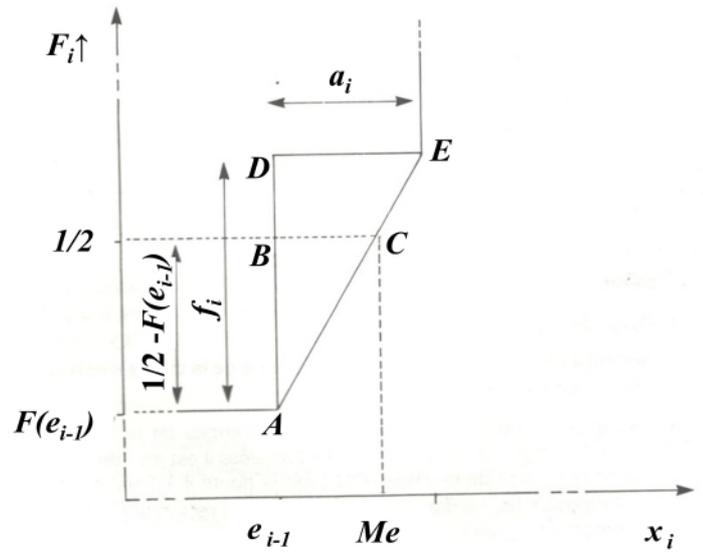
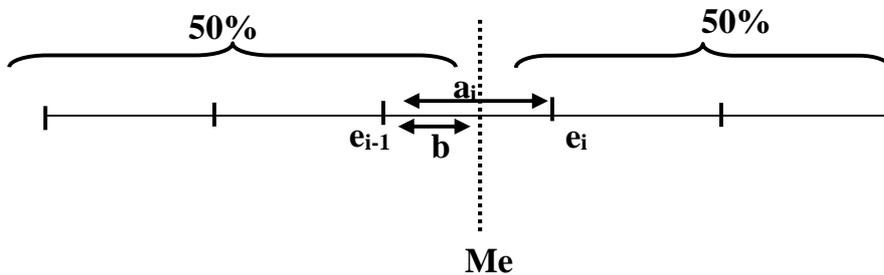


Figure 9 : **A :** Histogramme des fréquences cumulées. **B :** Agrandissement de la partie de l’histogramme renfermant la médiane.

Approche 2 : Calcul par la méthode d’interpolation

$Me = e_{i-1} + b$ Formule 12

Supposons $[e_{i-1}, e_i]$ la classe médiane :



$n_i \rightarrow a_i \Rightarrow b = \frac{a_i * n_i'}{n_i}$ Formule 13

$n_i' \rightarrow b$

7.1.3. La moyenne

La moyenne est la valeur unique que devraient avoir tous les individus d'une population pour que leur total soit inchangé. C'est un critère de position. Il s'agit d'une mesure statistique caractérisant les éléments d'un ensemble de quantités et elle exprime la somme de toutes les mesures divisée par l'effectif total de l'échantillon étudié ou bien une valeur telle qu'en la multipliant par l'effectif total on trouve la somme des mesures effectuées sur toute la population.

Il est à signaler que les notations de la moyenne diffèrent selon que l'étude s'effectue sur une population ou un échantillon (ceci est détaillé dans la section qui suit).

a. Cas discret

- *Etude de la moyenne d'une population*

Soit X une variable statistique discrète pouvant prendre les valeurs X_1, X_2, \dots, X_k auxquelles correspondent les effectifs N_1, N_2, \dots, N_k . La moyenne notée μ est de forme :

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i = \sum_{i=1}^N X_i \dots \dots \dots \text{Formule 14}$$

- *Etude de la moyenne d'une population*

Soit x une variable statistique discrète pouvant prendre les valeurs x_1, x_2, \dots, x_k auxquelles correspondent les effectifs n_1, n_2, \dots, n_k avec $\sum_{i=1}^k n_i = n$

La moyenne notée m est de forme :

$$m = \frac{1}{n} \sum_{i=1}^n x_i = \sum_{i=1}^n x_i \dots \dots \dots \text{Formule 15}$$

b. Cas continu

Dans ce cas on utilise les mêmes formule que dans la cas précédent sauf que les modalités seront les centres de classes.

avec : $c_i = \frac{e_{i-1} + e_i}{2}$ est le centre de la classe $[e_{i-1}, e_i[$ et n_i son effectif.

Remarque : Comparaison des valeurs centrales

Il n'y a pas de règle générale pour déterminer laquelle des mesures de tendance centrale est la plus pertinente pour caractériser une distribution.

Exemple :

Les nombres ci-dessous représentent le nombre d'enfants dans chacune de dix familles choisies au hasard : 3, 8, 1, 1, 4, 5, 2, 3, 0, 1 (il s'agit d'étude d'un échantillon).

✓ **Moyenne**

$$m = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10}}{n} = 1/10 [(3+8+1+1+4+5+2+3+0+1)] = 28 /$$

10

$$m = 2,8$$

✓ **Médiane**

Rangeons les données en ordre ascendant :

0, 1, 1, 1, 2, 3, 3, 4, 5, 8

Comme $n = 10$ (donc pair), $p = 5$

$$Me = (5^{\text{ème}} \text{ Observation} + 6^{\text{ème}} \text{ Observation}) / 2 = (2 + 3) / 2 = 2,5$$

✓ **Mode**

On voit immédiatement que $Mo = 1$

On remarque que les trois mesures sont différentes.

Laquelle est la plus «représentative»? (Voir ce qui suit.)

7.4.Valeurs centrales et forme des distributions

La comparaison des trois valeurs centrales ou l'examen de l'histogramme permettent de définir la forme des distributions. Selon la forme d'une distribution le meilleur résumé sera fourni par l'une ou l'autre des trois valeurs centrales.

7.4.1. Distributions bimodales ou multimodales

La distribution comporte plusieurs modes. Le mode principale est différent de la moyenne et de la médiane. Dans ce cas, ni la moyenne ni la médiane ne sont significatives.

Le meilleur résumé est alors donné par les modes principaux et secondaires de la distribution.

7.4.2. Distributions unimodales symétriques

Lorsque la distribution est unimodale et symétrique, on va trouver à peu près moyenne = médiane = mode.

Le meilleur résumé est alors donné par la moyenne car elle tient compte de toutes les observations et elle possède des propriétés statistiques intéressantes.

7.4.3. Distributions unimodales dissymétriques

Deux cas peuvent se présenter :

* $\text{mode} < \text{médiane} < \text{moyenne}$: la distribution est dissymétrique à gauche (ou la distribution est positivement dissymétrique (Figure 10 A). c'est à dire qu'il y a concentration pour les valeurs faibles et dispersion pour les valeurs fortes. C'est le cas le plus fréquent

* $\text{moyenne} < \text{médiane} < \text{mode}$: la distribution est dissymétrique à droite (ou la distribution est négativement dissymétrique (voir la distribution B), c'est à dire qu'il y a concentration pour les valeurs élevées et dispersion pour les valeurs faibles. Ce cas est plus rare que le précédent.

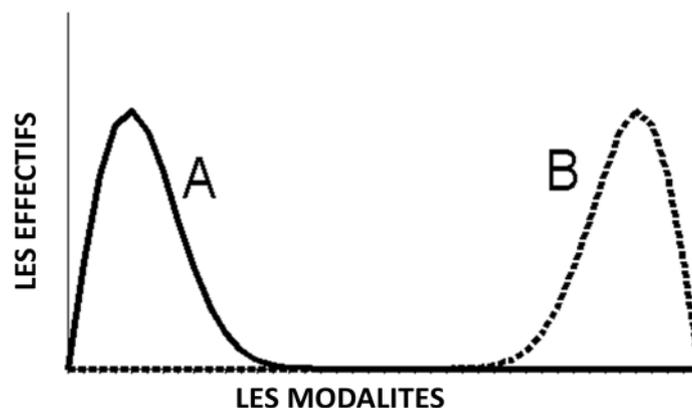


Figure 10 : Courbe illustrant les deux types de distribution. A : ditribution dissymétrique à gauche. B : distribution dissymétrique à droite.

Remarque :

Lorsqu'il y a une dissymétrie marquée, la médiane est généralement préférable à la moyenne car elle est moins influencée par les valeurs exceptionnelles qui sont souvent à l'origine de la dissymétrie. Si la dissymétrie est peu marquée, on pourra néanmoins utiliser la moyenne.

7.5. Paramètres de dispersion absolue

7.5.1. La fourchette des valeurs ou étendue

L'étendue E est la différence entre la plus grande valeur et la plus petite valeur. La formule générale est :

$$E = V_{\max} - V_{\min} \dots\dots\dots \text{Formule 16}$$

Où :

V_{\max} : C'est la valeur maximale.

V_{\min} : C'est la valeur minimale.

Remarque :

Peu intéressante, car ne prend en compte que les valeurs extrêmes.

Exemple :

1, 10, 11, 12, 10, 11, 12, 13, 11, 12, 11, 11, 12, 11, 12, 12, 19

$$E = 19 - 1 = 18$$

Valeur peu représentative.

En réalité, il existe une faible dispersion.

7.5.2. Ecart absolu moyen et l'écart absolu médian

L'écart absolu moyen est la moyenne de la valeur absolue des écarts à la moyenne. On peut calculer de la même manière l'écart absolu médian qui est la moyenne des écarts à la médiane. Formules de l'écart absolu moyen et de l'écart absolu médian.

$$EAM_{\sigma} = 1/n \sum n_i (x_i - m) \dots\dots\dots \text{Formule 17}$$

$$EAM_{\theta} = 1/n \sum n_i (x_i - Me) \dots\dots\dots \text{Formule 18}$$

7.5.3. Variance et écart type

On peut citer deux types de variances : σ^2 et S^2 .

La variance dans la population « σ^2 » : elle est immuable, on ne la connaît pas et on ne peut pas la déterminer avec précision, puisqu'il faudrait avoir un échantillon contenant toute la population à étudier. Puisqu'on ne peut pas la trouver, on se contente de l'estimer, à partir d'un échantillon.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \dots\dots\dots \text{Formule 19}$$

$$\sigma = \sqrt{\sigma^2} \dots\dots\dots \text{Formule 20}$$

L'estimateur de σ^2 « S^2 » : il s'agit-là d'une valeur propre à un échantillon ; elle est différente de σ^2 , mais tend à s'en approcher (plus l'échantillon est représentatif, et plus les deux valeurs sont proches).

S^2 exprime la moyenne des carrés des écarts entre chaque observation et la moyenne des observations.

L'écart type, noté s , est simplement la racine carrée de la variance s^2 .

Les formules sont :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \dots\dots\dots \text{Formule 21}$$

$$s = \sqrt{s^2} \dots\dots\dots \text{Formule 22}$$

Remarque

Si on choisit un individu de manière aléatoire dans une population et que l'on relève une valeur numérique sur cet individu, les valeurs possibles sont les valeurs présentes dans la population (et les probabilités associées sont les fréquences dans la population). De ce fait, la moyenne μ et l'écart-type σ des valeurs possibles sont égales à la moyenne \bar{x} et à l'écart-type s des valeurs prises par les individus de la population.

7.5.4. L'étendue interquartile

Les quartiles sont les trois valeurs qui permettent de découper la distribution en quatre classes d'effectifs égaux, on les note Q_1 , Q_2 et Q_3 .

Le premier quartile, noté Q_1 , est le point qui sépare la portion de 25% des valeurs les plus petites de la portion de 75% des valeurs les plus grandes.

Le troisième quartile, noté Q_3 , est le point qui sépare la portion de 25% des valeurs les plus grandes de la portion de 75% des valeurs les plus petites.

Le deuxième quartile Q_2 correspond tout simplement à la médiane.

L'intervalle interquartile (Q_3-Q_1) est un paramètre de dispersion absolue qui correspond à l'étendue de la distribution une fois que l'on a retiré les 25% des valeurs les plus faibles et les 25% des valeurs les plus fortes. 50% des observations sont donc concentrées entre Q_1 et Q_3 .

$$Iq = Q_3 - Q_1 \dots\dots\dots \text{Formule 23}$$

L'intervalle semi-interquartile (Q) correspond à la moitié de l'étendue interquartile. En formule :

$$Q = (Q_3 - Q_1) / 2 \dots\dots\dots \text{Formule 24}$$

Remarque :

Q_1 et Q_3 sont obtenus par les mêmes méthodes exposées plus haut pour la médiane.

7.6. Paramètres de dispersion relatifs

7.6.1. Le coefficient de variation

Le coefficient de variation noté CV est une mesure relative de l'écart type qui permet de prendre en compte l'ordre de grandeur de la moyenne. Autrement dit, ce coefficient correspond à l'écart type de la distribution exprimé en pourcentage de la moyenne de la distribution

La formule générale est :

$$CV = (s / m) * 100 \dots\dots\dots \text{Formule 25}$$

L'écart-type seul ne permet le plus souvent pas de juger de la dispersion des valeurs autour de la moyenne. Si par exemple une distribution a une moyenne de 0,1 et un écart-type de 1, elle sera beaucoup plus dispersée (CV de 1000%) qu'une distribution de moyenne 1000 et d'écart-type 10 (CV de 1%).

Ce nombre est sans unité, c'est une des raisons pour lesquelles il est parfois préféré à l'écart type. En effet, pour comparer deux séries de données d'unités différentes, l'utilisation du coefficient de variation est plus judicieuse.

7.6.2. Le coefficient interquartile relatif

C'est le rapport de l'écart interquartile par rapport à la médiane. La formule :

$$CIQ = (Q_3 - Q_1) / Q_2 \dots\dots\dots \text{Formule 26}$$

7.6.3. Ecart relatif moyen

C'est le rapport de l'écart absolu moyen par rapport à la moyenne. La formule :

$$ERM = EAM / m \dots\dots\dots \text{Formule 25}$$

7.6.4. Ecart relatif médian

C'est le rapport de l'écart absolu médian par rapport à la médiane. La formule :

$$ERMe = EAM / Me \dots\dots\dots \text{Formule 26}$$

7.5. Paramètres de dispersion absolue et valeurs centrales

Le résumé d'une distribution par une valeur centrale est souvent trompeur parce qu'il est incomplet : on connaît l'ordre de grandeur des valeurs mais on ignore la dispersion des valeurs autour de la valeur centrale de référence. Le premier rôle des paramètres de dispersion est donc :

D'accompagner et de préciser les résumés de distribution effectués à l'aide des valeurs centrales.

Un bon résumé statistique doit donc toujours comporter au moins deux paramètres : une valeur centrale et un paramètre de dispersion. L'appariement des valeurs centrales et des paramètres de dispersion ne peut toutefois pas s'effectuer de n'importe quelle manière et certaines associations peuvent se faire de façon privilégiée.

La moyenne peut-être accompagnée soit de l'écart-type, soit de de l'écart absolu moyen (puisque ces paramètres sont calculés par rapport à la moyenne).

La médiane peut-être accompagnée soit de l'écart absolu médian, soit de l'intervalle interquartile ou de l'intervalle inter décile.

Le mode, qui n'est véritablement utile que dans le cas des distributions multimodales, peut être accompagné de l'étendue (qui est le seul paramètre de dispersion conservant une signification dans le cas des distributions multimodales).

7.6. Les paramètres de forme

7.6.1. Le Coefficient d'asymétrie

Il s'agit de l'étalement à gauche ou à droite de la représentation d'une série statistique telle vue précédemment lors d'une explication en fonction des paramètres de tendance centrale ou de position

Le coefficient d'asymétrie est calculé ainsi :

$$S = \frac{m_3}{s^3} = \left(\frac{1}{s^3} \right) \frac{\sum_{i=1}^n (x_i - m)^3}{n} \dots\dots\dots \text{Formule 27}$$

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^3 \dots\dots\dots \text{Formule 28}$$

Le coefficient S évalue le défaut de symétrie d'une distribution (Voir figure 10 A et B).

Remarque :

- ✓ S est nul pour une distribution symétrique (une distribution normale).
- ✓ S est positif pour une distribution étalée à droite.
- ✓ S est négatif pour une distribution étalée à gauche.

7.6.2. Le coefficient d'aplatissement

Le coefficient d'aplatissement (K) évalue la dispersion des valeurs extrêmes par référence à la loi normale. Il permet d'étudier la forme plus ou moins pointue ou aplatie.

La moyenne centrée d'ordre 4 permet de calculer le degré d'aplatissement d'une distribution à une variable. Afin d'obtenir un nombre sans dimension, on le divise par le carré de la variance. L'indicateur obtenu est appelé coefficient d'aplatissement K de Pearson, ou kurtosis, sa formule est la suivante :

$$K - \frac{m_4}{s^4} = \left(\frac{1}{s^4} \right) \frac{\sum_{i=1}^n (x_i - m)^4}{n} \dots\dots\dots \text{Formule 29}$$

$$m_4 = \frac{1}{n} \sum_{i=1}^n (x_i - m)^4 \dots\dots\dots \text{Formule 30}$$

La kurtosis d'une loi normale (de Gauss) est égale à 3. Certains auteurs et logiciels retranchent 3 à la forme ci-dessus, ce qui conforte la loi normale dans son rôle de « loi étalon ». Le coefficient obtenu K' est dit « de Fisher » (ou *excess kurtosis*).

$$K' = K - 3 \dots\dots\dots\text{Formule 31}$$

Remarque :

- ✓ Le K' est grand quand il y a beaucoup de valeurs éloignées de la moyenne.
- ✓ $K' > 0$: indique que les queues comptent plus d'observations que dans une distribution normale.
- ✓ $K' < 0$: indique que les queues comptent moins d'observations que dans une distribution normale.
- ✓ $K'=0$: indique une loi normale.

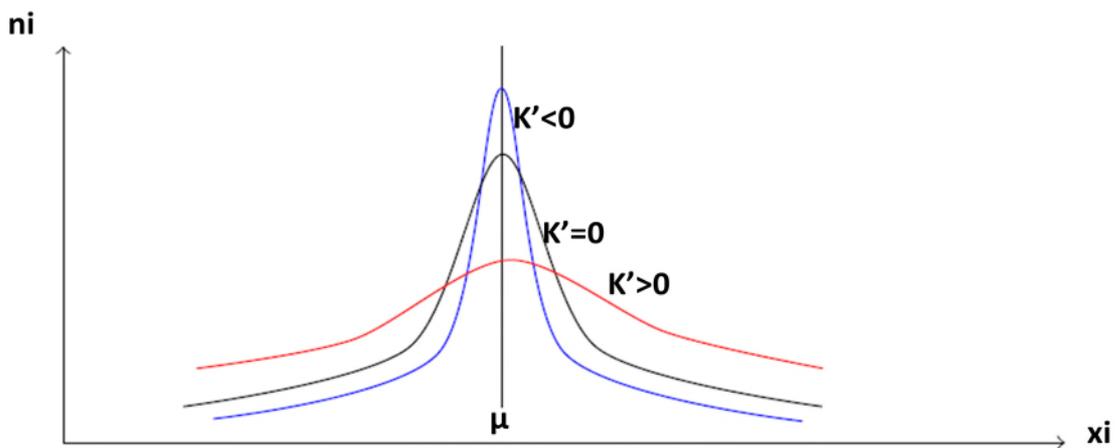


Figure 11 : Courbe illustrant les trois formes de distribution générées par la variation du coefficient d'aplatissement.

8. Notion de la loi normale

La loi normale est caractérisée par un coefficient d'asymétrie et un coefficient d'aplatissement nuls.

Soit X une variable, μ est une moyenne, σ un écart-type et Z la variable définie comme suit :

$$Z = (X - \mu) / \sigma \dots\dots\dots\text{Formule 31}$$

Si Z suit une loi normale, alors X suit une loi normale d'une moyenne μ et d'écart-type σ notée :

$$N - (\mu; \sigma)$$

En termes de variance, la fonction de densité de la loi normale s'exprime ainsi :

$$f(X) = (1/\sigma\sqrt{2\pi})\exp^{-1/2(x-\mu/\sigma)^2} \dots\dots\dots\text{Formule 32}$$

La loi normale centrée réduite est donc la loi normale dont les paramètres sont $\mu = 0$ et $\sigma = 1$.

L'allure de la courbe représentative d'une fonction de densité de loi normale centrée réduite est donnée ci-après, accompagnée des courbes représentant des densités de lois normales ayant d'autres écart-types pour une même moyenne nulle (la loi normale d'écart-type 2 ; la loi d'écart-type 0,5).

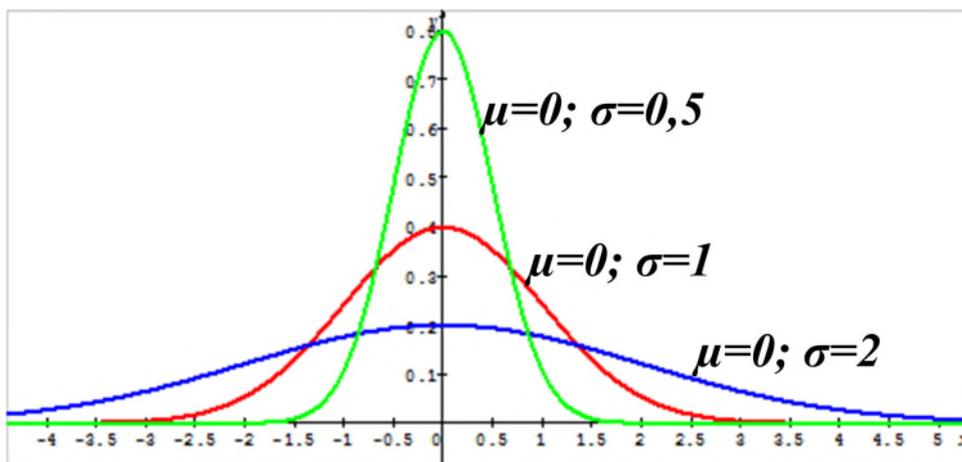
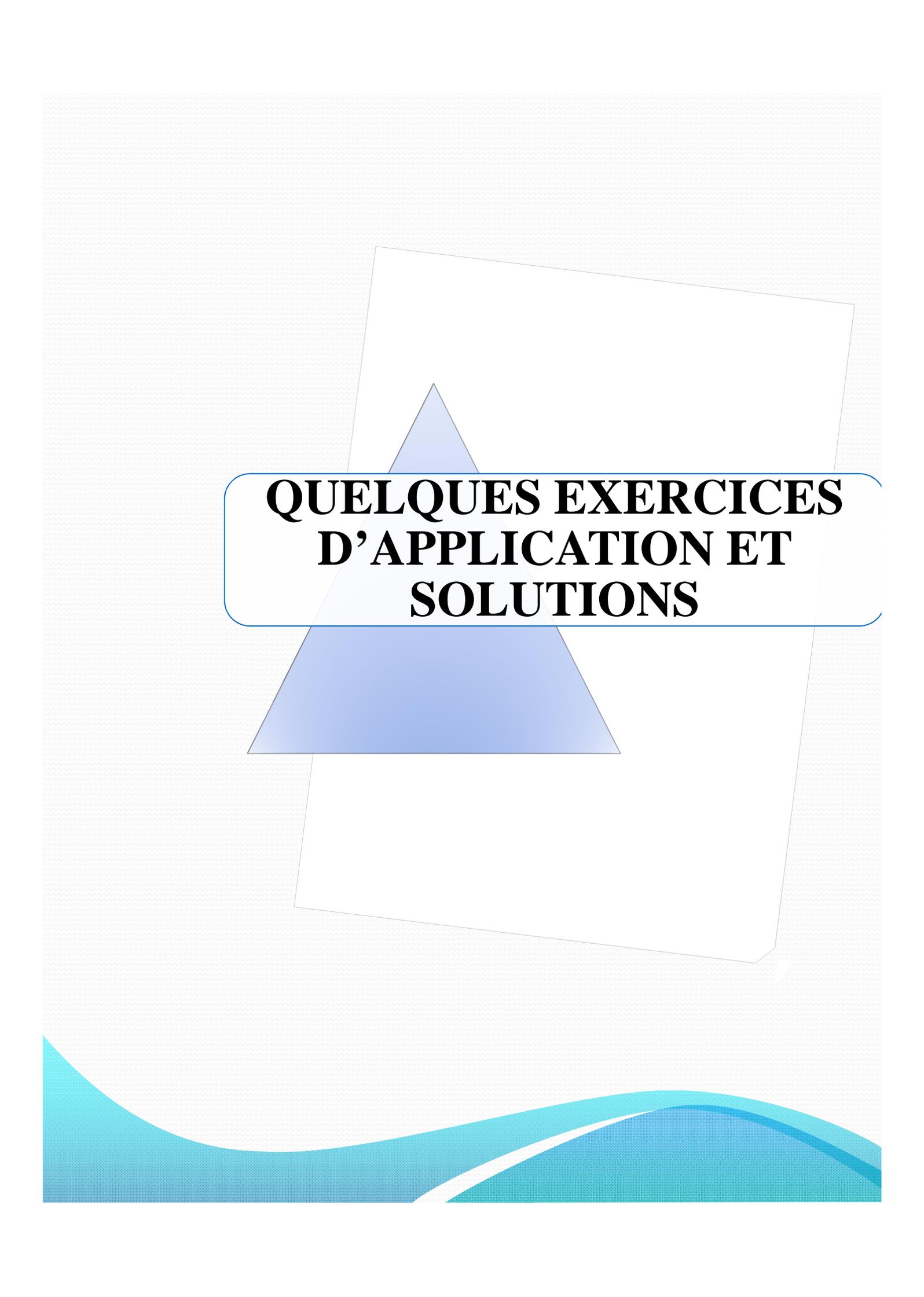


Figure 12 : Courbe illustrant les trois différentes allures des fonctions de densités des lois normales.

On remarque que plus la dispersion est élevée, plus la courbe est aplatie et plus elle est faible, plus la courbe est resserrée.



**QUELQUES EXERCICES
D'APPLICATION ET
SOLUTIONS**

QUELQUES EXERCICES D'APPLICATION

EXERCICE N°1

Un expérimentateur avait extrait à partir de 5 plantes les doses ($\mu\text{g/g}$) de polyphénols suivantes: **13, 15, 13, 14, 15**.

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Tracer le diagramme intégral par l'utilisation des effectifs cumulés.

SOLUTION

1. Les notions de base :

Population statistique	:	Les plantes
Unité statistique	:	Une plante
Echantillon	:	5 plantes ($n=5$)
Caractère	:	doses en polyphénols ($\mu\text{g/g}$)
Nature du caractère	:	Caractère quantitatif (continu, mesurable, de stock) assimilé à un cas discret.
Modalités	:	13, 14, 15
Taille (Cardinal)	:	5

2. Tableau statistique

Tableau I : Tableau de dénombrement des doses en polyphénols dans un échantillon de 5 plantes.

x_i	n_i	f_i	$N_i\uparrow$	$F_i\uparrow$
13	2	0,4	2	0,4
14	1	0,2	3	0,6
15	2	0,4	5	1
T	5	1	-	-

3. Tracer le diagramme intégral en utilisant les effectifs cumulés.

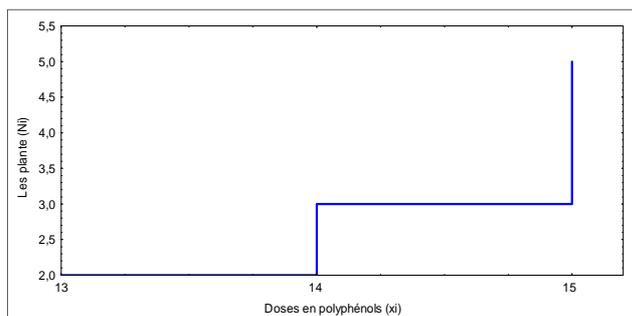


Figure 1 : Diagramme intégral des plantes en fonction des doses en polyphénols.

EXERCICE N°2

Un généticien s'est intéressé au séquençage de l'ADN de 10 espèces de poissons d'eau douce, les résultats obtenus sont : **AATCGCC**, **AACTCGG**, **AATCGCC**, **ACGTAAT**, **TCGATCG**, **AACTCGG**, **AATCGCC**, **TCGATCG**, **ACGTAAT**, **AATCGCC**.

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Tracer le graphe correspondant.

SOLUTION

1. Les notions de base :

Population statistique	: Les espèces de poissons d'eau douce
Unité statistique	: Une espèce
Echantillon	: 10 espèces (n=10)
Caractère	: Les séquences d'ADN
Nature du caractère	: Caractère qualitatif nominal
Modalités	: AATCGCC ; AACTCGG ; ACGTAAT ; TCGATCG
Taille (Cardinal)	: 10

2. Tableau statistique

Tableau II : Tableau de dénombrement des séquences d'ADN de 10 espèces de poissons d'eau douce.

x_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
AATCGCC	4	0,4	4	0,4
AACTCGG	2	0,2	6	0,6
ACGTAAT	2	0,2	8	0,8
TCGATCG	2	0,2	10	1
T	10	1	-	-

3. Tracer le diagramme intégral en utilisant les effectifs cumulés.

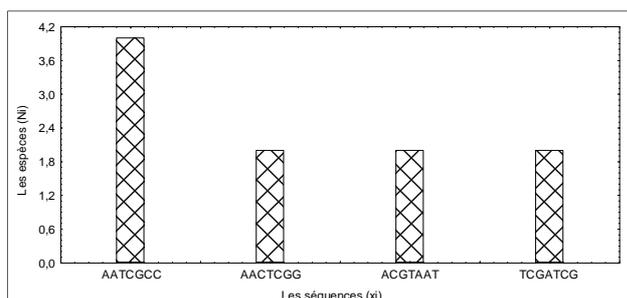


Figure 2 : Tuyaux d'orgue des espèces de poissons d'eau douce en fonction des séquences d'ADN.

EXERCICE N°3

Un enseignant a voulu savoir le niveau des étudiants de Biochimie en module de BMGG, pour ce faire il s'est intéressé aux notes obtenues au premier semestre par 50 étudiants. Les résultats sont : **12 / 10 / 12 / 13 / 10 / 15 / 14 / 10 / 11 / 11 / 12 / 15 / 12 / 10 / 12 / 13 / 10 / 15 / 14 / 10 / 12 / 10 / 12 / 13 / 10 / 15 / 14 / 10 / 11 / 11 / 12 / 15 / 12 / 10 / 12 / 13 / 10 / 15 / 14 / 10 / 11 / 12 / 13 / 12 / 10 / 11 / 12 / 13 / 12 / 10.**

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Tracer le diagramme intégral en utilisant les effectifs cumulés.

SOLUTION

1. Les notions de base :

Population statistique	: Les étudiants de Biochimie
Unité statistique	: Un étudiant
Echantillon	: 50 espèces (n=50)
Caractère	: Les notes de BMGG en premier semestre
Nature du caractère	: Caractère quantitatif (continue, mesurable, d'intensité) assimilé à un cas discret
Modalités	: 10 ; 11 ; 12 ; 13 ; 14 ; 15
Taille (Cardinal)	: 50

2. Tableau statistique

Tableau III : Tableau de dénombrement des notes d'examen de BMGG de 50 étudiants de Biochimie.

x_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
10	14	0,28	14	0,28
11	6	0,12	20	0,40
12	14	0,28	34	0,68
13	6	0,12	40	0,80
14	4	0,08	44	0,88
15	6	0,12	50	1
T	50	1	-	-

3. Tracer le diagramme intégral en utilisant les effectifs cumulés.

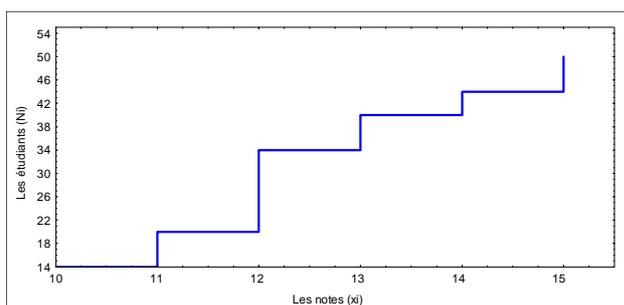


Figure 3 : Diagramme intégral des étudiants Biochimie en fonction de leurs notes en BMGG.

EXERCICE N°4

Afin de tester la toxicité d'une molécule synthétique sur les organes reproducteurs des souris blanches, un biologiste a préparé 10 lots de 10 souris chacun. Pour des doses ($\mu\text{g/g}$) différentes, il a obtenu, en fonction des lots, les taux de mortalité (%) suivants : **0, 12, 0, 50, 12, 50, 100, 12, 50, 100**.

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Tracer le diagramme intégral en utilisant les effectifs cumulés.

SOLUTION

1. Les notions de base :

Population statistique	:	Les lots de souris
Unité statistique	:	Un lot de souris
Echantillon	:	10 lots de souris
Caractère	:	Taux de mortalité
Nature du caractère	:	Caractère quantitatif (continue, mesurable, d'intensité) assimilé à un cas discret
Modalités	:	0 ; 12 ; 50 ; 100
Taille (Cardinal)	:	10

2. Tableau statistique

Tableau IV : Tableau de dénombrement des taux de mortalité de souris dans 10 lots de 10 souris chacun.

x_i	n_i	f_i	$N_i\uparrow$	$F_i\uparrow$
0	2	0,2	2	0,2
12	3	0,3	5	0,5
50	3	0,3	8	0,8
100	2	0,2	10	1
T	10	1	-	-

3. Tracer le diagramme intégral en utilisant les effectifs cumulés.

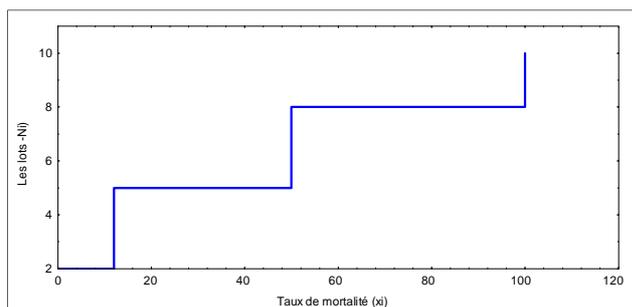


Figure 4 : Diagramme intégral des lots de souris en fonction de leurs taux de mortalité.

EXERCICE N°5

Dans un laboratoire de Biochimie, 20 étudiants ont élaboré une liste d'espèce de plantes qui seront sujettes d'une extraction des huiles essentielles, les résultats en nombre sont comme suit : 75 / 69 / 75 / 96 / 80 / 102 / 102 / 69 / 75 / 75 / 90 / 80 / 96 / 80 / 96 / 80 / 96 / 80 / 90 / 102.

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Calculer les effectifs cumulatifs croissants et les fréquences cumulatives croissantes et tracer une courbe en fonction de l'un des paramètres.
4. Calculer les paramètres de position (moyenne, mode, médiane)
5. Calculer les paramètres de dispersion (variance et écart-type)
6. Déduire à partir des paramètres de position le type de la distribution.

SOLUTION

1. Les notions de base :

Population statistique	: Les étudiants de Biochimie
Unité statistique	: Un étudiant
Echantillon	: 20 étudiants
Caractère	: Nombre de plantes à étudier au laboratoire
Nature du caractère	: Caractère discret, mesurable, de stock.
Modalités	: 69 ; 75 ; 80 ; 90 ; 96 ; 102
Taille (Cardinal)	: 20

2. Tableau statistique et calcul des cumuls

Tableau V : Tableau de dénombrement des doses en polyphénols de cinq plantes.

x_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
69	2	0,1	2	0,1
75	4	0,2	6	0,3
80	5	0,25	11	0,55
90	2	0,1	13	0,65
96	4	0,2	17	0,85
102	3	0,15	20	1
T	20	1	-	-

3. Représentation graphique

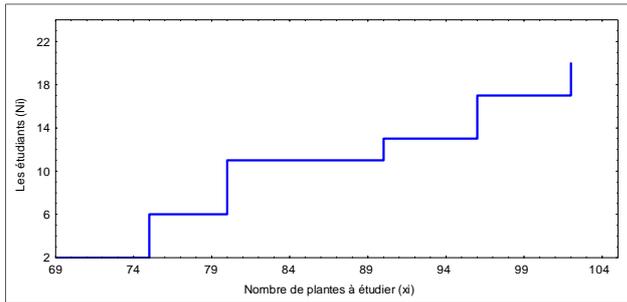


Figure 5 : Diagramme intégral des étudiants en fonction du nombre de plantes à étudier.

4. Calcul des paramètres de position

a. La moyenne

$$m = [(0,1 \times 69) + (0,2 \times 75) + (0,25 \times 80) + (0,1 \times 90) + (0,2 \times 96) + (0,15 \times 102)]$$

$$m = 85,4 \text{ plantes}$$

b. Le mode

$$Mo = 80 \text{ plantes}$$

c. La médiane

$$P = 10$$

$$Me = [x_{(10)} + x_{(11)}] / 2$$

$$Me = (80 + 80) / 2$$

$$Me = 80 \text{ plantes}$$

5. Calcul des paramètres de dispersion

a. La variance

$$s^2 = 1/19 [2(69 - 85,4)^2 + 4(75 - 85,4)^2 + 5(80 - 85,4)^2 + 2(90 - 85,4)^2 + 4(96 - 85,4)^2 + 3(102 - 85,4)^2]$$

$$s^2 = 134,88 \text{ plantes}$$

b. Ecart-type

$$s = 11,61 \text{ plantes.}$$

6. Type de la distribution

$$m > Me = Mo$$

La distribution est positivement dissymétrique

EXERCICE N°6

Au laboratoire, lors des tests de toxicité, un biochimiste s'est intéressé au nombre de lapins décédés soumis à des concentrations différentes d'une toxine. Afin d'atteindre au parfait l'objectif recherché, ce chercheur a effectué 100 expériences. Les résultats sont les suivants :
 4 / 1 / 1 / 3 / 3 / 4 / 2 / 2 / 2 / 3 / 3 / 4 / 1 / 1 / 1 / 2 / 2 / 3 / 4 / 4 / 2 / 3 / 3 / 5 / 0 / 2 / 2 / 3 / 2 / 2 /
 3 / 2 / 3 / 2 / 0 / 3 / 2 / 2 / 1 / 2 / 4 / 3 / 4 / 2 / 3 / 2 / 2 / 2 / 1 / 1 / 4 / 3 / 2 / 1 / 3 / 5 / 4 / 1 / 2 /
 2 / 4 / 1 / 3 / 3 / 4 / 2 / 2 / 3 / 1 / 4 / 3 / 2 / 0 / 2 / 4 / 1 / 3 / 5 / 0 / 3 / 2 / 2 / 2 / 3 / 4 / 2 / 2 / 3 / 2 /
 / 3 / 3 / 4 / 1 / 3 / 2 / 3 / 2 / 3 / 1 / 2.

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Calculer les effectifs cumulatifs croissants et les fréquences cumulatives croissantes et tracer une courbe en fonction de l'un des paramètres.
4. Calculer les paramètres de position (moyenne, mode, médiane)
5. Calculer les paramètres de dispersion (variance et écart-type)
6. Dédurre à partir des paramètres de position le type de la distribution.

SOLUTION

1. Les notions de base :

Population statistique	: Les expériences
Unité statistique	: Une expérience
Echantillon	: 100 expériences
Caractère	: Nombre de mortalité
Nature du caractère	: Caractère quantitatif discret, mesurable, de stock
Modalités	: 0 ; 1 ; 2 ; 3 ; 4 ; 5
Taille (Cardinal)	: 100

2. Tableau statistique et calcul des cumuls

Tableau VI : Tableau de dénombrement des mortalités en lapins dans 100 expériences.

x_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
0	4	0,04	4	0,04
1	15	0,15	19	0,19
2	35	0,35	54	0,54
3	28	0,28	82	0,82
4	15	0,15	97	0,97
5	3	0,03	100	1
T	100	1	-	-

3. Représentation graphique

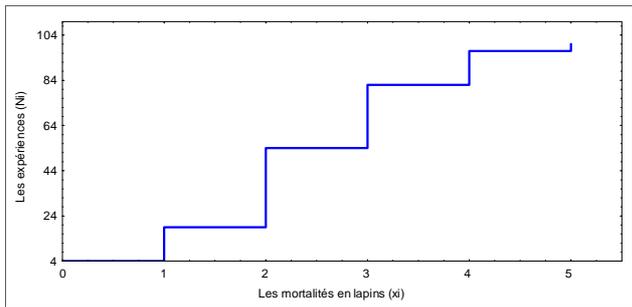


Figure 6 : Diagramme intégral des expériences en fonction du nombre de lapins décédés.

4. Calcul des paramètres de position

a. La moyenne

$$m = [(0,04 \times 0) + (0,15 \times 1) + (0,35 \times 2) + (0,28 \times 3) + (0,15 \times 4) + (0,03 \times 5)]$$

$$m = 2,44 \text{ mortalités}$$

b. Le mode

$$Mo = 2 \text{ mortalités}$$

c. La médiane

$$P = 50$$

$$Me = [X_{(50)} + X_{(51)}] / 2$$

$$Me = (2 + 2) / 2$$

$$Me = 2 \text{ mortalités}$$

5. Calcul des paramètres de dispersion

a. La variance

$$s^2 = 1/99 [4(0 - 2,44)^2 + 15(1 - 2,44)^2 + 35(2 - 2,44)^2 + 28(3 - 2,44)^2 + 15(4 - 2,44)^2 + 3(5 - 2,44)^2]$$

$$s^2 = 0,66 \text{ mortalités}$$

b. Ecart-type

$$s = 0,81 \text{ mortalités}$$

6. Type de la distribution

$$m > Me = Mo$$

La distribution est positivement dissymétrique

EXERCICE N°7

Une insémination artificielle a été exercée sur cinquante vaches, après une période il a été enregistré l'efficacité de ce dispositif par comptage du nombre de descendants. Les résultats sont : 2 / 2 / 3 / 5 / 2 / 1 / 4 / 2 / 3 / 5 / 3 / 2 / 3 / 3 / 4 / 1 / 2 / 4 / 2 / 2 / 4 / 2 / 3 / 2 / 3 / 3 / 2 / 2 / 4 / 2 / 1 / 4 / 2 / 3 / 2 / 2 / 3 / 1 / 3 / 3 / 2 / 3 / 2 / 2 / 3 / 4 / 3 / 2 / 3 / 2.

1. Faites ressortir les notions de base de la statistique.
2. Donner le tableau statistique.
3. Calculer les effectifs cumulatifs croissants et les fréquences cumulatives croissantes et tracer une courbe en fonction de l'un des paramètres.
4. Calculer les paramètres de position (moyenne, mode, médiane)
5. Calculer les paramètres de dispersion (variance et écart-type)
6. Déduire à partir des paramètres de position le type de la distribution.

SOLUTION

1. Les notions de base :

Population statistique	: Les vaches
Unité statistique	: Une vache
Echantillon	: 50 vaches
Caractère	: Nombre de descendants
Nature du caractère	: Caractère quantitatif discret, mesurable, de stock
Modalités	: 1 ; 2 ; 3 ; 4 ; 5
Taille (Cardinal)	: 50

2. Tableau statistique et calcul des cumuls

Tableau VII : Tableau de dénombrement des descendance de 50 vaches.

x_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
1	4	0,08	4	0,08
2	21	0,42	25	0,50
3	16	0,32	41	0,82
4	7	0,14	48	0,96
5	2	0,04	50	1
T	50	1	-	-

3. Représentation graphique

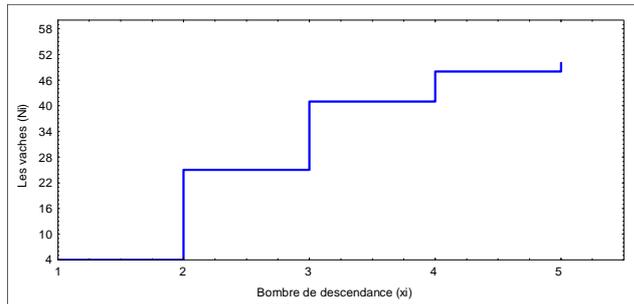


Figure 7 : Diagramme intégral des vaches en fonction du nombre de descendants.

4. Calcul des paramètres de position

a. La moyenne

$$m = [(0,08 \times 1) + (0,42 \times 2) + (0,32 \times 3) + (0,14 \times 4) + (0,04 \times 5)]$$

$$m = 2,64 \text{ vaches}$$

b. Le mode

$$Mo = 2 \text{ vaches}$$

c. La médiane

$$P = 25$$

$$Me = [x_{(25)} + x_{(26)}] / 2$$

$$Me = (2 + 3) / 2$$

$$Me = 2,5 \text{ vaches}$$

5. Calcul des paramètres de dispersion

a. La variance

$$s^2 = 1/49 [4(1 - 2,64)^2 + 21(2 - 2,64)^2 + 16(3 - 2,64)^2 + 7(4 - 2,64)^2 + 2(5 - 2,64)^2]$$

$$s^2 = 0,92 \text{ vaches}$$

b. Ecart-type

$$s = 0,95 \text{ vaches}$$

6. Type de la distribution

$$m > Me > Mo$$

La distribution est positivement dissymétrique

EXERCICE N°8

La TSHus (Thyréostimuline Hormone) est une hormone sécrétée au niveau de la glande thyroïdienne par l'hypothalamus. Stimulant ainsi, à ce niveau, la production de deux autres hormones de base ; T3 (Triiodthyronine) et T4 (Thyroxine). Le bon fonctionnement de la glande, qui repose sur l'ensemble des hormones, assure une régulation efficace du métabolisme en générale. Par contre, le dysfonctionnement de la glande engendre plusieurs maladies : Hyperthyroïdie, Hypothyroïdie, Le nodule et Le goitre... Pour cela, Un biochimiste s'est intéressé aux résultats des bilans, effectués à une clinique privée, des patients, les résultats sont les suivants :

Patients	Âge	[TSHus] (μ UI/ml)	Normes (μ UI/ml)	*Diagnostic I ^{aire}
Patient 1	30	0,15	0.35 – 4.94	A
Patient 2	29	0,10		A
Patient 3	21	0,23		A
Patient 4	33	0,50		B
Patient 5	25	1,20		B
Patient 6	58	60,50		A
Patient 7	43	1,40		B
Patient 8	19	72,50		A
Patient 9	21	61,55		A
Patient 10	19	3,48		B
Patient 11	45	4,35		B
Patient 12	37	5,28		A
Patient 13	25	1,55		B
Patient 14	21	0,23		A
Patient 15	19	61,54		A

***A** = Dysfonctionnement de la glande, **B**= Bon fonctionnement de la glande.

1. Faites ressortir les notions de base.
2. Donner le tableau statistique avec le calcul des cumuls croissants des effectifs et des fréquences.
3. Calculer les paramètres de position.
4. Calculer la variance, l'étendu interquartile, l'intervalle semi-interquartile et le coefficient interquartile relatif.
5. Précisez la forme de la distribution de la série par l'utilisation des trois démarches : paramètres de position, coefficients d'asymétrie et d'aplatissement.

SOLUTION

1. Les notions de base

Dans cette problématique il existe trois variables différentes :

Population statistique : Les patients
Unité statistique : Un patient
Echantillon : 15 patients
Taille (Cardinal) : 15

VARIABLE 1

Caractère : Age
Nature du caractère : Caractère quantitatif continu, mesurable, d'intensité
Modalités : 5 classes (voir tableau ci-après)

VARIABLE 2

Caractère : La quantité de la TSHus (Thyréostimuline Hormone)
Nature du caractère : Caractère quantitatif continu, mesurable, d'intensité
Modalités : 5 classes –voir tableau ci-après)

VARIABLE 3

Caractère : Etat de la glande
Nature du caractère : Caractère qualitatif ordinal
Modalités : A, B

2. Tableau statistique et calcul des cumuls

Nombre de classe : $k = 1 + 3,3 \text{ Log } 15 = 4,88$

$$\boxed{k = 5 \text{ classes}}$$

VARIABLE 1

Amplitude : $a = (58 - 19) / 5 = 7,8$

$$\boxed{a = 8 \text{ ans}}$$

Tableau VIII : Tableau des classes d'âges des patients étudiés.

Classes	c_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
[19 - 27 [23	8	0,53	8	0,53
[27 - 35 [31	3	0,20	11	0,73
[35 - 43 [39	1	0,07	12	0,80
[43 - 51 [47	2	0,13	14	0,93
[51 - 59 [55	1	0,07	15	1
T	-	15	1	-	-

VARIABLE 2

Amplitude : $a = (72,5 - 0,1) / 5 = 14,48$

$$a = 14,48 \mu\text{UI/ml}$$

Tableau IX : Tableau des classes de la quantité en TSHus des patients étudiés.

Classe	c_i	n_i	f_i	$N_i\uparrow$	$F_i\uparrow$
[0,1 – 14,58 [7,34	11	0,73	11	0,73
[14,58 – 29,06 [21,82	0	0	11	0,73
[29,06 – 43,58[36,3	0	0	11	0,73
[43,58 – 58,02[50,78	0	0	11	0,73
[58,02 – 72,5]	65,26	4	0,27	15	1
T	-	15	1	-	-

VARIABLE 3

Tableau X : Tableau des diagnostics de la glande thyroïdienne.

x_i	n_i	f_i	$N_i\uparrow$	$F_i\uparrow$
A	9	0,6	9	0,6
B	6	0,4	15	1
T	15	1	-	-

3. Calcul des paramètres de position

VARIABLE 1

a. La moyenne

$$m = 1/15 [(8 \times 23) + (3 \times 31) + (1 \times 39) + (2 \times 47) + (1 \times 55)]$$

$$m = 31 \text{ ans}$$

b. Le mode

Approche 1

$$Mo = 23 \text{ ans}$$

La classe modale est [19 – 27[ans

Approche 2

$$Mo = 19 + 8 [8 / (8 + 5)] = 19 + (8 \times 0,61)$$

$$Mo = 23,88 \text{ ans}$$

c. La médiane

Approche 1

$$Me = 19 + 8 [(7,5 - 0) / (8 - 0)]$$

$$Me = 26,50 \text{ ans}$$

Approche 2

$$Me = 19 + b$$

$$b = (7,5 \times 8) / 8 = 7,5$$

$$Me = 26,50 \text{ ans}$$

VARIABLE 2

a. La moyenne

$$m = 1/15 [(11 \times 7,34) + (0 \times 21,82) + (0 \times 36,3) + (0 \times 50,78) + (4 \times 65,26)]$$

$$m = 22,78 \text{ } \mu\text{UI/ml}$$

b. Le mode

Approche 1

$$Mo = 7,34 \text{ } \mu\text{UI/ml}$$

La classe modale est [0,1 – 14,58[$\mu\text{UI/ml}$]

Approche 2

$$Mo = 0,1 + 14,48 [11 / (11 + 11)]$$

$$Mo = 7,34 \text{ } \mu\text{UI/ml}$$

c. La médiane

Approche 1

$$Me = 0,1 + 14,48 [(7,5 - 0) / (11 - 0)]$$

$$Me = 9,97 \text{ } \mu\text{UI/ml}$$

Approche 2

$$Me = 0,1 + b$$

$$b = (7,5 \times 14,48) / 11 = 9,87$$

$$Me = 9,97 \text{ } \mu\text{UI/ml}$$

4. Calcul des paramètres de dispersion

VARIABLE 1

a. La variance

$$s^2 = 1/14 [8 (23 - 31)^2 + 3 (31 - 31)^2 + 1 (39 - 31)^2 + 2 (47 - 31)^2 + 1 (55 - 31)^2]$$

$$s^2 = 118,85 \text{ ans}$$

b. Ecart-type

$$s = 10,90 \text{ ans}$$

c. Etendu interquartile

Q_1 ?

Approche 1

$$Q_1 = 19 + 8 [(3,75 - 0) / (8 - 0)]$$

$$Q_1 = 22,75 \text{ ans}$$

Approche 2

$$Q_1 = 19 + b$$

$$b = (3,75 \times 8) / 8 = 3,75$$

$$Q_1 = 22,75 \text{ ans}$$

Q_3 ?

Approche 1

$$Q_3 = 35 + 8 [(11,25 - 11) / (12 - 11)]$$

$$Q_3 = 37 \text{ ans}$$

Approche 2

$$Q_3 = 35 + b$$

$$b = (0,25 \times 8) / 1 = 2$$

$$Q_3 = 37 \text{ ans}$$

$$Iq = 37 - 22,75$$

$$Iq = 14,25 \text{ ans}$$

d. L'intervalle semi-interquartile

$$Q = 14,25 / 2$$

$$Q = 7,12 \text{ ans}$$

e. *Le coefficient interquartile relatif*

$$CIQ = 14,25 / 26,50$$

$$CIQ = 0,53$$

VARIABLE 2

a. *La variance*

$$s^2 = 1/14 [11 (7,34 - 22,78)^2 + 0 (21,83 - 22,78)^2 + 0 (36,3 - 22,78)^2 + 0 (50,78 - 22,78)^2 + 4 (65,26 - 22,78)^2]$$

$$s^2 = 702,88 \mu\text{UI/ml}$$

b. *Ecart-type*

$$s = 26,51 \mu\text{UI/ml}$$

c. *Etendu interquartile*

Q_1 ?

Approche 1

$$Q_1 = 0,1 + 14,48 [(3,75 - 0) / (11 - 0)]$$

$$Q_1 = 5,03 \mu\text{UI/ml}$$

Approche 2

$$Me = 0,1 + b$$

$$b = (3,75 \times 14,48) / 11 = 4,93$$

$$Q_1 = 5,03 \mu\text{UI/ml}$$

Q_3 ?

Approche 1

$$Q_3 = 43,58 + 14,48 [(11,25 - 11) / (15 - 11)]$$

$$Q_3 = 44,48 \mu\text{UI/ml}$$

Approche 2

$$Q_3 = 43,58 + b$$

$$b = (0,25 \times 14,48) / 4 = 0,90$$

$$Q_3 = 44,48 \mu\text{UI/ml}$$

$$Iq = 44,48 - 5,03$$

$$Iq = 49,51 \mu\text{UI/ml}$$

d. L'intervalle semi-interquartile

$$Q = 49,51 / 2$$

$$Q = 24,75 \mu\text{UI/ml}$$

e. Le coefficient interquartile relatif

$$CIQ = 49,51 / 9,97$$

$$CIQ = 4,96$$

5. Type de la distribution

VARIABLE 1

a. Par la comparaison des paramètres de position

$$m > Me > Mo$$

La distribution est positivement dissymétrique

b. Par le calcul du coefficient d'asymétrie

$$S = m_3 / s^3$$

$$m_3 = 1/15 [8 (23 - 31)^3 + 3 (31 - 31)^3 + 1 (39 - 31)^3 + 2 (47 - 31)^3 + 1 (55 - 31)^3] =$$

1228,8 ans

$$s^3 = (10,90)^3 = 1295,02 \text{ ans}$$

$$S = 0,94$$

S est positif, alors la distribution étalée à droite.

c. Par le calcul du coefficient d'aplatissement

$$K = m_4 / s^4$$

$$m_4 = 1/15 [8 (23 - 31)^4 + 3 (31 - 31)^4 + 1 (39 - 31)^4 + 2 (47 - 31)^4 + 1 (55 - 31)^4] =$$

33314,13 ans

$$s^4 = (10,90)^4 = 14115,81 \text{ ans}$$

$$K = 2,36$$

$$K' = -0,64$$

$K' < 0$, donc les queues comptent moins d'observations que dans une distribution normale.

VARIABLE 2

d. Par comparaison des paramètres de tendance centrale

$$m > Me > Mo$$

La distribution est positivement dissymétrique

e. Par le calcul du coefficient d'asymétrie

$$S = m_3 / s^3$$

$$m_3 = 1 / 15 [11 (7,34 - 22,78)^3 + 0 (21,83 - 22,78)^3 + 0 (36,3 - 22,78)^3 + 0 (50,78 - 22,78)^3 + 4 (65,26 - 22,78)^3] = 17742,69 \mu\text{UI/ml}$$

$$s^3 = (26,51)^3 = 18630,70 \mu\text{UI/ml}$$

$$S = 0,95$$

S est positif, alors la distribution étalée à droite.

f. Par le calcul du coefficient d'aplatissement

$$K = m_4 / s^4$$

$$m_4 = 1 / 15 [11 (7,34 - 22,78)^4 + 0 (21,83 - 22,78)^4 + 0 (36,3 - 22,78)^4 + 0 (50,78 - 22,78)^4 + 4 (65,26 - 22,78)^4] = 910050,34 \mu\text{UI/ml}$$

$$s^4 = (26,51)^4 = 493899,86 \mu\text{UI/ml}$$

$$K = 1,84$$

$$K' = -1,16$$

$K' < 0$, donc les queues comptent moins d'observations que dans une distribution normale.

Remarque

VARIABLE 3 : Puisque les modalités sont exprimées en qualité alors la description numérique n'est pas possible.

EXERCICE N°9

Dans une étude sur les malformations morphologiques des poissons d'eau douce suite à leur exposition à la pollution, il a été analysé 33 individus. Les résultats obtenus sont ainsi :

Anomalies	2	4	5	8	10	11	12	14	15	18	20
Effectifs	1	2	1	4	2	7	6	3	4	2	1

- Déterminer l'étendue et le mode de cette série.
- Construire un tableau donnant les effectifs cumulés, les fréquences et les fréquences cumulées.
- Calculer la moyenne de cette série.
- Déterminer la médiane de cette série.
- Quel est le nombre de poissons ayant un nombre d'anomalie strictement inférieur à 8 ?
- Quel est le pourcentage de poissons ayant un nombre d'anomalie supérieur ou égal à 10 ?

SOLUTION

1. Etendue et mode

a. Etendue

$$E = 20 - 2 = 18$$

$$E = 18 \text{ anomalies}$$

b. Le mode

$$Mo = 11 \text{ anomalies}$$

2. Tableau statistique

Tableau XI : Tableau des effectifs et des fréquences des anomalies relatives aux 33 poissons analysés.

x_i	2	4	5	8	10	11	12	14	15	18	20	T
n_i	1	2	1	4	2	7	6	3	4	2	1	33
f_i	0,03	0,06	0,03	0,12	0,06	0,22	0,18	0,09	0,12	0,06	0,03	1
$N_i \uparrow$	1	3	4	8	10	17	23	26	30	32	33	-
$F_i \uparrow$	0,03	0,09	0,12	0,24	0,30	0,52	0,70	0,79	0,91	0,97	1	-

3. La moyenne

$$m = 1/33 [(1 \times 2) + (2 \times 4) + (1 \times 5) + (4 \times 8) + (2 \times 10) + (7 \times 11) + (6 \times 12) + (3 \times 14) + (4 \times 15) + (2 \times 18) + (1 \times 20)]$$

$$m = 11,33 \text{ anomalies}$$

4. La médiane

$$P = 16$$

$$Me = x_{(17)}$$

$$\boxed{Me = 11 \text{ anomalies}}$$

5. D'après le tableau des effectifs cumulés croissants, il y a 4 poissons qui ont un nombre d'anomalie strictement inférieur à 8.
6. Toujours d'après le tableau, 8 poissons sur 33 ont un nombre d'anomalie strictement inférieur à 8, donc :

$$\boxed{33 - 8 = 25 \text{ poissons ont un nombre d'anomalie supérieur ou égal à 10.}}$$

Soit un pourcentage égal à :

$$\boxed{\%tage = (25 / 33) * 100 = 75,75 \%}$$

EXERCICE N°10

On prélève 20 poulets dans un élevage et on mesure le taux de dioxine ($\mu\text{g/l}$) contenu dans leur viande afin d'estimer le taux moyen pour tout l'élevage. Les résultats sont donnés dans le tableau ci-dessous :

Poulets	1	2	3	4	5	6	7	8	9	10
Taux	0,34	0,23	0,11	0,42	0,22	0,33	0,16	0,12	0,11	0,21
Poulets	11	12	13	14	15	16	17	18	19	20
Taux	0,14	0,44	0,27	0,36	0,43	0,38	0,34	0,22	0,32	0,11

1. Donner le tableau statistique.
2. Calculer les effectifs cumulatifs croissants et les fréquences cumulatives croissantes.
3. Calculer les paramètres de position (moyenne, mode, médiane)
4. Calculer les paramètres de dispersion (variance et écart-type)

SOLUTION

1. Tableau statistique

Nombre de classes

$$k = 1 + 3,3 \text{ Log } 20$$

$$\boxed{k = 5 \text{ classes}}$$

Amplitude

$$a = (0,44 - 0,11) / 5$$

$$\boxed{a = 0,07 \mu\text{g/l}}$$

Tableau XII : Tableau des classes de quantité en dioxine retrouvée chez 20 poulets.

Classe	c_i	n_i	f_i	$N_i \uparrow$	$F_i \uparrow$
[0,11 – 0,18 [0,145	6	0,3	6	0,3
[0,18 – 0,25 [0,215	4	0,2	10	0,5
[0,25 – 0,32[0,285	1	0,05	11	0,55
[0,32 – 0,39[0,355	6	0,3	17	0,85
[0,39 – 0,46]	0,425	3	0,15	20	1
T	-	20	1	-	-

2. Les effectifs croissants et les fréquences croissantes sont donnés dans le tableau ci-dessus.

3. Les paramètres de position

a. La moyenne

$$m = 1/20 [(6 \times 0,145) + (4 \times 0,215) + (1 \times 0,285) + (6 \times 0,335) + (3 \times 0,425)]$$

$$m = 0,271 \mu\text{g/l}$$

b. Le mode

C'est une série statistique bimodale

Approche 1

$$Mo_1 = 0,145 \mu\text{g/l}$$

$$Mo_2 = 0,355 \mu\text{g/l}$$

La classe modale 1 est [0,11 – 0,18[$\mu\text{g/l}$

La classe modale 2 est [0,32 – 0,39[$\mu\text{g/l}$

Approche 2

$$Mo_1 = 0,11 + 0,07 [6 / (6 + 2)] = 0,11 + (0,07 \times 0,75)$$

$$Mo_1 = 0,162 \mu\text{g/l}$$

$$Mo_2 = 0,32 + 0,07 [5 / (5 + 3)] = 0,32 + (0,07 \times 0,625)$$

$$Mo_2 = 0,363 \mu\text{g/l}$$

Remarque

Cette méthode (approche 2) est appelée méthode de calcul approché cela explique alors la différence entre les résultats obtenus par les deux approches.

c. La médiane

Approche 1

$$Me = 0,18 + 0,07 [(10 - 6) / (10 - 6)]$$

$$Me = 0,25 \mu\text{g/l}$$

Approche 2

$$Me = 0,18 + b$$

$$b = (10 \times 0,07) / 10 = 0,07$$

$$Me = 0,25 \mu\text{g/l}$$

4. Les paramètres de dispersion

a. La variance

$$s^2 = 1/19 [6 (0,145 - 0,271)^2 + 4 (0,215 - 0,271)^2 + 1 (0,285 - 0,271)^2 + 6 (0,335 - 0,271)^2 + 3 (0,425 - 0,271)^2]$$

$$s^2 = 0,011 \mu\text{g/l}$$

b. Ecart-type

$$s = 0,10 \mu\text{g/l}$$

EXERCICE N°11

Des agriculteurs ont inventorié le nombre de terrains cultivés dans un village à Chemini, suivant la surface en m². Les résultats sont ainsi :

Surface	[400, 800[[800, 1000[[1000, 2500]
Effectif	2613	928	3379

- Déterminer la surface moyenne m d'après ce regroupement en classes.
- Sachant que la surface totale de répartition est de 6739000 m², calculer la surface moyenne d'un terrain cultivé. Comparer avec la valeur obtenue à la question 1.

SOLUTION

1. Calcul de la surface moyenne

Tableau XIII : Tableau du nombre de terrains cultivés en fonction des centres de classe des surfaces en m².

c_i	600	900	1750
n_i	2613	928	3379

$$m = 1/6920 [(2613 \times 600) + (928 \times 900) + (3379 \times 1750)]$$

$$m = 1201,77 \text{ m}^2$$

- Il y a au total :

$$2613+928+3379=6920 \text{ terrains cultivés}$$

Si la surface totale de répartition est de 6739000 m², la surface moyenne d'une terrain cultivé est égale à

$$6739000/6920 = 973.84 \text{ m}^2$$

Cette dernière valeur est bien inférieure à celle obtenue à la question 1.

EXERCICE N°12

Une compétition entre étudiants, à l'université de Béjaïa, a été organisée avec des gains en points. Le tableau ci-dessous résume les points perçus par les étudiants :

Points	100	200	300	400	500	600	700	800	900	1000
Etudiants	2	1	1	3	2	2	3	5	0	1

1. Calculer le nombre total d'étudiants.
2. Calculer le nombre moyen en points.
3. Calculer la médiane et les quartiles de cette série statistique, et déterminer l'étendu et le coefficient interquartile.
4. Calculer l'écart type de la série
5. L'université de Béjaïa envisage une augmentation des points.
 - a. La première hypothèse envisagée consiste à augmenter tous les gains de 217 points. Dans ce cas, comment varient la moyenne ; l'écart-type et la médiane ?
 - b. La deuxième hypothèse envisagée consiste à multiplier tous les gains par 1,2. Dans ce cas, comment varient la moyenne ; l'écart type et la médiane ?

SOLUTION

1. Le nombre d'étudiants à la compétition est égal à :

$$n = 2+1+1+3+2+2+3+5+0+1=20$$

$$\boxed{n = 20 \text{ étudiants}}$$

2. **La moyenne**

$$m = 1/20 [(2 \times 100) + (1 \times 200) + (1 \times 300) + (3 \times 400) + (2 \times 500) + (2 \times 600) + (3 \times 700) + (5 \times 800) + (0 \times 900) + (1 \times 1000)]$$

$$\boxed{m = 560 \text{ points}}$$

3. **La médiane, les quartiles, l'étendue interquartile, le coefficient interquartile**

- a. **La médiane**

$$P = 10$$

$$Me = [x_{(10)} + x_{(11)}] / 2$$

$$Me = (600 + 600) / 2$$

$$\boxed{Me = 600 \text{ points}}$$

b. Les quartiles

$$Q_1 ?$$

$$P = 5$$

$$Q_1 = [x_{(5)} + x_{(6)}] / 2$$

$$Q_1 = (400 + 400) / 2$$

$$\boxed{Q_1 = 400 \text{ points}}$$

$$Q_3 ?$$

$$P = 15$$

$$Q_3 = [x_{(15)} + x_{(16)}] / 2$$

$$Q_3 = (800 + 800) / 2$$

$$\boxed{Q_3 = 800 \text{ points}}$$

c. L'étendue interquartile

$$Iq = Q_3 - Q_1 = 800 - 400$$

$$\boxed{Iq = 400 \text{ points}}$$

d. Le coefficient interquartile

$$Q = Iq / Me = 400 / 600 \text{ (} Me = Q_2 \text{)}$$

$$\boxed{Q = 0,66}$$

4. L'écart-type

a. Variance

$$s^2 = 1/19 [2 (100 - 560)^2 + 1 (200 - 560)^2 + 1 (300 - 560)^2 + 3 (400 - 560)^2 + 2 (500 - 560)^2 + 2 (600 - 560)^2 + 3 (700 - 560)^2 + 5 (800 - 560)^2 + 0 (900 - 560)^2 + 1 (1000 - 560)^2]$$

$$\boxed{s^2 = 65684,21 \text{ points}}$$

b. Ecart-type

$$\boxed{s = 256,28 \text{ points}}$$

5. L'université de Béjaïa envisage une augmentation des points.

a. Hypothèse 1

• **La moyenne**

Elle augmente également avec 217 points, et devient donc :

$$m = 560 + 217$$

$$\boxed{m = 777 \text{ points}}$$

- *L'écart-type*

Il est inchangé car la dispersion des valeurs autour de la moyenne n'est pas modifiée par cette augmentation.

- *La médiane*

Elle se trouve également augmentée de 217 points, et devient :

$$Me = 600 + 217$$

$$\boxed{Me = 817 \text{ points}}$$

b. Hypothèse 2

- *La moyenne*

Elle est également multipliée par 1,2 et devient :

$$m = 560 \times 1,2$$

$$\boxed{m = 672 \text{ points}}$$

- *L'écart-type*

Il est également multiplié par 1,2 et devient

$$s = 256,28 \times 1,2$$

$$\boxed{s = 307,53 \text{ points}}$$

- *La médiane*

Elle s'en trouve également multipliée par 1,2 et devient

$$Me = 600 \times 1,2$$

$$\boxed{Me = 720 \text{ points}}$$

EXERCICE N°13

Un contrôle de mortalité a été effectué pendant 100 jours d'observation sur deux sites d'élevage des crustacés destinés pour l'exportation. Certains individus présentent une infestation parasitaire qui les rend inexploitable. On a relevé le nombre de spécimens infestés constatés durant chaque jour :

Site 1 « El Kala »

Nombre de spécimens infestés	0	1	2	3	4	5	6	7
Nombres de jours	13	42	38	2	2	1	1	1

Site 2 « Annaba »

Nombres de spécimens infestés	0	1	2	3	4	5
Nombres de jours	35	40	1	1	10	13

1. Calculer le nombre moyen m_1 de spécimens infestés pendant les 100 jours observés dans le premier site. Calculer ensuite la variance s^2_1 .
2. Calculer le nombre moyen m_2 de spécimens infestés pendant les 100 jours observés dans le site 2. Calculer ensuite la variance s^2_2 .
3. Déterminer la médiane puis l'étendue interquartile dans le cas 1. Calculer l'étendue.
4. Déterminer la médiane puis l'étendue interquartile dans le cas 2. Calculer l'étendue.
5. Parmi la moyenne, l'écart type, la médiane, l'étendue interquartile ou l'étendue, quels sont les paramètres qui mesurent la dispersion ?
6. Quel(s) paramètre(s) semble(nt) le(s) plus intéressant(s) à exploiter pour comparer ces deux sites ? Justifier.

SOLUTION

1. La moyenne et la variance dans le cas 1

a. La moyenne

$$m_1 = 1/100 [(13 \times 0) + (42 \times 1) + (38 \times 2) + (2 \times 3) + (2 \times 4) + (1 \times 5) + (1 \times 6) + (1 \times 7)]$$

$$\boxed{m_1 = 1.5 \text{ spécimens}}$$

b. La variance

$$s^2_1 = 1/99 [13(0 - 1,5)^2 + 42(1 - 1,5)^2 + 38(2 - 1,5)^2 + 2(3 - 1,5)^2 + 2(4 - 1,5)^2 + 1(5 - 1,5)^2 + 1(6 - 1,5)^2 + 1(7 - 1,5)^2]$$

$$\boxed{s^2_1 = 1.30 \text{ spécimens}}$$

2. La moyenne et la variance dans le cas 2

a. La moyenne

$$m_2 = 1/100 [(35 \times 0) + (40 \times 1) + (1 \times 2) + (1 \times 3) + (10 \times 4) + (13 \times 5)]$$

$$\boxed{m_2 = 1.5 \text{ spécimens}}$$

c. La variance

$$s^2_2 = 1/99 [35(0 - 1,5)^2 + 40(1 - 1,5)^2 + 1(2 - 1,5)^2 + 1(3 - 1,5)^2 + 10(4 - 1,5)^2 + 13(5 - 1,5)^2]$$

$$s^2_2 = 3,16 \text{ spécimens}$$

3. La médiane, l'étendue interquartile et l'étendue dans le cas 1

a. La médiane

$$P_1 = 50$$

$$Me_1 = [X_{(50)} + X_{(51)}] / 2$$

$$Me_1 = (1 + 1) / 2$$

$$Me_1 = 1 \text{ spécimen}$$

b. L'étendue interquartile

$$Q_1$$

$$P_1 = 25$$

$$Q_1 = [X_{(25)} + X_{(26)}] / 2$$

$$Q_1 = (1 + 1) / 2$$

$$Q_1 = 1 \text{ spécimen}$$

$$Q_3$$

$$P_1 = 75$$

$$Q_3 = [X_{(75)} + X_{(76)}] / 2$$

$$Q_3 = (2 + 2) / 2$$

$$Q_3 = 2 \text{ spécimens}$$

$$Iq_1 = Q_3 - Q_1 = 2 - 1$$

$$Iq_1 = 1 \text{ spécimen}$$

c. Etendue

$$E_1 = 7 - 0$$

$$E_1 = 7 \text{ spécimens}$$

4. La médiane, l'étendue interquartile et l'étendue dans le cas 1

a. La médiane

$$P_2 = 50$$

$$Me_2 = [X_{(50)} + X_{(51)}] / 2$$

$$Me_2 = (1 + 1) / 2$$

$$Me_2 = 1 \text{ spécimen}$$

b. L'étendue interquartile

$$Q_1$$

$$P_2 = 25$$

$$Q_1 = [X_{(25)} + X_{(26)}] / 2$$

$$Q_1 = (0 + 0) / 2$$

$$Q_1 = 0 \text{ spécimen}$$

$$Q_3$$

$$P_2 = 75$$

$$Q_3 = [X_{(75)} + X_{(76)}] / 2$$

$$Q_3 = (1 + 2) / 2$$

$$Q_3 = 1,5 \text{ spécimens}$$

$$Iq_2 = Q_3 - Q_1 = 1,5 - 1$$

$$\boxed{Iq_2 = 0,5 \text{ spécimen}}$$

c. Etendue

$$E_2 = 5 - 0$$

$$\boxed{E_2 = 5 \text{ spécimens}}$$

5. Les paramètres qui mesurent la dispersion sont l'écart type et l'étendue interquartile. Plus ils sont petits, et plus la série est regroupée. Plus ils sont grands, et plus la série est dispersée.

6. Dans le cas des deux sites ci-dessus, puisque leurs moyennes sont identiques, les variances (et donc les écarts-types qui en sont leurs racines carrées) nous indique que les valeurs du site 2 sont plus dispersées que celles du site 1. Il en est de même de l'étendue interquartile. Le site 1 semble donc plus homogène que le site 2.