Université de Lubumbashi

FACULTE DE DROIT

Département de Droit économique et social

STATISTIQUE DESCRIPTIVE :

Notes de cours destinées aux étudiant de troisième année graduat Droit

Par Jean Pierre BAKATUAMBA

Professeur

Chef de Département de Droit économique et social

Secrétaire du Laboratoire d'analyse économique des lois

Lubumbashi, 2021

INTRODUCTION

1. OBJET

Le domaine d'application de la statistique est trop vaste. Beaucoup de sciences recourent à la statistique, c'est notamment la médecine, l'agronomie, la biologie, l'économie, la chimie, la psychologie, la santé, le sport, la physique, le droit,…. Elle s'appréhende comme des ***faits numériques*** concernant plusieurs choses telles que le nombre d'étudiants dans plusieurs facultés, les points obtenus par les étudiants dans différents cours, l'augmentation ou la diminution des prix des denrées alimentaires de première nécessité, les buts marqués par une équipe dans un championnat de foot, le nombre d'accidents de circulation dans une ville. Bref, la statistique nous aide au travers la numérisation, la détermination de tous ces faits sociaux. La statistique nous permet de tirer profit de ses observations et de dégager les lois à partir des phénomènes observés. Ces phénomènes sont traduits en chiffres, en %, en tableaux et présentés sous forme des graphiques.

2. IMPORTANCE

La statistique est utilisée pour résoudre des problèmes en vue d'obtenir des faits qui amènent à une nouvelle connaissance et une pénétration dans un domaine spécifique d'étude. Elle rend les gens capables de sonder les décisions basées sur des faits, faire des prévisions et conclure à partir des faits présentés.

- Un commerçant avisé est toujours en quête d'informations sur les articles qu'il achète ou qu'il vend. Il doit savoir à tout moment ce qui reste en stock et la quantité qu'il faut ajouter et où le trouver.
- La police nationale routière prélève chaque jour les informations sur les différents accidents, les causes et éventuellement la situation des accidentées pour en déduire une projection par mois, par trimestre, par semestre ou par année et déterminer en même temps la bonne ou la mauvaise période ou au besoin connaître l'in fraction la plus récurrente.

Exemples :

La statistique s'oriente vers une description des phénomènes économiques et sociaux et aborde l'organisation, la présentation et la synthèse des séries, des indices, des séries chronologiques ainsi que les problèmes de corrélation et de régression statistiques.

3. OBJECTIF SPECIFIQUE

A l'issue de cet enseignement chaque apprenant sera en mesure de :

- Définir et expliquer certains concepts statistiques.
- Collectionner les données selon les méthodes fiables.
- Classer les données dans les formats compréhensibles et concis (ou tabulation des données et autres formes de résumés des données) selon leur nature.
- Représenter les données au moyen des tableaux et des graphiques.
- Analyser les données au moyen des techniques statistiques appropriées par le calcul d'une corrélation.
- Interpréter les données par l'élaboration d'une droite ou d'une courbe de régression.

4. CONTENU

Point I. Introduction

- Définitions et terminologie

  Population, Individu, variable, caractère, modalités, échantillon et série statistique.
- Fondamentaux
  - Droite, segment de droite, demi-droite, intervalle ouvert, fermé, semi-ouvert ou semi fermé, Axes (X, Y).
  - Ensemble, N, Z, Q, R, $\leq$, $\geq$, $=$, pair, impair.
  - Moyenne arithmétique, valeur absolue
  - Fraction, notation décimale, en pourcentage, proportion et règle de trois simples.

Point II. Collecte et classification des données

- Collecte des données.
- Classification des données (sériation).

Point III. Représentations graphique

- Diagramme à bâtons (barre)
- Histogramme (diagramme en bandes)
- Polygone des effectifs cumulés croissants et décroissants

- Diagramme circulaire

Regroupement en classes.

Point IV. Analyse des données

- Paramètres de position (mode, moyenne, médiane, quartiles)
- Paramètres de dispersion
- Indicateurs statistiques

5. BIBLIOGRAPHIE

- Barrussaud, G. et Touzer, A., Mathématique, Baccalauréats professionnels tertiaires, Foucher, 1996.
- Bontemps, G., Cie, Mathématique seconde, Collection, Fractale, édition Natalie Carnellana, Bordas, Paris, 1990.
- Derro, A., et le charpentier, F., Activités maths secteur tertiaire, Hachette technique, 2000.
- Bringuier, G., Marie-claude Dairé et Eric Sirot, Mathématiques et traitement de données, Hachette technique, 2003.
- Fredon, D., Gay, M., Jensen, C., Mathématiques classe terminale D, Algèbre-Probabilité, Statistique, Armand Colin, Paris, 1983.
- Bwalankay. Pa-Kome, Statistique, Kinshasa, 1998.
- Christian Labrousse, Statistiques : Exercices corrigés, Tome I, Dunod, Paris, 1972.
- Bwalankay. Pa-Kome, Statistique, Ed. P@Kome, Collection New Math, Kinshasa, 2012.
- Dangnelie, P., Statistique théorique et appliqué, Tomme II, Paris, 1998.
- Keth, Lockyer, Guide de gestion de stock, Paris, 1974.
- Huguier M ; Flahaut, A., Biostatistiques au quotidien, Paris, Mai 2002.

6. METHODOLOGIE.

L'approche sera ex. cathedra avec interaction pour faciliter la compréhension de la matière.

L'apprenant devra s'interdire de considère l'analyse de bilan comme analyse financière qui du reste est de la faculté d'économie.

7. EVALUATION.

L'évaluation sera journalière au début de chaque séance d'enseignement. La présence au cours étant obligatoire nous tiendrons compte de cela dans l'évaluation. A l'issue une épreuve sera organisée.

8. RESSOURCES.

Outre notre présence, un support sera mis à la disposition de chaque apprenant.

9. DISCIPLINE.

Le formateur accède le dernier à l'auditoire et quitte le premier. Les phones ne seront si pas fermés, sous mode vibreur.

10. DU PROFESSEUR.

Le professeur se nomme **Jean Pierre BAKATUAMBA**, Docteur en Droit économique de l'UNILU.

Contacts : +243997029994, +243851620059.

E- mail : Jpbakatuamba@gmail.com.

I. Préambule

1. **Définitions**

1.1. Une statistique est un ensemble d'observations (exprimer en nombre) concernant un fait ou un sujet donné. Elle est une branche des mathématiques appliquées qui fournit la méthode d'étude statistique. Elle se divise en deux parties :

- La statistique descriptive qui se limite à la description des données. Elle est une statistique de constatation et d'élaboration des tableaux et graphiques.
- La statistique analytique qui procède à l'analyse, à l'interprétation de données et à la généralité des faits considérées à partir d'un échantillon. Elle cherche les explications, teste la signification des résultats et s'efforce d'en découvrir les causes.

1.2. **On appelle population**, l'ensemble étudié. Les unités ou les individus composant la population doivent être de même nature. Il s'agit de l'homogénéité, c'est-à-dire population homogène.

Exemples : - Etudiants de G_3 Droit.
- Habitants d'une rue.
- Professeurs de la faculté.

Un individu (ou comité statistique) est un élément de la population qui à son tour présente une ou plusieurs variables (ou caractères) qui peuvent être quantitatives ou qualitatives.

Exemples :

- Etudiants de sexe féminin de G_3 Droit
- Etudiants intelligents de G_3 Droit.

1.3. L'échantillon est un sous-ensemble étudié. Ce sous-ensemble peut être pris de manière hasardeuse, il s'agit d'un échantillon aléatoire ou de manière intéressée, il sera question d'un échantillon raisonné.

Exemples :

- Filles de G_3 Droit : Echantillon aléatoire.
- Filles de G_3 Droit de grande taille : Echantillon raisonné.

Lorsque l'échantillon vise la quantité, il s'agit d'un échantillon quantitatif c'est-à-dire mesurable et si c'est la qualité qui est prise en compte, on parle d'un échantillon qualitatif.

Exemples :

- Si l'ensemble à étudier est une université, alors la faculté en est un échantillon. Idem, la promotion est un échantillon de la faculté.
- Pour connaître le taux de fréquentation des filles dans la ville de Lubumbashi, on peut considérer une ou deux communes pour mener cette étude. Les communes ciblées constituent un échantillon de Lubumbashi.

1.4. La variable dite aussi caractère est l'élément sur lequel porte l'étude. Cette variable peut viser la quantité (mesure) ou la qualité (valeur, le seuil ou 1 et/ou 0). La variable est un trait commun à tous les individus étudiés.

Exemples :

- L'âge, le poids, la taille des étudiants de G3 Droit.
- Le nombre de fautes d'une dictée française posée en G3 Droit. (Ce nombre qui désigne la quantité vise la qualité : plus de 10 fautes : médiocre, moins de 5 fautes : Bon, moins de 2 fautes : TB, etc...).

- Lorsque la variable quantitative prend la valeur entière de la population ou de l'échantillon, elle est dite discrète ou discontinue.

Exemples :

- Nombre d'étudiants de G3 Droit.
- L'âge d'une personne.
- Nombre de frères et sœurs dans une famille.

- Lorsque la variable quantitative prend n'importe quelle valeur dans un intervalle fini ou infini, elle est dite continue.

  Exemple :

- Le poids et la taille d'une personne c'est-à-dire en visant le poids et la taille, les autres éléments de la personne sont mis de côté. (force, intelligence,...)

- Etudier une série statistique à une variable consiste à étudier l'ensemble des éléments, on aurait pu étudier aussi d'autres aspects tels que "taille", "sexe", "poids" etc.

- En visant un caractère exclusif ou exhaustif des éléments d'un échantillon, on établit ce qu'on appelle modalité, ce qui signifie qu'un individu doit présenter une seule modalité et une seule variable.

Exemple : Modalité du caractère "sexe" (féminin ou masculin). Il sera hors de question de trouver un individu ni femme ni homme.

- Si deux variables sont supposées influencer la mesure d'une autre variable, elles sont dites indépendantes entre elles.

Exemple : Si le sexe et l'âge influencent la mesure de l'intelligence, le sexe et l'âge sont indépendants.

- Lorsque l'évolution d'une variable règle ou dérègle l'autre variable, ces variables sont dites dépendantes.

Exemple : Le poids et la santé d'une personne.

Bonne santé → Bon poids

Mauvaise santé → Mauvais poids.

1.5. Série statistique

Les éléments observés peuvent avoir entre eux en lieu, c'est-à-dire si à chacun des éléments de l'ensemble considéré on fait correspondre la valeur de la variable (caractère) étudié, on obtient une série statistique en forme soit d'une liste énumérative ou d'une collection des fiches.

Exemple : Soit l'ensemble A= {1, 2, 5, 7} des années de Servitude Pénales Principales, avec comme caractère le double des peines (SPP) de A.

La série sera : 2, 4, 10, 14, 4, 8, 20, 28, 8, 16, 40, 56,….

- Lors de l'établissement d'une liste énumérative ou d'une collection des fiches, il s'affiche des effectifs ou des fréquences dites absolues. L'effectif est donc le nombre de répétition d'un caractère étudié. Cet effectif peut-être simple ou cumulé.

A= {2, 4, 10,14} Effectif simple A= {2, 4,10,14,4 ,8,20, 28, 8, 16,40, 56 ,…} Effectif cumulé.

1.6. Effectif

Le nombre d'éléments pour lesquels la variable a une valeur donnée s'appelle l'effectif de cette variable.

Exemples :

Soit un groupe de prisonniers, que chacun décline le motif de son arrestation.

Escroquerie notée A	représente 2
Viol noté	B représente 3
Abus de confiance	F représente 0
Coups et blessures	M représente 11
Vol	K représente 16

Donc, l'effectif de la variable A est 2.

l'effectif de la variable B est 3

l'effectif de la variable F est 0

l'effectif de la variable M est 11

l'effectif de la variable K est <u>16</u>

32

A la fin du remplissage du tableau, il faut vérifier si le total des effectifs correspond à la valeur de la population étudiée.

2) Si j'ai 200.000 Fc en billets différents, je peux connaître combien de billets de chaque coupure. A la fin, les effectifs de différentes coupures correspondent à des montants qui doivent correspondre à 200.000 Fc

2. Les fondamentaux

Les fondamentaux est un rappel des notions d'arithmétique et de géométrie qui reviennent de manière récurrente tout au long de l'étude statistique.

2.1. Ensembles

N : Ensemble des naturels {0, 1, 2, …., ∞}

Z : ensemble des entiers relatifs {-∞, -1, 0, 1, 2, …., ∞}

Q : Ensemble des décimaux (rationnels) ex : 0,5 ou ½.

R : Ensemble des réels. Ex. : $\sqrt{3}$

C : Ensemble des nombres complexes. Ex : $\sqrt{-3}$

$\mathbb{C} \supset \mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z} \supset \mathbb{N}$.

- Dans ces ensembles existent les éléments qui sont comparables les uns des autres.

Exemples

- 2 est inférieur à 3 ($2 < 3$) qu'on peut aussi noter $3 > 2$. (3 supérieur à 2).
- Rappeler aussi l'usage de au moins et au plus. Au moins $10 = 10, 11, 12,\ldots$

Lorsque l'on travaille dans un ensemble, il faut que le résultat soit de cet ensemble. D'où le recours à l'arrondissement. $1,3 = 1$: $1,5 = 2$; $1,7 = 2$; $1,0 = 1$. Si on trouve dans $\mathbb{N}$ ou $\mathbb{Z}$

2.2. Lors d'une énumération, la moyenne arithmétique notée $\bar{M}$ est la moyenne de la somme des valeurs par le nombre de valeurs. $\bar{X} = \frac{\sum n_i x_i}{\sum n_i}$=N

Remarque : Si les données sont regroupées en classes.

x_i : est le centre de classe.

n_i : effectif correspondant au centre de classe.

N : la somme des effectifs.

Exemples :

1. Quatre étudiants ont reçu les cotes suivantes : A : 7/10 ; B : 6/10 ; C : 3/10 ; D : 4/10. Quelle est la moyenne arithmétique de ces côtes ?
2. Réponse : - Il y a 4 étudiants

- Somme des côtes = 7+6+3+4 = 20.

  La moyenne notée $\bar{M} = \frac{20}{4} = 5$.

  $\bar{M} = 5$.

3. Au cours de la même année, j'ai enregistré 70 Kgs, 65 Kgs, 50 Kgs, 73 Kgs, 68 Kgs. Quelle est la moyenne de ma masse ?

   Réponse : Somme → 70 +65 + 50 +73 +68 = 326

   Combien de mesure : 5

   La moyenne notée $\bar{M} = \frac{S}{M} = \frac{326}{5} = 65{,}2$ kgs

<u>Attention</u>

Dans la terminologie quantitative, la notion moyenne désigne parfois la moitié.

Exemple : Moyenne de 10 égale 5.

2.3. Logarithme

a∈R_+^*\ {1} et b∈R_+^*, le réel x est le logarithme de b dans la base a si et seulement si x est l'exposant qu'il faut attribuer à a pour obtenir b.

On note $log_a^b = X \Leftrightarrow a^x = b$.

log_3^9 =?, x qu'il faut avoir comme exposant de 3 pour avoir 9.

$3^x = 9 \Longrightarrow 3^x = 3^2 \Longrightarrow x = 2$.

$log_3^9 = 2$

Lorsque la base est décimale (10), on note Log=Ln

<u>Propriétés</u>

1. $Log_a^{b.c} = log_a^b + log_a^c$

Exemple :

$log_2^8 = log_2^{4.2} = log_2^4 + log_2^2 =$

$= log_2^{2.2} + log_2^2 == log_2^2 + log_2^2 + log_2^2 = 1+1+1 = 3$

$log_2^8 = 3$ parceque $2^3 = 8$.

2. $Log_a^{b^n} = n\, log_a^b$

Exemple : $log_2^8 = log_2^{2^3} = 3log_2^2 = 3.1 = 3$

3. $log_a^{\frac{b}{c}} = \frac{log_a^b}{log_a^c}$

4. $log_a^{\sqrt[n]{b}} = \frac{1}{n} log_a^b$

5. si a =b$\Longrightarrow log_a^b = 1$.

Exercices. Calculez :

a) $log_4^{16} = log_4^{4^2} = 2\, log_4^4 = 2.1 = 2.$

b) $log\, 200 = log 10^2.2 = log 10^2 + \log 2$

$= 2 \log 10 + \log 2$

$= 2.1 + 0{,}247$

$=2+0.247 = 2{,}247$

c) $log\, 80 = log 2^3.10 = log 2^3 + \log 10$

$= 3 \log 2 + \log 10$

$= 3.0{,}247 + 1$

$= 1{,}741$

- La proportionnelle est la représentation unitaire ou la participation de chaque individu dans la série, la suite ou l'ensemble quelconque.

Exemples :

1. Trois bananes coûtent 2.000 Fc, quel est le prix d'une banane, 2 banane. Représenter cela en pourcentage et en fraction. Dire combien de bananes il y a dans 1.200 Fc.

<u>Solution</u>

1) 3 bananes $=$ 2.000 Fc

   1 banane $=$ 2.000 Fc/3

   2 bananes $= \frac{2.000}{3} \times 2$

2) 2.000 Fc $=$ 100% ou 3 bananes $=$ 100%

   Donc : 1 Fc $= \frac{100\%}{2.000}$ ou une banane $= \frac{100\%}{3}$

   $= \frac{100\%}{2.000}$ ou 2 banane $= \frac{100\%}{3}.2$

   $= \frac{1}{20}\%$

   $= \frac{1.10^{-1}}{2}\%$

   $= 0{,}05\%$

   1.000 Fc $=$ 0,05% x 1000

   $=$ 50%

   1.200 Fc $=$ 0,05% x 1.200

   $=$ 0,05% x 1.200

   $=$ 60%.

3) Combien de bananes dans 1.200 Fc, pur quelle proportion.

   1 banane $= \frac{2000\ Fc}{3}$; 1 Fc $= \frac{100\ Fc}{2.000}$, 1 Fc $= \frac{3}{2.000} \times b$

   1.200 Fc $= \frac{3}{2.000} \times 1.200\ bananes$; $\quad$ $1.200\ Fc\ = \frac{100}{2.000} \times 1.200\% = 60\%$.

4) Si 15 étudiants identiques pèsent globalement 825Kgs. Quelle est la masse de 7 étudiants, 12 étudiants en nombre et en pourcentage ?

<u>Solution 1</u>

15 étudiants $=$ 825 Kgs, 100% $=$ 825, 15 $=$ 100Y ; 825=100% ; 1Kg $= \frac{100\%}{825}$

1 étudiant $= \frac{825}{15} = 55Kgs$; 1 étudiant$= \frac{100\%}{15}$; 55Kg$= \frac{100\%}{825} \times 55kgs$

$1\% = \frac{825}{100} Kg$

<u>Solution 2</u>

Si $1\% = \frac{825}{1000}$, 1 étudiant $= \frac{100}{15}\%$, 1 étudiant $= 55\ Kgs$, donc 12 étudiants : $= 55.12 =$ 660 Kgs

12 étudiants : $= \frac{100}{15} . 12\% = 80\%$.

Si nous avons $\frac{X}{Y}$ comme rapport de participation d'une variable en pourcentage cela conduit à noter $\frac{X}{Y} \times 100\%$.

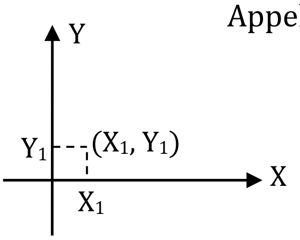
Droite : — : infinie dans les deux sens.

Semi-droite : ⊢ : infinie dans un sens

Demi-droite : ⊢ : infinie dans un sens

Segment : ⊢⊣ : limité dans deux sens

Axe : X : abscisse Y : ordonnée et le croisement abscisse-ordonnée donne Appelle coordonnées.



[2 4] : 2 et 4 inclus

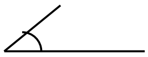
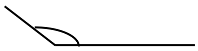
[2, 4[: 2 inclus, 4 exclu

]2, 4] : 2 exclu et 4 inclus

]2, 4[: 2 et 4 exclus.

1.4. Les angles, le cercle et le degré.

L'intersection des lignes crée les angles. Au nombre des angles il y a :

- Angle droit : 90° 
- Angle aigu < à 90° 
- Angle obtu >à 90° 
- Angle plat. 180°

Dans un cercle nous avons :

1 cercle = 360° = 400 grades = 2π radians.

Pour dessiner les effectifs dans le cercle, on procède de la manière suivante :

L'angle de la valeur $n_i = \frac{Nombre\ de\ degrés\ du\ cercle}{Effectif\ total\ (N).n_i} = \frac{360°.n_i}{N}$

Exemple : Dans un auditoire de 60 étudiants, le professeur constitue des groupes selon la couleur d'habits portés ce jour. Il obtient ce tableau statistique :

Couleur d'habits	Blanc	Rouge	Bleu	Jaune
Effectif n_i	5	10	20	25

Dessiner le diagramme circulaire de cette série statistique.

<u>Solution</u>

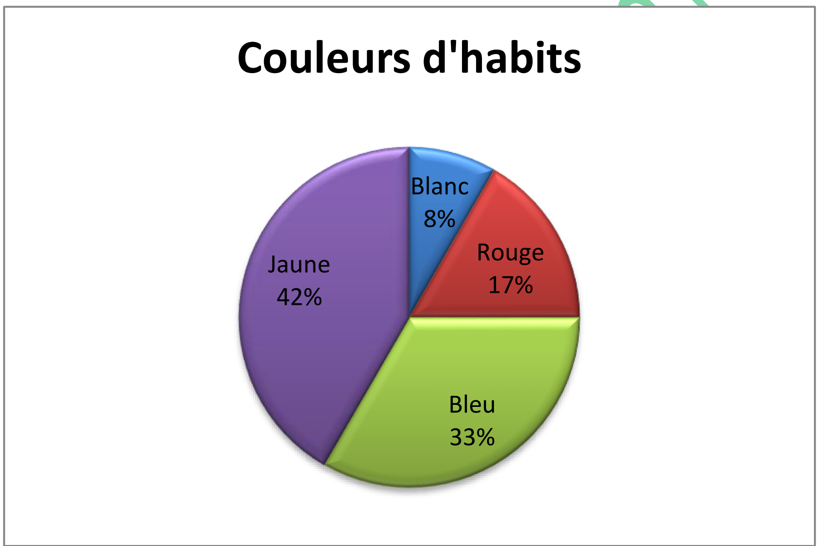
N= 60, n blanc = 5, n rouge = 10, n bleu = 20 et n jaune = 25.

L'angle de la valeur « blanc » $= \frac{5}{60} * 360° = 30°$

L'angle de la valeur « Rouge » $= \frac{10}{60} * 360° = 60°$

L'angle de la valeur « Bleu » $= \frac{20}{60} * 360° = 120°$

L'angle de la valeur « Bleu » $= \frac{25}{60} * 360° = 150°$.



Exercices :

1. Une entreprise fabrique des stylos à billes de couleur différente. On a effectué les observations suivantes sur le lot de 150 pièces produites.

Couleur des pièces	Vert	Jaune	Rouge	bleu	Total
Effectif	36	42	54	18	150

Construire le diagramme circulaire en pourcentage pour cette série statistique.

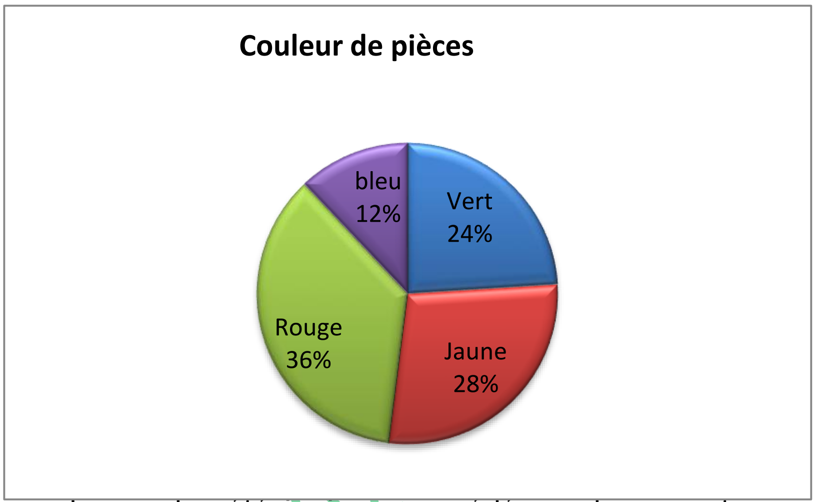
Solution : N = 150

Couleur verte : $\frac{36}{150} \times 100 = 24\%$

Couleur jeune : $\frac{42}{150} \times 100 = 28\%$

Couleur rouge : $\frac{54}{150} \times 100 = 36\%$

Couleur bleue : $\frac{18}{150} \times 100 = 12\%$



2. Dans un poulailler de 720 pondeuses, le vétérinaire a procédé au classement selon la couleur de plume et a obtenu le tableau statistique suivant :

Couleur	Leghorn blanche	Issex blanche	Issa brwon	Susseix hyght	Rhad Isaland Red	Low man brwon
Effectif	210	120	60	90	180	60

En degré, dessiner le diagramme circulaire de cette série statistique.

II. COLLECTE ET CLASSIFICATION DES DONNEES

II.1. Collecte des données

Il existe plusieurs méthodes pour collecter les données. Malgré que chacune d'elles peut présenter ses spécificités, elles ont toutes des attributs pour se rendre chacune plus utile que les autres dans certaines situations.

Au nombre de ces méthodes, nous avons entre autre :

- Interview ;
- Enquêtes téléphoniques ;
- Opinion écrite ou orale sur une matière spécifique ;
- Questionnaires ;
- Tests standardisés, …

De toutes ces méthodes, c'est le questionnaire qui est le plus usité. Une fois la méthode retenue, il reste un autre problème majeur, celui du choix du groupe spécifique sur lequel portera la recherche. Le groupe spécifique, dit répondant est supposé fournir les renseignements requis. Ainsi donc, le chercheur devra déterminer le groupe afin de collecter ou d'investiguer. Il s'agit donc de choisir l'échantillonnage.

Plus l'échantillon est représentatif, mieux le résultat obtenu est fiable. On estime cet échantillon à plus ou moins 20% de la population cible, mais il existe de situations où 20% sont énormes.

L'échantillonnage rassure que les résultats des membres sélectionnés à partir de la population, produira une image représentative et fiable du groupe à investiguer au point de l'étendre (généraliser) à la population.

Exemple : Soit une population de 12.500 candidats au test d'entrée à l'université de Lubumbashi. Pour déterminer la dynamique de ce groupe, il y a 2.500 copies choisies en fonction : - garçons et filles ; - même année ou année antérieure ; - garçon de Lubumbashi ou d'intérieur ; - fille de Lubumbashi ou d'intérieur ;… qui nous renseignent sur l'objectif recherché.

Lorsque les variables (caractères) sont mal choisies, il y a plus de chance de ne pas atteindre l'objectif, c'est-à-dire la fiabilité. Toujours éviter de se faire influencer par le caractère dominant, qui peut ne pas être la vérité.

Exemples :

- S'il faut enquêter sur le pourquoi les étudiants trichent, pas besoin de ne prendre que les étudiants du fond supposés être paresseux.

- Avant de déduire sur la propreté, il faut étaler l'étude sur les différentes heures de la journée.

II.2. Classification des données

II.2.1. Définition

La classification dite dépouillement est une étape qui vient après la collecte des données. Cette étape peut-être réalisé à la main ou à la machine dans une bonne organisation faite au moyen des tableaux de format clair, de pointage, de carreaux, de pâturage qui du reste sont concis et compréhensifs.

Exemple :1. Soit G_3 Droit une promotion des majeurs dont l'âge varie entre : 21 ans, 22 ans, 23 ans, 24 ans, 25 ans et plus. Après collecte, nous avons la situation suivante de 142 étudiants :

21 ans :

22 ans :

23 ans :

24 ans :

25 ans :

Plus :

Représenter ces effectifs par pointage, carreaux ou bâtonnets.

21 ans :

22 ans :

23 ans :

24 ans :

25 ans :

Plus :

Voici le tableau synthétique.

Age	21	22	23	24	25	Plus	Total
Nombre							

- Combien d'étudiants ont l'âge compris entre 22 ans et 25 ans ?]22 25[.
- Combien d'étudiants ont plus de 24 ans ?]24 →.
- Combien d'étudiants ont 23 ans et plus ? [23, →[.
- Combien d'étudiants ont moins de 22 ans ?] → 22[
- Combien d'étudiants ont au moins de 22 ans ?→ 22]
- Combien d'étudiants ont au plus 25 ans ?]25-→[

- Combien d'étudiants ont l'âge ≥ 24 ans ?
- Combien d'étudiants ont l'âge ≤ 21 ans ?
- Combien de filles dans chaque catégorie ?
- Ecrire le rapport de chaque valeur :

$21\ ans = \frac{}{142} = \frac{}{142} \times 100\%$; $23\ ans = \frac{}{142} =$; $25\ ans = \frac{}{142}$;

$22\ ans = \frac{}{142}$; $24\ ans = \frac{}{142}$; $Plus = \frac{}{142}$;

R (21) + R (22)+ R (23)+ R (24)+ R (25)+ R (plus)= Effectif total

Si : Effectif total = N avec n_i = Effectif de valeur (21, 22, ….., plus).

$N = n_1 + n_2 + n_3 + n_4 + n_5 + n_6$.

D'où : $N = n_1 + n_2 + \ldots n_i$ avec *i* le nombre de caractères ou variables. Pour notre exemple i = 6.

2. Si nous considérons les tranches d'âge ci-dessous :

A : 21 à 23 exclu.

B : 22 à 23 exclu.

C : 23 à 24 exclu.

D : 24 à 25 exclu.

E : 25 à plus.

Déterminer les effectifs des valeurs ci-dessous :

22 ans = 21 + 22 ans = Effectif cumulé de 22

23 ans = 21 + 22 + 23 ans = Effectif cumulé de 23

24 ans = 21 + 22 + 23 + 24ans = Effectif cumulé de 24

25 ans = 21 + 22 + 23 + 24 + 25ans = Effectif cumulé de 25

II.2.2. Effectif cumulé croissant ou décroissant

a) Effectif cumulé décroissant

L'expression « avoir X ans ou moins « introduit l'effectif cumulé decroissant.

Exemple : quel est l'effectif cumulé décroissant de la valeur "24 ans".

Réponse : 23 ans + 22 ans + 21 ans.

Dans le cas des séries à caractère quantitatif, l'effectif cumulé décroissant correspond à une valeur ***n*** somme des effectifs des valeurs précédentes, valeur ***n*** exclue.

b) Effectif cumulé croissant

L'expression avoir plus introduit l'effectif cumulé croissant.

Si n est la valeur d'une série la somme des effectifs cumulés croissants correspond aux valeurs qui suivent la valeur, n exclue.

Exemple : Quel est l'effectif cumulé croissant de la valeur "24 ans" ?

Réponse : 25 ans + plus.

II.2.3. Fréquence

a) Définition : la fréquence d'une valeur est le quotient de l'effectif de cette valeur par l'effectif total de la variable. La préférence est d'exprimer cette valeur en pourcentage (%). Toutefois, cette fréquence est toujours comprise entre 0 et 1. Le total en pourcentage sera inférieur ou égal à 100%. Le mieux c'est d'avoir 100%, raison pour laquelle la négligence même d'un dix millième se fait remarquer au total.

Si ni est l'effectif correspondant à une valeur n, la fréquence de la valeur n_i d'une suite N est : $f_i = \frac{n_i}{N}$; avec ni : valeur de la variable, N : valeur de la variable et f_i : la fréquence de la valeur i de l'échantillon.

La détermination de la fréquence absolue peut se faire dans un tableau à simple entrée ou à double entrée.

a. Tableau statistique à simple entrée.

Ce tableau statistique à simple entrée se caractérise par le recours à deux colonnes réservées respectivement :

- L'inscription des valeurs de la variable dans la colonne dite X.
- L'inscription des fréquences absolues correspondantes dans la colonne Y.

Exemple : Soit le tableau ci-dessous traduisant les tranches d'âges des étudiants de G3 Droit.

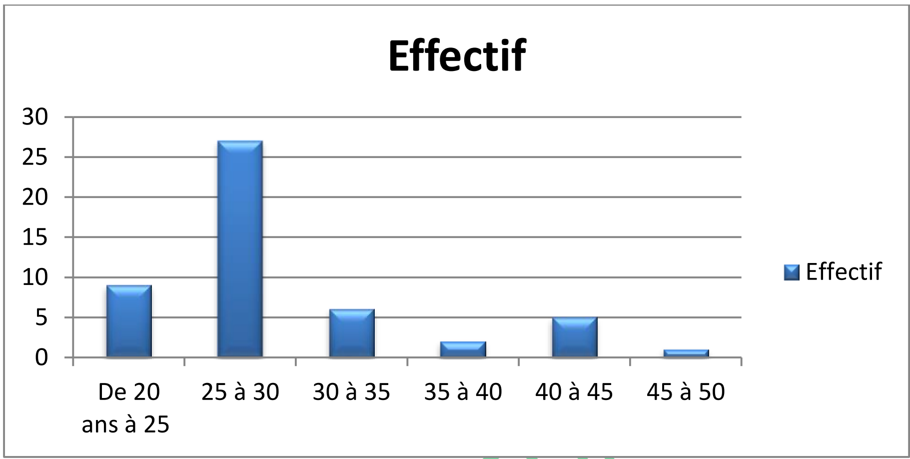
Classe d'âge	Fréquence absolue
De 20 ans à 25	9
25 à 30	27
30 à 35	6
35 à 40	2
40 à 45	5
45 à 50	1
Total	**50**

- Nous pouvons compléter ce tableau en déterminant les fréquences relatives.

Nous aurons :

Fréquence absolue
9/50 = 0,18
0,54
0,12
0,04
0,1
0,02
Total =1

En représentant nous aurons :



Cette représentation est dite en colonne ou histogramme.

b) Tableau statistique à double entrée

Il est toujours possible d'avoir sur une même unité statistique observée deux ou plusieurs variables.

Exemple sur un même étudiant on détermine son âge et son poids. La représentation se fera en recourant au tableau à double entrée.

Ages / Masse	20 à 25	25 à 30	30 à 35	35 à 40	40 à 45	45 à 50	Total X
Moins de 50	2	-	-	-	1	-	**3**
50 à 65	3	5	-	1	-	1	**10**
65 à 75	-	3	8	4	4	3	**22**
Plus de 75	-	1	-	5	7	2	**15**
Total Y	**5**	**9**	**8**	**10**	**12**	**6**	**50**

<u>Interprétation</u> :

- 8 étudiants de 30 à 45 ans pèsent entre 65 et 75 Kgs.

- 7 étudiants de 40 à 45 pèsent plus de 75 Kgs.
- 2 étudiants de 20 et 25 et 1 entre 40 et 45 pèsent moins de 50 Kgs.

Exemples : 1. Lors d'une interrogation 50 étudiants ont eu la côte suivante sur 20.

3 ont eu 0 ; 4 ont eu 1 ; 10 ont eu 6 ; 9 ont eu 8 ; 7 ont eu 9 ; 6 ont eu 10 ; 5 ont eu 11 ;

1 a eu 12 ; 2 ont eu 13 ; 2 ont eu 14 et 1 a eu 16.

Voici comment se présente le tableau de ces 50 étudiants.

Points obtenus	0	1	6	8	9	10	11	12	13	14	16	**Total**
Effectif	3	4	10	9	7	6	5	1	2	2	1	**50**
Fréquence	0,06	0,08	0,2	0,18	0,14	0,12	0,1	0,02	0,04	0,04	0,02	**1**
Fréquence en %	6	8	20	18	14	12	10	2	4	4	2	**100%**

2. Lors d'une interrogation, trente étudiants obtiennent la côte ci-dessous sur 20, 1 a eu 0.

5 ont eu 6 ; 3 ont eu 12 ; 4 ont eu 14 ; 2 ont eu 11 ; 1 a eu 16 ; 4 ont eu 9 ; 9 ont eu 10 ;

1 a eu 8.

Voici le tableau synthétique.

Points obtenus	0	6	8	9	10	11	12	14	16	**Total**
Effectif	1	5	1	4	9	2	3	4	1	**30**
Fréquence	0,03	0,166	0,033	0,133	0,3	0,066	0,1	0,133	0,033	**+0,991**
Fréquence en %	3,3	16,6	3,33	13,33	30	6,66	10	13,33	3,333	**+3,3=99,9%**

b. Fréquence cumulée

On appelle fréquence cumulée de la valeur n, le quotient de l'effectif cumulé de la valeur n par l'effectif total.

Exemple : Soit l'exemple 1 dont l'effectif est 50, quelle est la fréquence cumulée de la valeur n=11.

$f_{11} = 3 + 4\ 10 + 9 + 7 + 6 + 5 = 44.$

Donc la fréquence $f_{11} = \frac{44}{50} = 0{,}88$ ou 88%.

FCC de 11 dans le même exemple donne $1 + 2 + 2 + 1 = 06$.

$f_{11} = \frac{06}{50} = 0{,}12 = 12\%.$

Total $FCC_{11} + FCD_{11} = 100\% \Longrightarrow 88 + 12 = 100\%.$

Exercices : 1. Un sondage réalisé auprès de 80 étudiants de la faculté de droit de l'université de la ville, portant sur le nombre de téléviseurs par famille a donné les résultats suivants :

Nbre de télé par famille	Effectif n_i	Fréquence f_i	Effectif cumulé croissant	FCC
0	2			
1	25			
2	28			
3	15			
4	8			
5	2			

Calculez f_i, ECC et FCC.

2. Dans cet auditoire de G3 Droit, le relevé des notes, sur 20, attribuées après un contrôle de statistique à la première session indique :

4	11	14	6	8	8	18	10	5	5
10	10	12	2	17	5	125	14	4	17
8	13	15	18	14	15	4	9	17	13
5	11	4	16	2	12	13	10	17	5
15	8	11	14	8	6	15	19	9	13

a. Combien d'entre eux ont obtenu la note 10 ?, la note 14 et la note 19 ?

b. A l'aide d'un tableau, déterminer les effectifs, les fréquences, les ECD, les FCD et les FCC.

c. Dites quelle est la FCC de la valeur 10 ?

Solutions.

III. REPRESENTATIONS GRAPHIQUES

a) Définition : les graphiques sont les représentations imagées des résultats. Elles sont une illustration des données contenues dans le tableau et sont plus parlant dans la facilitation de la compréhension pour les non-initiés. La maîtrise de la géométrie étant difficile partant des instruments usités, les graphiques finissent par être moins précises d'où l'exigence des chiffres.

b) Il y a plusieurs façons de représenter les graphiques dont les plus usitées sont :

Qualitatif :

- Par diagramme à colonne (Histogramme) ou en bande.
- Par diagramme circulaire.

Quantitatif :

- Par diagramme en bâtons (bâtonnets).
- Par polygone
- Représentation quelconque.

III.1 Graphiques des caractères qualitatifs

1° Histogramme ou diagramme en bandes

Un histogramme est constitué de rectangles accolés ayant pour bases les amplitudes des classes et dont l'aire est proportionnelle aux effectifs ou aux fréquences.

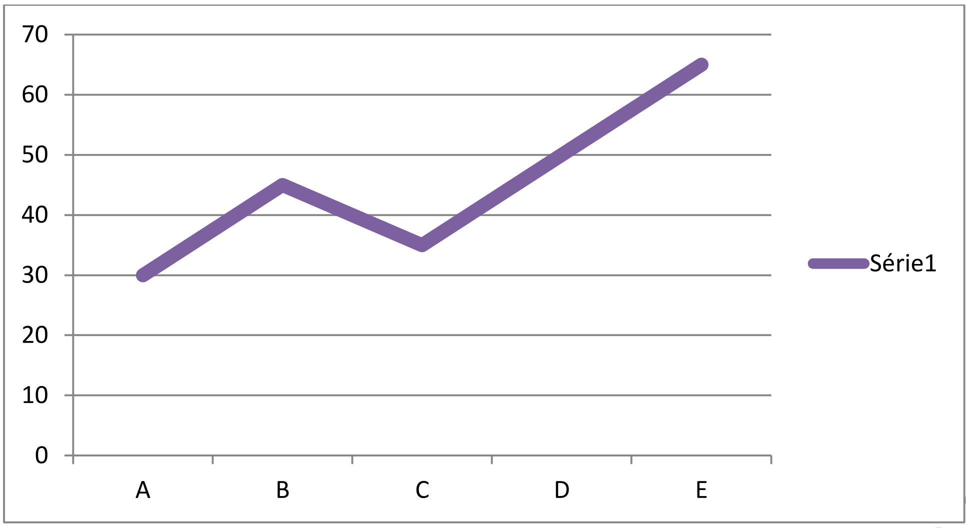
En principe si toutes les classes ont la même amplitude, tous les rectangles auront la même base et la même hauteur. L'exemple de l'âge des étudiants de G3 Droit pris ci-haut en est une illustration.

Pour lire un histogramme, il faut faire correspondre les valeurs de l'abscisse (X) sur celles de l'ordonnée (Y).

- Lorsque les valeurs de l'abscisse sont unitaires nous aurons une courbe à partir des points des jonctions abscisse –ordonnée.

Exemple : Si la masse de 5 étudiants est traduit par les valeurs ci-dessous.

A : 30 Kgs ; B : 45 Kgs ; C : 35 Kgs ; D : 50 Kgs et E : 65 Kgs ;



- Lorsque les valeurs de l'abscisse sont des tranches de valeurs quelconques nous aurons les colonnes.

Exemple : 15 poids pris au hasard pèsent respectivement : 19, 11, 13, 15, 13, 16, 9, 10, 14, 15, 20, 12, 18, 15 et 16 Kgs.

Les valeurs extrêmes sont : 9 et 20.

L'intervalle contenant ces valeurs sera : $[9, 20]$ ou $]8, 21[$ ou $[9, 21[$ ou $]8, 20]$.

L'amplitude de l'intervalle sera : 21 – 9 = 12 ou 20 -8 = 12 et non 20 -9 ou 21 – 8 parce que les bornes sont exclues.

La longueur de l'intervalle égale l'amplitude moins la borne inférieure. 12 – 9 =3.

Le nombre de classe sera égal au nombre d'intervalle contenu dans le tableau.

9 plus 3 = 12
12 plus 3 = 15 d'où le nombre de classe = 4.
15 plus 3 = 18
18 plus 3 = 21.

Le tableau des effectifs et des fréquences sera :

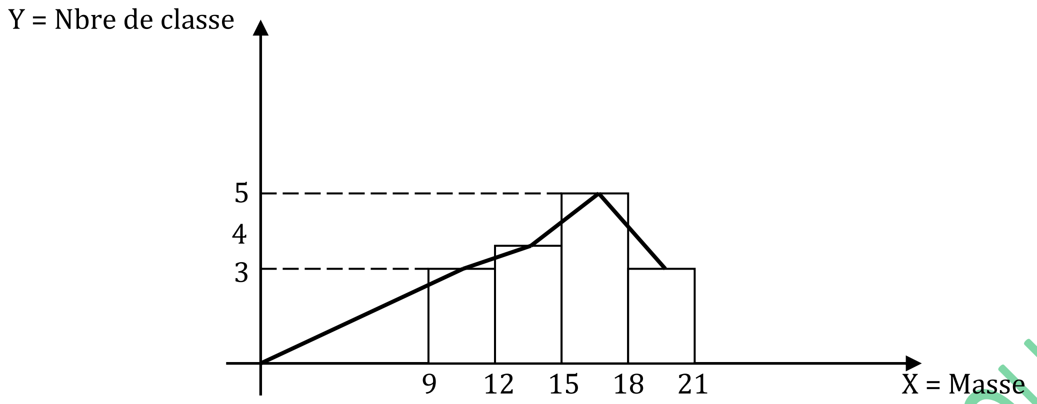
$[9, 12[\ = 9, 10, 11\ d'où\ 3\ \Longrightarrow \frac{3}{15} = 0{,}2$ ou 20%

$[12, 15[\ = 12, 13{,}13, 14\ d'où\ 4\ \Longrightarrow \frac{4}{15} = 0{,}27$ ou 27%

$[15, 18[\ = 15, 15, 15, 16, 16\ d'où\ 5\ \Longrightarrow \frac{5}{15} = 0{,}33$ ou 33%

$[18, 21[\ = 18, 19, 20\ d'où\ 3\ \Longrightarrow \frac{3}{15} = 0{,}2$ ou 20%.

Histogramme

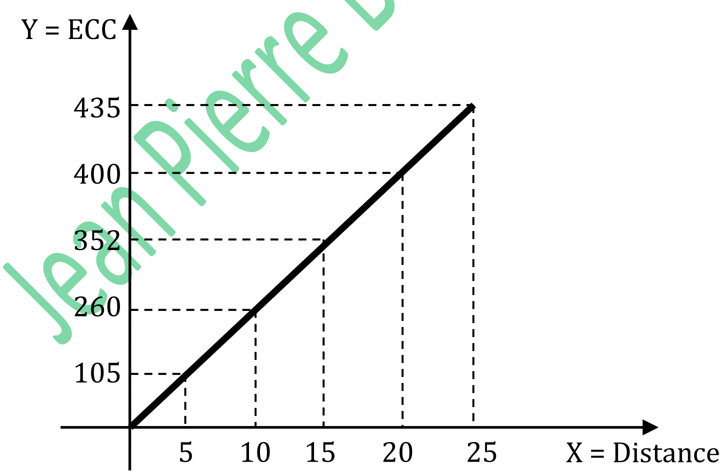


Dans la représentation, il y a lieu d'obtenir les effectifs cumulés croissants ou décroissants.

Exemple : Soit le tableau ci-dessous d'amplitudes 5 représentants la distance domicile université de 435 étudiants.

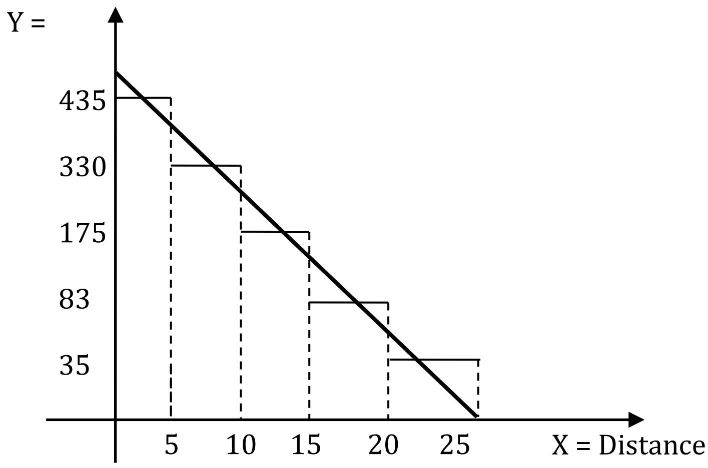
Classe (en Km)	Effectif n_i	Fréquence f_i	Fréquence %	ECC	ECD
A [0, 5[	105	0,24	24	105	435
B [5, 10[	155	0,36	36	260	330
C [10, 15[	92	0,21	21	352	175
D [15, 20[	48	0,11	11	400	83
E [20, 25[	35	0,08	8	435	35
TOTAL	435	1	100		

Polygone croissant



N.B. : La courbe est d'autant vraie que les indications en axes sont vraies.

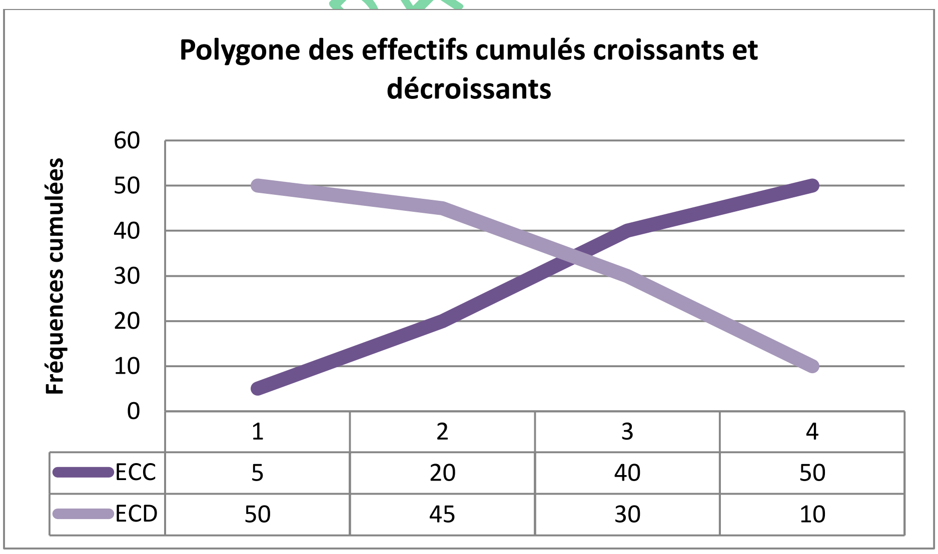
Polygone décroissant.



Exercice

- Si les classes n'ont pas la même amplitude, il faut le faire ressortir dans le schéma.
- Soit le tableau ci-dessous de 50 individus en série statistique « âge » avec amplitude différentes.

Classe (en Km)	Effectif n_i	ECC	ECD
[5, 10[	5	5	50
[10, 20[	15	20	45
[20, 30[	20	40	30
[30, 50[	10	50	10
TOTAL	50		



Le polygone des effectifs cumulés croissants est formé des segments joignant les points ayant pour :

- Abscisse, la borne supérieure de la classe.
- Ordonnée, l'effectif cumulé croissant de la classe.

Le premier point étant placé sur l'axe des abscisses et ayant référence la borne inférieure de la première classe.

En situation du polygone des effectifs cumulés décroissants, les segments joignent les points ayant pour :

- Abscisse, la borne inférieure de la classe.
- Ordonnée, l'effectif cumulé décroissant de la classe.

Ici, le premier point est placé sur l'axe des ordonnées et a point pur abscisse la borne inférieure de la première classe.

2° Diagramme circulaire (semi-circulaire)

Devant cette situation de cercle, se profile deux solutions :

- Représenter les valeurs en pourcentage. A ce moment-là tout le cercle représente 100%. $Portion = \frac{Effectif\ .100}{Echantillon\ de\ la\ série}$%.

Exemple : Dans une série de valeur X une couleur Y représente une valeur Z. quel est le pourcentage.

$Y = \frac{z.100}{X}\%$

- Une série statistique de 150 stylos, la couleur verte représente 36 stylos. Quelle est la proportion.

$Y = \frac{36.100}{150} = 24\%$. 24% représente près de ¼ du cercle.

- Représenter les valeurs en degré. A ce moment on considère que le cercle représente 360°.

Exemple : $Portion = \frac{36\ .360}{150} = 86{,}4°$. 86,4° représente aussi près de ¼ du cercle.

$Portion = \frac{Proportion\ en\ \%\ .360°}{x}$= Proportion en degré.

Exercice : Dans une formation médicale sur 38.853 examens de Labo.

15.420 sont de la goutte épaisse (GE)

18.031 sont pour l'hémoglobine (S)

5.402 sont pour la vitesse de sédimentation (VS)

Calculer les angles au centre correspondant aux activités de ce labo en degré et en pourcentage. Dresser le diagramme.

$$GE = \frac{15.420 \times 100}{38.853} = 39{,}69\%$$

$$GE = \frac{15.420 \times 360}{38.853} = 142{,}92°$$

$$S = \frac{18.041 \times 100}{38.853} = 46{,}41\%$$

$$S = \frac{18.041 \times 360}{38.853} = 167{,}07°$$

$$VS = \frac{5.402 \times 100}{38.853} = 13{,}9\%$$

$$S = \frac{5.402 \times 360}{38.853} = 50{,}05°$$

IV. GRAPHIQUES DES CARACTERES QUANTITATIFS.

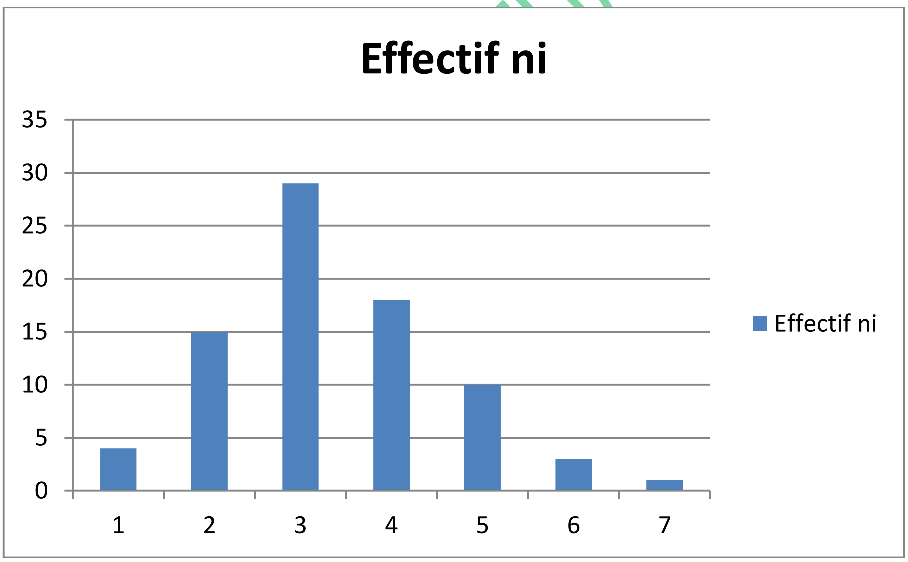
Dans la représentation graphique à caractères quantitatifs, deux possibilités s'offrent à nous :

a) Graphique des caractères quantitatifs discontinus.

L'histogramme est constitué des bâtons libres les uns des autres.

Exemple : A la faculté de droit, on détermine le nombre d'enfants à charge pour 80 corps scientifique et académique. On trouve la série statistique ci-dessous :

Nbre d'enfants à charge	Effectif n_i	Fréquence f_i	Effectif cumulé croissant
0	4	0,050	4
1	15	0,188	19
2	29	0,363	48
3	18	0,0225	66
4	10	0,125	76
5	3	0,038	79
6	1	0,013	80
TOTAL	80	1	



b) Graphiques des caractères quantitatifs continus

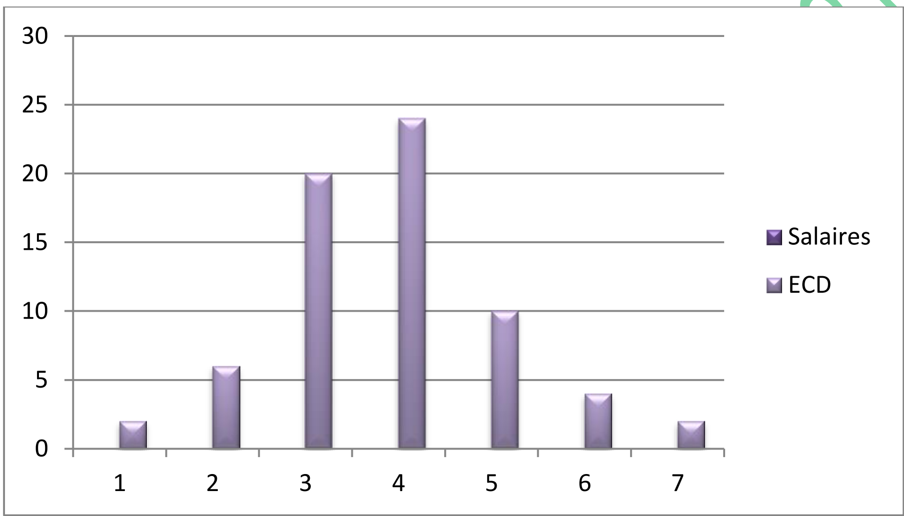
L'histogramme est constitué d'une suite de rectangle dont les bases coïncident avec les classes divisant le domaine de variation de la variable et dont les hauteurs soient telles qu'effective (amplitude). Effectif amplitude $= \frac{n_i\,(efefctifs)}{Amplitude}$

Exemple : Soit une entreprise de 250 ouvriers recevant des salaires dans les tranches ci-dessous d'amplitudes variables.

Salaires	Amplitude de base des classes	Effectif	EC par classe
[47,5 – 52,50[	5	10	2
[52,5 – 57,50[	5	30	6
[57,5 – 60,50[	3	60	20
[60,5 – 63,50[	3	72	24
[63,5 – 67,50[	4	40	10
[67,5 – 73,50[	6	24	4
[73,5 – 80,50[	7	14	2
TOTAL		$\sum_{n_i} = 250$	

Graphique

EC. Par classe



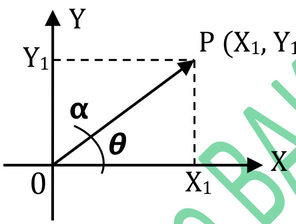
N.B. : Obtenir que les colonnes aient une base correspondant à la valeur de l'amplitude.

V. Représentations quelconques

A ce niveau, nous avons le choix entre la représentation en coordonnées polaires, circulaires, logarithmiques, barres, carré ou triangle.

a) Coordonnées polaires

On part de système cartésien (X, Y)



$\boldsymbol{\theta}$ et $\boldsymbol{\alpha}$ noté $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ sont les coordonnées polaires de P. $P = (\boldsymbol{\theta}, \boldsymbol{\alpha})$.

Si N est l'effectif de la série quelconque avec $Y_1 = \boldsymbol{\beta}$.

Angle $\boldsymbol{\theta} = \frac{360°}{\beta}$

Angle $\alpha = 90° - angle\ \theta$.

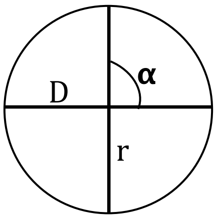
b) Coordonnées logarithmiques

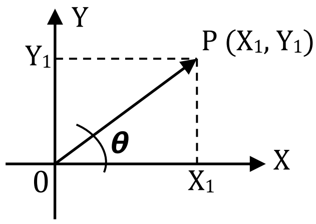
Supposons qu'on veut représenter les ventes exprimées en milliers (000) de francs en fonction des mois, on peut utiliser deux axes cartésiens pour les mois logarithmiques et pour les ventes.

On parle alors de l'échelle semi-logarithmique.

c) Coordonnées circulaires (diagramme circulaire ou secteur)

La représentation part du principe selon lequel un cercle couvre 360°, 4 fois les angles droits contenus dans des diamètres étant le double du rayon partant du centre du cercle.





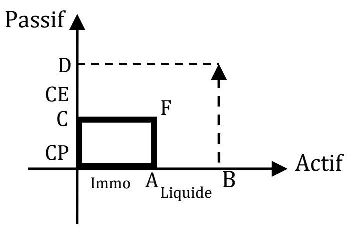
d) Diagramme carré

Soit une entreprise dont l'actif reprend 40% d'immobilisé et 60% de liquide avec un passif de 30% de capitaux propres pour 70% des capitaux étrangers.

Actif immobilisé 40% = A — Passif Capitaux propres 30% = C

Liquide 60% = B — Capitaux étrangers 70% = D

Le diagramme carré sera :



F représente A, B, C et D soit 4 grandeurs de l'entreprise.

e) Diagramme triangulaire

Le recours à la représentation triangulaire est conditionné que si les mesures des phénomènes à présenter peuvent être ventilées en trois composantes dont le total des mesures reste constant.

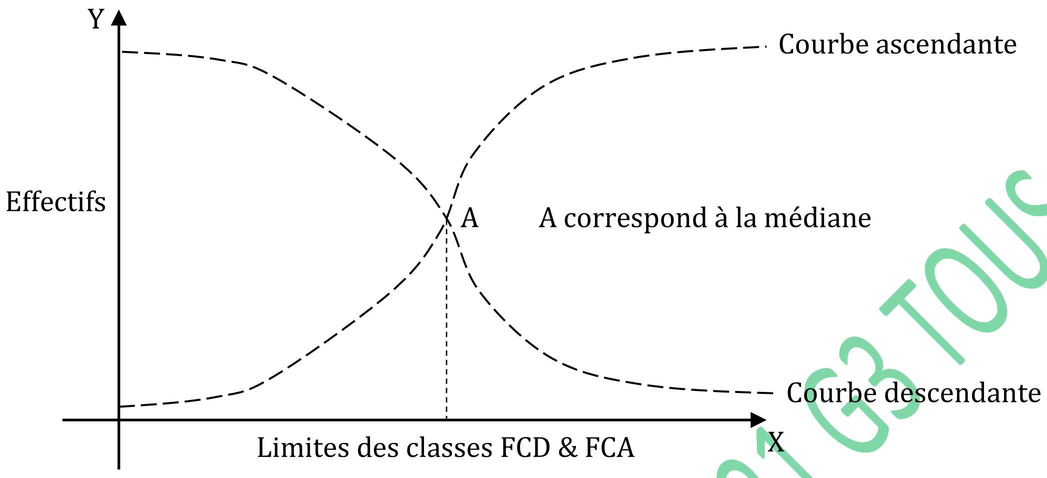
Exemple : Le passif du bilan reprend (pas toujours) les capitaux propres, les dettes à long terme et les dettes à court terme.

f) Courbe des fréquences cumulées dites ogives

Cette courbe est obtenue à partir de deux fréquences :

- Fréquences cumulées descendantes (FCD) qui sont les cumuls des fréquences des différentes classes de la distribution partant de la 1ère classe à la dernière.
- Fréquences cumulées ascendantes (FCA) qui sont, les cumuls des fréquences des différentes classes de la distribution partant de la classe « *n* », la dernière, à la première classe.

Les limites inférieures des classes sont portées sur l'axe des abscisses X pour la courbe des fréquences cumulées descendantes et les limites supérieures des classes sur l'axe des ordonnées des effectifs cumulés pour les fréquences cumulées ascendantes.



VI. REGROUPEMENT EN CLASSES OU ELEMENTS CARACTERISTIQUES DES SERIES STATISTIQUES

a) Notions

Il n'est pas toujours facile d'utiliser les données non groupées dans une étude statistique.

Pour rendre la classification possible, il faut recourir parfois à leur nature, à leur positionnement bref à la manière que chaque élément est pris par rapport aux autres.

Partant d'un auditoire de 50 étudiants où chacun a obtenu une côte différente dans un test (sur 50 points) de mathématique, l'effectif de chaque valeur dans un tableau pour les données non groupées sera égal à 1. La crainte est que ce tableau soit très grand et encombrant si les données ne sont pas regroupées.

Exemples : Dans un auditoire de 50 étudiants, le professeur décide de les répartir en groupe. Pour cela, il mesure la taille (en mètre) de ses étudiants et a trouvé :

1,54	1,53	1,57	1,59	1,54	1,55	1,60	1,63	1,59	1,67
1,61	1,63	1,67	1,69	1,68	1,69	1,70	1,72	1,73	1,64
1,74	1,78	1,55	1,76	1,75	1,79	1,66	1,77	1,67	1,69
1,59	1,76	1,64	1,67	1,69	1,79	1,76	1,59	1,74	1,78
1,73	1,68	1,65	1,71	1,78	1,65	1,57	1,58	1,65	1,54

Il décide de constituer les groupes et en réalise trois :

- Le groupe des « petits » formé des étudiants dont la taille appartient à [1,50 – 1,60[.
- Le groupe des « moyens » formé des étudiants dont la taille appartient à [1,60 – 1,70[.
- Le groupe des « grands » formé des étudiants dont la taille appartient à [1,70 – 1,80[.

Chaque groupe constitue une classe. D'où trois classes et obtient une série statistique groupée ci-dessous :

Classe	[1,50 – 1,60[	[1,60 – 1,70[	[1,70 – 1,80[	Total
Effectif n_i	13	20	17	50
Fréquence en %	26	40	34	100%

Les valeurs extrêmes sont : 1,53 et 1,79.

L'intervalle contenant ces valeurs est [1,50 – 1,80[.

L'amplitude de cet intervalle est : 1, 80 – 1,50 = 0,30.

L'amplitude de la classe : 1,60 – 1,50 ; 1,70 -1,60 ; 1,80-1,70 = 0,10.

36	54	23	48	54	40	53	50
20	52	26	63	51	51	34	50
19	59	28	43	75	29	17	65
18	68	25	40	80	27	32	34
30	74	20	36	73	57	17	25
35	76	32	44	23	52	49	25
40	82	86	35	17	42	42	51
42	51	72	46	15	27	23	86

Déterminer les classes, l'amplitude de cet intervalle, le centre de classe X_i.

Solution :

- Intervalle de classe proposé égal à 8.
- L'intervalle contenant ces valeurs peut être noté [15; 87]
- L'amplitude de cet intervalle sera : 87 -15 = 72.
- La longueur de l'intervalle de la classe égale :

$$\frac{Amplitude\ de\ l'intervalledes\ valeurs\ extrêmes}{Nombre\ d'intervallesde\ classes\ proposées} = \frac{71}{8} = \sim 9$$

La longueur d'intervalle est de 9.

D'où :

Ages	Effectif n_i	Centre de classe Xi
$[15 - 24[$	11	$19{,}5 = \frac{15+24}{2}$
$[24 - 35[$	13	29,5
$[35 - 44[$	12	39,5
$[44 - 55[$	14	49,5
$[55 - 64[$	3	59,5
$[64 - 75[$	5	69,5
$[75 - 84[$	4	79,5
$[84 - 95[$	2	89,5

Attention :

- Pour calculer le centre de la classe jamais faire a +b : 2 mais faire a +b = réponse et cette réponse divisée par 2 c'est-à-dire $\frac{a+b}{2}$
- Dans le cas des séries à caractère quantitatif, l'effectif cumulé décroissant correspondant à une classe est la somme des effectifs des classes précédentes ; classe considérée incluse.

Exemple : la valeur 40 est dans l'intervalle $[35 - 44[$.

= 12+14+3+5+4+2 = 40.

- De la même manière l'effectif cumulé décroissant correspondant à une classe est le quotient des effectifs cumulés de la classe par des classes précédentes, classe considérée incluse, par l'effectif total.

Exemple : Classe $[35-44[$

- Le même raisonnement est appliqué pour les séries à caractère quantitatif à l'effectif cumulé croissant.

b) Eléments caractéristiques des séries statistiques

Les éléments caractéristiques d'une série statistique sont de plusieurs ordres. Les plus remarquables sont : le mode, la médiane, la moyenne, la médiale et les centiles.

1) Le mode $(\hat{X})$ dit aussi la dominante est la valeur de la variable qui représente l'effectif le plus élevé, c'est-à-dire la fréquence la plus élevée.

Pour déterminer la répartition des données en classes on recourt à deux méthodes que sont :

- Méthode de l'amplitude entière.

L'étendue d = X_{max} - X_{min}

Le nombre de classe K dans l'intervalle $[7,12]$ est donné par : $\frac{d}{K}$.

Pour recourir à cette méthode, il faut que le résultat soit entier. Dans le cas contraire, la méthode ne sera pas d'application.

- Méthode de STURGES dite aussi de LIORZOU lorsque le nombre de classes est plus élevé et pour éviter de ne pas atteindre l'entier on utilise la formule de LIORZOU par la relation ci-dessous :
  - Le nombre de classes est donné par $K = 1 + \frac{10\ log_n}{3}$
  - L'amplitude $a = \frac{d}{K-1}$ avec d = X_{max} - X_{min} (extrêmes)
  - La limite inférieure de la première classe: $L_{inf} = X_{min}$-a/2.

Dans le cas d'une série continue, le mode est déterminé par la relation : $\frac{1}{X} L_i + a \frac{\Delta}{\Delta_{1+}\Delta_2}$

Li : désigne la limite inférieure de la classe modale.

Δ_1 : Différence entre les effectifs de la classe modale et ceux de la précédente.

Exercice. Les notes finales en statistique descriptive obtenues par 80 étudiants de G3 Droit sont reportées dans le tableau suivant. Regroupez cette série en classes en utilisant la technique de regroupement de LIORZOU.

68	84	75	82	68	90	62	88	76	93
73	79	88	73	60	93	71	59	85	75
61	65	75	87	74	62	95	78	63	72
66	78	82	75	94	77	69	74	68	60
96	78	89	61	75	95	60	79	83	71
79	62	67	97	78	85	76	65	71	75
65	80	73	57	88	78	62	76	53	74
86	67	73	81	72	63	76	75	85	77

<u>Solution</u>

N = 80

X_{max} : 97 et X_{min} : 53

$$K = 1 + \frac{10\ log_n}{3} = 1 + \frac{10\ log_{80}}{3} = 7{,}34 \sim 7$$

$$a = \frac{d}{K-1} = \frac{44}{7-1} = 7{,}33 \sim 7$$

$$L_{inf} = X_{min} - \frac{a}{2} = 53 - \frac{7}{3} = 49{,}5$$

D'où le tableau statistique ci-dessous :

Intervalles des classes	Centre de classe X_i	Effectif n_i	f_i	f_i en %
[49,5 – 56,5[	53	1	0,0125	1,25
[56,5 – 63,5[	60	13	0,1625	16,25
[63,5 – 70,5[	67	13	0,1625	16,25
[70,5 – 77,5[	74	28	0,03375	33,75
[77,5 – 84,5[	81	12	0,1500	15,
[84,5 – 91,5[	88	9	0,1125	11,25
[91,5 – 98,5[	95	5	0,0625	6,25
		80	1	100

2. Soit une série statistique ci-dessous relative au prix du même article dans la ville de Lubumbashi observé dans 20 points de vente.

47, 28, 51, 38, 45, 29, 31, 32, 54, 38, 39, 33, 41, 34, 34, 34, 34, 34, 34, 34, 35, 47, 53, 40, 47, 47, 48, 48, 35, 35, 35, 35, 35, 36, 50, 36, 36, 36, 36, 36, 36, 36, 37, 52, 38, 37, 37, 37, 31, 31, 37, 37, 37, 37, 37, 37,38, 35, 35, 38, 38, 38, 41, 41, 41, 41, 32, 41, 38, 38, 45, 45, 39, 45, 47, 30, 40, 40, 40, 39, 39, 44, 44, 40, 40, 41, 41, 33, 33, 41, 36, 41, 41, 41, 41, 42, 42, 30, 38, 33, 34, 41, 41, 41, 38, 38, 38, 42, 42, 42, 42, 42, 36, 43, 36, 42, 42, 42, 42, 42, 42, 40, 40, 40, 40, 42, 42, 43, 32, 43, 43, 43, 43, 43, 43, 45, 43, 43, 43, 44, 44, 44, 32, 33, 33, 44, 26, 27, 35, 39, 39, 39, 35, 44, 44, 44, 44, 45, 46, 45, 40, 40, 40, 40, 45, 39, 39, 39, 39, 45, 46, 37, 39, 39, 39, 39, 39, 45, 37, 37, 46, 46, 46, 46, 46, 46, 47, 48, 48, 44, 40, 40, 43, 32, 49, 49, 49.

Regroupez cette série en classe en utilisant la technique de regroupement de LIORZOU.

Solution

- Extrêmes : 54 et 26.
- Nombre de classes : $K = 1 + \frac{10\, log_{200}}{3} = 9$
- L'étendue de la série : d = 54 -26= 28
- L'amplitude $a = \frac{d}{K-1} = \frac{28}{8} = 3{,}5$
- $L_{inf} = X_{min} - \frac{a}{2} = 26 - \frac{3{,}5}{2} = 24{,}25$

Ainsi on formera les classes ci-après : $[24{,}25 - 27{,}75[$, $[27{,}75 - 31{,}25[$, $[31{,}25 - 34{,}75[$, $[34{,}75 - 38{,}25[$, $[38{,}25 - 41{,}75[$, $[41{,}75 - 45{,}25[$, $[45{,}25 - 48{,}75[$, $[48{,}75 - 52{,}25[$, $[52{,}25 - 55{,}75[$.

Tableau de dépouillement

Classes	Pointage	Effectifs
$[24{,}25 - 27{,}75[$		2
$[27{,}75 - 31{,}25[$		7
$[31{,}25 - 34{,}75[$		19
$[34{,}75 - 38{,}25[$		49
$[38{,}25 - 41{,}75[$		48
$[41{,}75 - 45{,}25[$		49
$[45{,}25 - 48{,}75[$		18
$[48{,}75 - 52{,}25[$		6
$[52{,}25 - 55{,}75[$		2
TOTAL		200

3. Soit la distribution de la pression systolique pour 63 individus observés.

102, 138, 190, 122, 128, 112, 128, 116, 134, 104, 116, 152, 134, 132, 130, 114, 118, 136, 108, 108, 128, 118, 134, 178, 134, 162, 120, 98, 144, 118, 114, 118, 138, 134, 108, 96, 142, 122, 146, 126, 176, 104, 112, 140, 102, 142, 98, 146, 92, 112, 152, 116, 118, 128, 116, 134, 108, 134, 124, 124, 114, 154.

Regroupez cette série en classe par la technique de LIORZOU.

Solution

- Extrêmes : 190 et 92.
- Nombre de classes : $K = 1 + \frac{10\, log_{63}}{3} = 4$
- L'étendue de la série : d = 190 -92 = 98
- L'amplitude $a = \frac{d}{K-1} = \frac{98}{4-1} = 32{,}66$
- $L_{inf} = X_{min} - \frac{a}{2} = 92 - \frac{32{,}66}{2} = 75{,}67$

Calcul du Mode

1. Cas discret

Dans la situation d'un cas discret c'est-à-dire dont les données sont en vrac, le mode $\hat{X}$ est la valeur qui correspond au maximum de la courbe de fréquence.

Exemple : 2, 2, 5,7, 9, 9, 9, 9, 10, 10, 11 ,12 ,18. Le mode de cette suite est égal à 9 parce que 9 est plus fréquent, il se répète 4 fois.

2. Cas continu

Lorsque les données sont regroupées en classe, l'application du mode ne se fait qu'à condition d'attribuer au mode la valeur de la classe à plus grand effectif appelée "classe modale". Si la distribution a plusieurs modes, on parle de la distribution "plurimodale".

La valeur du mode s'obtient à partir de la distribution d'effectifs ou d'histogramme dont la procédure est :

- La détermination de la classe dominante dite aussi classe à plus grand effectif.
- Par interpolation des points intermédiaires à la courbe, déterminer la valeur du mode à partir de la formule $\hat{X} = L_i + a\frac{\Delta_1}{\Delta_1+\Delta_2}$

  Li : limite inférieure de classe modale

  a= amplitude de la classe modale

  Δ_1= différence entre les effectifs de la classe modale et ceux de la classe précédente.

  Δ_2= différence entre les effectifs de la classe modale et ceux de la classe suivante.

N.B. : Si la distribution a des amplitudes différentes, l'amplitude à considérer est celle de la classe modale.

Exemple : Calculer le mode de la distribution suivante.

Limites réelles	n_i	fcd
[9,3 – 9,7[	0	0
[9,7 – 10,1[	2	2
[10,1 – 10,5[	5	7
[10,5 – 10,9[	12	19

[10,9 – 11,3[	17	36
[11,3 – 11,7[	14	50
[11,7 – 12,1[	6	56
[12,1 – 12,5[	3	59
[12,5 – 12,9[	1	60
[12,9 – 13,3[	0	60

Solution

L_{sup} - L_{inf} = 9,7 – 9,3 = 0, 4
a= 0,4
n_i = 17
Δ_1= 17 -12 = 5
Δ_2= 17 -14 = 3
L_i = 10,9

$$\hat{X} = L_i + a\frac{\Delta_1}{\Delta_1 + \Delta_2} = 10{,}9 + 0{,}4\frac{5}{3+5} = 11{,}5$$

Explications :

- Pourquoi 17, c'est la fréquence la plus grande de cette distribution.
- Pourquoi 10,9, c'est la borne inférieure correspondant à l'intervalle qui contient la classe modale.
- Pourquoi 0,4, c'est la différence entre la limite supérieure et la limite inférieure d'une classe.
- Pourquoi 12, c'est la valeur qui précède celle de la classe modale.
- Pourquoi 14, c'est la valeur qui suit celle de la classe modale.

L'avantage : la détermination de la valeur modale est aisée dans un tableau statistique et le mode nous permet de voir la valeur qui revient souvent.

L'inconvénient : lorsque la fréquence correspondante du mode est nettement supérieure aux autres fréquences, le mode n'a aucune signification. Le mode perd toute sa valeur (signification) dans une série bimodale ou multimodale.

2° La médiane ($\tilde{X}$)

Elle est une valeur de la variable aléatoire qui, après séparation, se trouve au milieu de la série. Soit n le nombre d'éléments de la série :

- Si n est impair : la médiane est donné par la relation $\tilde{X} = \frac{X_{n+1}}{2}\frac{=n+1}{2}$

Exemple : 9, 10, 11, 12, 9, 11, 10. Quelle est la médiane ?

Solution

$$n = 7\ ;\ n+1 = 8 \text{ et } \frac{n+1}{2} = \frac{8}{2} = 4 \Longrightarrow \tilde{X} = X_4$$

La médiane est à la position 4, question de reprendre la série par ordre croissant pour déterminer la médiane. 9, 9, 10, 10, 11, 11, 12, la position 4 correspond à 10 qui est la médiane.

- Si n est pair la médiane est déterminée par la relation $\tilde{X} = L_i + a\left[\frac{\frac{n}{2} - fcd_i}{n_i(med)}\right]$ ou $L_S - a\left[\frac{\frac{n}{2} - fcd_i}{n_i(med)}\right]$

Avec Li= limite inférieure

Ls = Limite supérieure

a = amplitude de la classe médiane

Fca = fréquences cumulées ascendantes

Fcd = fréquences cumulées descendantes

Exemple : 9, 8, 7, 10, 9, 10.

n = 6 (pair). Quelle est la médiane ?

<u>Solution</u>

$L_i = 7$

$a = \frac{d}{k-1} = \frac{10-7}{k-1} = \frac{3}{k-1} \Longrightarrow a = \frac{d}{k} \Longrightarrow d = 3; k = 6 \Longrightarrow \frac{3}{6-1} = 0{,}6$

La médiane peut-être déterminée graphiquement en traçant les courbes des fréquences cumulées ascendantes et descendantes, leur intersection donne la médiane. Cette façon de faire présente des avantages dans la facilité de calcul après classement des données et donne l'idée d'une valeur centrale.

Toutefois, se dépendance au rang des unités statistiques constitue un inconvénient.

Calcul de la médiane

- **Cas d'une distribution non groupée**

1° Déterminer la médiane $\tilde{X}$ de la série : 9, 10, 11, 12, 9, 11, 14, 15, 13, 16, 14.

<u>Solution</u>

Procéder à la sériation c'est-à-dire ordonné la série. 9, 9, 10, 11, 11, 12, 13, 14, 14, 15, 16.

$n = 11 \Longrightarrow n+1 = 12 \Longrightarrow \frac{n+1}{2} = \frac{11+1}{2} = 6$

$\widetilde{Xi} = 6 \Longrightarrow \tilde{X} = 12.$

6 est la position dans la série, qui correspond à 12.

2° Si le nombre d'éléments de la distribution est paire, la médiane s'obtient en faisant la moyenne de deux termes du milieu $X_{\frac{n}{2}}$ et $X_{\frac{n+1}{2}} \Longrightarrow \widetilde{X} = \frac{X_{\frac{n}{2}}+X_{\frac{n}{2}+1}}{2}$

$[X_{\frac{n}{2}} ; X_{\frac{n+1}{2}}]$est dit intervalle médian.

Exemple : calculer la $\tilde{X}$ pour 15,18,16,17,15,18,21,22,19,20,22,13.

<u>Solution</u>

Sériation : 15,15,16,17,18,18,19,20,21,22,22,23.

n = 12 ⟹ n/2 = 6.   6$^{\text{ème}}$ position dans la série qui correspond à 18.

$\frac{n+1}{2} = \frac{13+1}{2}$=7, c'est-à-dire 7$^{\text{ème}}$ position qui correspond à 19. D'où $\widetilde{X} = \frac{18-19}{2} = 18{,}5$.

[18, 19] est appelé intervalle médian.

3° Lorsque la variable est discrète et que le nombre des termes est impair (2n+1), on adopte le même raisonnement que pour la variable continue, si par contre le nombre des termes est pair, la méthode d'interpolation ne convient pas, d'où la distribution aura deux médianes. En revanche le choix du terme du milieu parmi ces deux valeurs semble représentatif pour l'ensemble d'observations.

• **Cas d'une distribution groupé**

La recherche de la $\tilde{X}$ se fait par interpolation ou par la méthode graphique.

Exemple

Intervalles des classes	X_i	n_i	fcd	fca
[9,25 – 9,75[	9,5	2	2	60
[9,75 – 10,25[	10	5	7	58
[10,25 – 10,75[	10,5	12	**19**	53
[10,75 – 11,25[	11	**17**	36	41
[11,25 – 11,75[	11,5	14	50	24
[11,75 – 12,25[	12	6	56	10
[12,25 – 12,75[	12,5	3	59	4
[12,75 – 13,25[	13	1	60	1
Total		**60**		

Li = limite inférieure = 10,75

Ls = limite supérieure = 11,25 dans l'intervalle médian [10,75 – 11,25[

a = amplitude de la classe médiane = 0,5

fca = fréquences cumulées ascendantes : 41

fcd ; fréquences cumulées descendantes = 19   N.B. : pas prendre 36.

n = effectif des variables (fréquences) = 60

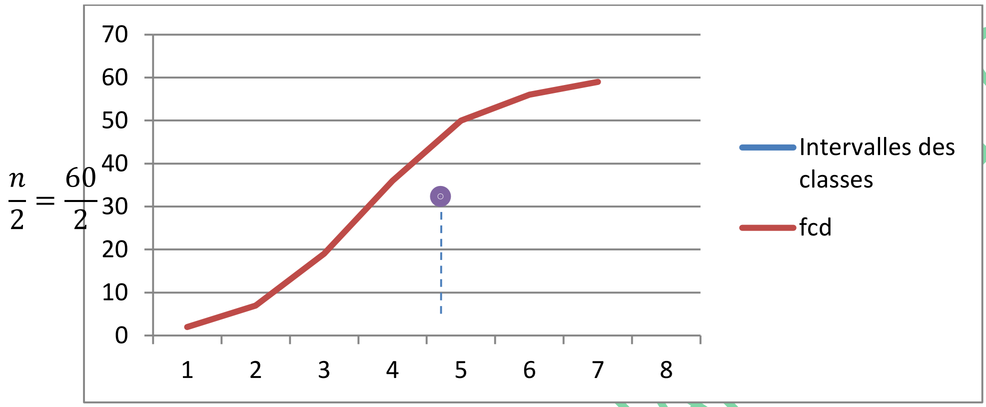
ni = valeur de l'intervalle médian = 17

$$\tilde{X} = L_i + a\left[\frac{\frac{n}{2} - fcd_i}{n_i(med)}\right] = 10{,}75 + 0{,}5\left[\frac{\frac{60}{2} - 19}{17}\right] = 11{,}07352$$

$$\tilde{X} \Longrightarrow \frac{\Delta Y_1}{\Delta Y_2} = \frac{\Delta X_1}{\Delta X_2} \Longrightarrow \frac{\frac{n}{2} - fcd_{inf}}{fcd_i - fcd_{inf}} \frac{\tilde{X} - X_{inf}}{a}$$

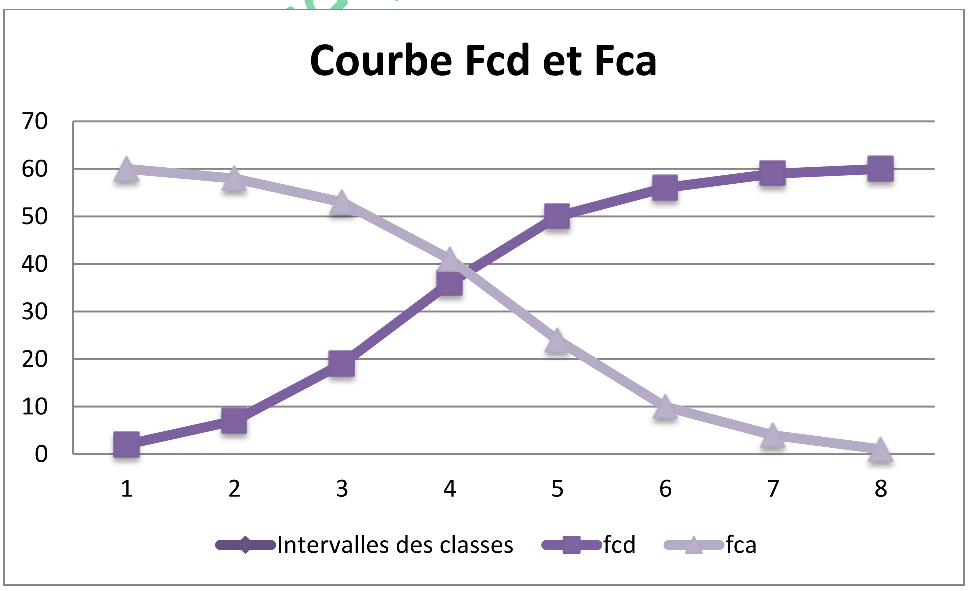
$$\frac{\frac{60}{2} - 19}{36 - 19} = = \frac{\tilde{X} - 10{,}75}{11{,}25 - 10{,}75} \Longrightarrow 11 \times 0{,}5 = 17\tilde{X} + 10{,}75 \times 17 \Longrightarrow \tilde{X} = \frac{188{,}25}{17} = 11{,}07352.$$

Résolution par la méthode graphique



Si les mesures sont bonnes la projection du point de rencontre de $\frac{n}{2}$ avec la courbe sur l'axe des abscisses doit correspondre à 11,07352 cherché.

De même en traçant la courbe de fcd et celle de fca, l'intersection donne la médiane.



3° Moyennes

a) Moyenne arithmétique $\bar{x}$: c'est la valeur qui est couramment utilisée en statistique pour la valeur centrale d'un ensemble.

$$\bar{X} = \sum_{i=1}^{n} n_i X_i$$

X_i : est le centre de classe

n_i : effectif correspondant au centre de classe

Σ : somme des effectifs.

N.B.

– Si le nombre d'observations est très élevé, il faut regrouper les données en classe ou changer d'origine. L'origine 0 devient 0' d'où $L_i = X_i\text{-}X_0$.

$\frac{\sum L_i}{n} = \frac{\sum X_i}{n} = X_0, \bar{e} = \bar{X} - X_0$ avec $\bar{X} = X_0 + \bar{e}$

- Si la distribution en classe a des intervalles égaux de même amplitude, on peut prendre comme déviation.

$$L_i = \frac{\bar{X} - X_0}{a} \Longrightarrow L_i.a = X_i - X_0$$

$$a\sum L_i = \sum X_i - nX_0, \quad a\bar{e} = X_i - X_0 \Longrightarrow \bar{X} = X_0 + a\bar{e}$$

En prenant la déviation, l'avantage est que la moyenne remplit la condition de YULE, c'est-à-dire le paramètre de position car la moyenne n'est pas une donnée mais une statistique. Sauf à la condition 5 une observation extrême (exceptionnellement élevée ou faible) peut avoir une forte incidence sur sa valeur.

Remarques : la somme de déviation par rapport à la moyenne est nulle.

$$\sum_{i=1}^{k} n_i(X_i - \bar{X}) \Longrightarrow \sum(\mathrm{X_i} - \bar{X}) = \sum \mathrm{X_i} = \mathrm{n}\bar{\mathrm{X}} = \frac{\sum \mathrm{X_i}}{\mathrm{n}} = \mathrm{n}\bar{\mathrm{X}}$$

$n\bar{X} - a\bar{X} = 0$, avec $\bar{X}$ qui est le cente de gravité ou le barycentre.

b) Moyenne arithmétique pondérée

C'est une moyenne arithmétique dont chaque valeur possède un poids appelé coefficient de pondération qui est un nombre positif non nul donnant la valeur de l'observation.

- La somme algébrique des écarts des valeurs d'une variable statistique à sa moyenne arithmétique est nulle.

$$\sum_{i=1}^{k} n_i(X_i - \bar{X}) = 0$$

- Lorsqu'on fait subir à une variable statistique X une transformation affine, c'est-à-dire un changement d'origine et d'unité $Y = ax + x_0$, sa moyenne arithmétique subit la même transformation $\bar{Y} = a\bar{X} + X_0$.

La moyenne présente un avantage, celui d'obtenir le paramètre compréhensif et de rendre le calcul très simple. Toutefois, elle présente un inconvénient celui d'avoir un paramètre insuffisant qui nécessite d'être complété par d'autre car dans les distributions à faibles effectifs, elle st très influencées par les valeurs les plus grandes.

Si les valeurs se retrouvent dans un tableau d'effectifs, la moyenne arithmétique sera calculée par la formule suivante :

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} n_i X_i$$

n_i : effectif correspondant au centre de classe

X_i = le centre de classe d'ordre i

N = la somme des effectifs.

Exemples : 1. Calculer la moyenne arithmétique pour la distribution ci-dessous :

X_i	n_i	$n_i x_i$
18	2	36
21	9	189
24	11	264
27	13	351
30	9	270
33	4	132
36	2	72
	50	**1.314**

Solution

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} n_i X_i = \frac{1}{50} \times 1.314 = 26{,}28.$$

2. Calculer le salaire moyen lorsque dans une entreprise la distribution des salaires se présente comme suit :

Limites des classes	X_i	n_i	$n_i x_i$
50 – 59,99	54,995	8	439,960
60 – 79,99	69,995	26	1.819,870
80 – 89,99	84,995	15	1.274,925
90 – 109,99	99,995	15	1.499,925
100 – 149,99	129,995	4	519,980
150 – 179,99	164,995	2	329,99
		70	**5.884,650**

$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} n_i X_i = \frac{1}{70} \times 5.884,650 = 84,0664$. Comme salaire moyen.

$\overline{X_p} = \frac{\sum P_i . X_i}{\sum P_i}$, p_i étant le poids.

c) Moyenne harmonique (H)

Elle est utilisée souvent pour évaluer la vitesse moyenne, elle est la moyenne des inverses des éléments d'un ensemble ou nombre.

En cas de ; - distribution non groupée, elle se calcule partant de $\frac{1}{H} + \frac{1}{n}\sum \frac{1}{X_i} \Longrightarrow H = \frac{1}{\frac{1}{n}}\sum \frac{1}{X_i}$

Données groupées

$H = \frac{1}{\frac{1}{N}}\sum \frac{n_i}{X_i}$ avec n_i : fréquence absolue et X_i : centre des classes.

d) Moyenne géométrique (G)

Elle est la racine carrée des produits des x_i.

$$G = \sqrt[n]{\frac{\frac{n}{\pi}}{i = 1} . X_i}$$

$$log\ G = \frac{1}{n}\sum_{i=1}^{n} logX_i \Longrightarrow G = \sqrt[n]{\frac{\frac{n}{\pi} . X_i . n_i}{i = 1}}$$

Elle est une moyenne pondérée avec $X_0 > 0$.

e) Moyenne quadratique (Q)

Elle est la racine carrée de la moyenne arithmétique des carrés des nombres.

$Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$ pour les données groupées.

N.B. : <u>H < G < $\bar{X}$ <Q</u>

f) Médiale (Me)

La médiale est la valeur de la variable aléatoire qui partage la masse des $n_i x_i$ cumulées croissantes.

$$M_e = L_i + a . \frac{\frac{\sum n_i X_i - fc_{ni} . X_{is}}{2}}{fc_{nime}}$$

g) Les centiles ou percentiles

Sont des valeurs de la variable aléatoire qui fournissent les informations sur la manière dont les observations sont réparties dans l'intervalle entre la plus petite et la plus grande valeur des observations. Calcul de percentiles.

a) *Cas des séries groupées en classes*

- $i = \frac{P}{100}.n$
- $C_p = L_{inf} + a\left(\frac{i-fcd_{inf}}{n_{icent}}\right)$

b) *Cas des séries non groupées*

- Classer les éléments par ordre croissant
- Calculer l'indice $i = \frac{P}{100}.n$
- Si i n'est pas un nombre entier, l'arrondir à l'excès pour la position de P qui correspondra à l'entier supérieur à i.
- Si i est un nombre entier, la position du $P^{ème}$ est la moyenne de i et i+1 (i correspondant aux valeurs des observations).

Exemple 1. Déterminer le 85ème et le 50ème percentile de la série ci-après : 2255, 2210, 2825, 2380, 2440, 2390, 2550, 2350, 2380, 2420, 2450, 2630.

<u>Solution</u>

Sériation : 2210,2255, 2350, 2380, 2390, 2420, 2440, 2450, 2550, **2630**, 2825.

n = 12 et P = 85.

$i = \frac{P}{100}.n = \frac{85}{100}.12 = 10{,}2 \Rightarrow i \approx 11$

Donc, la position du 85ème percentile correspond au nombre entier supérieur à 10,2 soit la 11ème position. D'où le 85ème percentile est égal à 2630.

2. Le 50ème percentile est : $i = \frac{50}{100} \times 12 = 6$. Comme i est un entier, la position p^{ème} sera égale à la moyenne de 6ème et 7ème observations : $\frac{2420+2440}{2} = 2430$

Le 50ème percentile correspond toujours à la médiane de la série statistique.

Exemple 2. Soit la distribution reprenant les salaires des ouvriers attribués dans un secteur précis. Calculer le 92ème centile.

Intervalles de classes	n_i	fcd
[0 – 100[	1300	1300
[100 – 125[	608	1908
[125 – 150[	998	2906

$[150 - 175[$	1508	4414
$[175 - 200[$	2496	6910
$[200 - 225[$	3050	9960
$[225 - 250[$	4580	14540
$[250 - 300[$	10200	24740
$[300 - 350[$	16600	41340
$[350 - 400[$	11984	53324
$[400 - 500[$	13632	66956
$[500 - 600[$	10200	77156
$[600 - 800[$	14320	91476
$[800 - 1000[$	8520	99996

<u>Solution</u>

$$C_p = L_{inf} + a\left(\frac{i - fcd_{inf}}{n_{icent}}\right)$$

$$i = \frac{99.996}{100} \times 92 = 91.996{,}32$$

- L'amplitude a doit être celle la plus grande c'est-à-dire 200.
- L'intervalle correspond est $[800 - 1000[$, Limite inférieure de cet intervalle est Linf = 800.
- fcd étant 99.996, fcd_{inf} est 91.476
- fcd de 99.996 correspond à n_{icent} de 8.520
- $c_{92} = 800 + 200\left(\frac{91.996{,}32+91476}{8520}\right) = 812{,}214$

Les quartiles notés aussi quantiles sont des percentiles particuliers. Les étapes de calcul des percentiles peuvent être directement appliquées au calcul de quartiles.

Les quartiles divisent la distribution en trois parties, ces nombres sont Q_1, Q_2 et Q_3.

- ❖ Q_1 dit quantile inférieur est la valeur de la variable aléatoire, laissant à sa gauche 25% des valeurs inférieures au quantile des observations et dont 75% lui sont supérieures. (1/4)
- ❖ Q_2 dit quantile central est la valeur de la variable aléatoire, supérieure à 50% d'observations de la série et inférieure au 50% restant. Q_2 correspond à la médiane (1/2).
- ❖ Q_3 dit quantile supérieur est la valeur de la variable aléatoire supérieure à 75% des observations et inférieures à 25%. (3/4)

Exemple

Intervalles des classes	n_i	fcd

$[800 - 900[$	4	4
$[900 - 1000[$	20	24
$[1000 - 1100[$	107	131
$[1100 - 1200[$	168	299
$[1200 - 1300[$	122	421
$[1300 - 1400[$	48	469
$[1400 - 1500[$	21	490
$[1500 - 1600[$	10	500

N =500

i de $Q_1 \ = \frac{n}{4} = \frac{250}{2} = 125$

i de $Q_2 \ = \frac{n}{2} = \frac{500}{2} = 250$

i de $Q_1 \ = \frac{3n}{4} = \frac{1500}{4} = 375$

$n_i > 168$ qui correspond à l'intervalle $[1100 - 1200[$

$f_i = 1100$

$f_i <= 1000$

$f_i > = 1200$

$$Q_1 = L_i + a\left[\frac{\frac{n}{4} - fcd_i}{n_i Q_1}\right] = 1000 + 100\left[\frac{125 - 24}{107}\right] = 1.094{,}39$$

$$Q_3 = L_i + a\left[\frac{\frac{n}{2} - fcd_i}{n_i Q_2}\right] = 1100 + 100\left[\frac{250 - 131}{168}\right] = 1.170{,}83$$

$$Q_1 = L_i + a\left[\frac{\frac{3n}{4} - fcd_i}{n_i Q_3}\right] = 1200 + 100\left[\frac{375 - 299}{122}\right] = 1.262{,}29$$

N.B. : fcd_i= 24 pour Q_1, fcd_i = 131 pour Q_2 et fcd_i = 299 pour Q_3.

N.B :

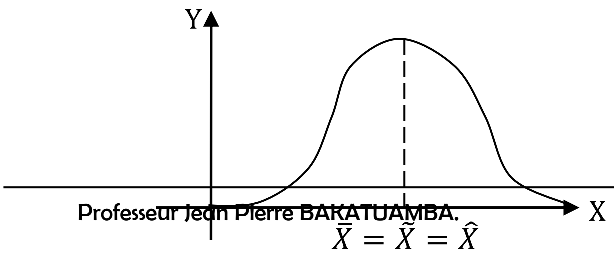
- En cas de quartiles ou déciles, si Q_1 = 25% alors Q_2 = 50% et Q_3 = 75%.

- s'il faut établir la relation entre moyenne, médiane et mode dans une distribution normale.

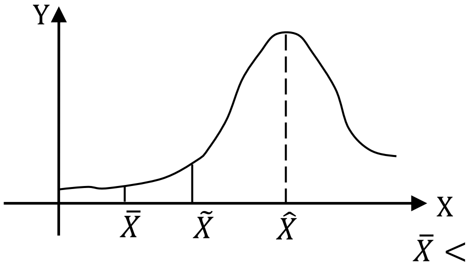
$\bar{X} = \tilde{X} = \hat{X} \Longrightarrow$ Moyenne = Médiane = Mode.

Cette relation est dite empirique parce qu'utilisée de moins à moins aujourd'hui.

Graphiquement la situation doit être la suivante :

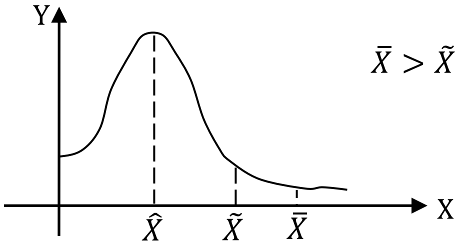


Si elles t étalée à gauche et biaisée à droite, elle est dite dissymétrique



$\bar{X} < \tilde{X} < \hat{X}$

Idem lorsqu'elle est étalée à droite et biaisée à gauche.



$\bar{X} > \tilde{X} > \hat{X}$

PEARSON admet que $\bar{X} = \hat{X} = 3(\bar{X} - \tilde{X})$