Université Mohamed Boudiaf - M'sila

Faculté des Sciences et Technologies

Départements de Hydraulique

Licence (2/HYDROLIQUE)

Année Universitaire 2020/2021

Module : (Probabilités et Statistiques)

U.E Methodologique; Crédits: 04; Coefficient: 02

Contrôle continue 40%; Examen: 60%

Partie A: Statistiques

Chapitre 1

Définitions de base-Séries statistiques à une variable

1.1 Généralités :

Le but de statistique descriptive est de structurer et de représenter l'information contenue dans les données, et analyse et synthèse numérique et graphique d'un ensemble de données.

Individus ou unités statistiques:

Chacune des "personnes" étudieés, ex: personne humaine, pays, notes.

Population :

Ensemble des individus observés; Ex: les étudiants de 12-25 ans.

La Taille de la population :

Est le nombre d'individus.

L'échantillon :

Est un sous ensemble de la pupulation considérée.

Exemple 1.1.1 *Population : les étudiants de* $2^{\text{ème}}$

unités statistiques: étudiant

L'échantillon : groupe de TD

Caractère (variable statistique):

Est la propriété étudiée sur ces sujets;

Exemple 1.1.2 *Variables statistiques: sexe, taille, nombre d'enfants...*

Les modalités:

Sont les différentes valeurs qu'une variable peut prendre.

Exemple 1.1.3 *Variable statistique: la note de TD; les modalités* $\in [0; 20]$.

Variable statistique: sex; les modalités: masculin et féminin:

Types de Caractères:

1-Variables qualitatives: Une variabe statistique est dit qualitative lorsque ses modalités ne sont pas mesurables. Elle peut être de type:

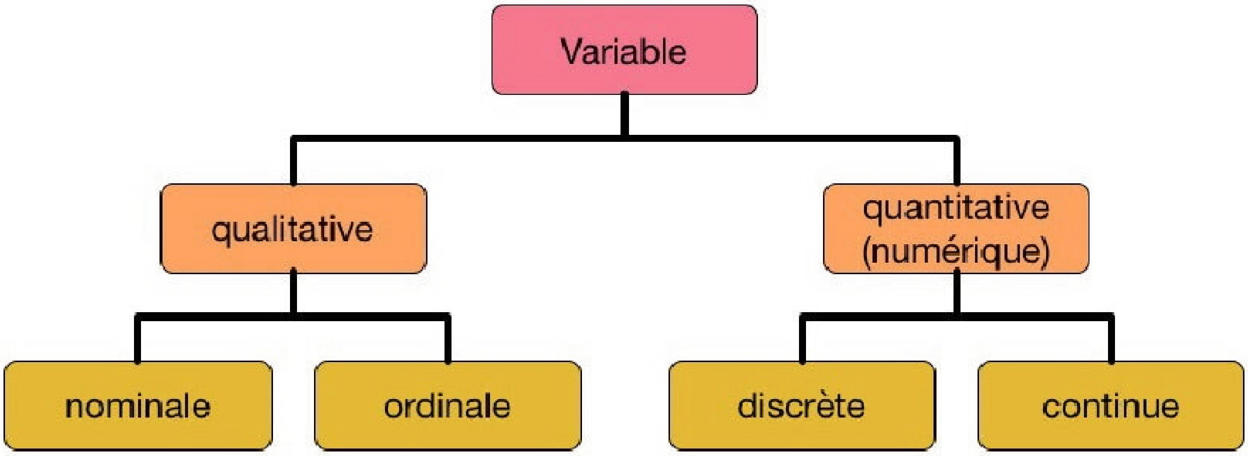
a- Nominale: les groupes ne sont pas ordonnés, par exemple: couleurs des yeux.

b- Ordinale: les groupes sont ordonnés, par exemple: mention au bac, (passable, assez-bien, bien)

2-Variables quantitatives: Une variabe statistique est dit quantitative lorsque ses modalités sont mesurables. Elle peut être de type:

a- Discrète: Les modalités sont dénombrables, par exemple: nombre d'enfants par famille.

b- Continue: Les modalités sont définies sur définies sur un intervalle continu, par exemple: la taille, le poids.



Exercice 1.1.1 *Quel est le type de variation statistique dans les cas suivants?*

La variable statistique (couleur de maisons d'un quartier) est-elle: qualitative.

La variable statistique (les notes d'étudiants) est-elle: quantitative discrète.

La variable statistique (les tailles d'étudiants) est-elle: quantitative continue.

1.2 Séries statistiques à une variable

1.2.1 Comment organiser les données:

On regroupe toutes les données de la série statistique dans un tableau indiquant la répartition des individus selon le caractère étudié. Le regroupement s'effectue par classes :

• Si le caractère est qualitatif ou discontinu, une classe contient tous les individus ayant la même modalité ou la même valeur du caractère.

• Si le caractère est continu, une classe est un intervalle.

■ Pour construire ces intervalles, on respecte les règles suivantes :

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

2. Chaque fois que cela est possible, les amplitudes des classes sont égales.

3. Chaque classe (sauf la dernière) contient sa borne inférieure mais pas sa borne supérieure.

■ Dans les calculs, une classe sera représentée par son centre, qui est le milieu de l'intervalle.

■ Que faut-il indiquer pour chaque classe ?

1. **L'effectif** : nombre d'individus de la classe : on le note n_i (i est l'indice de la classe).

2. **La fréquence** : proportion d'individus de la population ou de l'échantillon appartenant à la classe : on la note f_i.

f_i et n_i ont liés par :

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

où N est le nombre total d'individus dans la population.

Remarque : On peut remplacer f_i par $f_i \times 100$ qui représente alors un pourcentage.

On a toujours :

$$\sum_{i=1}^{k} n_i = N, \ \sum_{i=1}^{k} f_i = 1, \ 0 \leq f_i \leq 1$$

où k représente le nombre de classes.

3. **L'effectif (ou la fréquence) cumulé (e) :**

Etant donné une série statistique.

On appelle effectifs cumulés croissants associés à une valeur la somme des effectifs des valeurs inférieures.

On appelle effectifs cumulés décroissants associés à une valeur la somme des effectifs des valeurs supérieures.

Les fréquences cumulées croissantes et décroissantes sont les proportions de chacun de ces effectifs par rapport à l'effectif total.

Exemple 1.2.1

Valeur	*0*	*1*	*2*	*3*	*4*	*5*
Effectif	*3*	*3*	*12*	*6*	*3*	*3*
Fréquence	*0.1*	*0.1*	*0.4*	*0.2*	*0.1*	*0.1*
Effectif cumulé croiss	*3*	*6*	*18*	*24*	*27*	*30*
Effectif cumulé décroiss	*30*	*27*	*24*	*12*	*6*	*3*
Fréquence cumulé croiss	*0.1*	*0.2*	*0.6*	*0.8*	*0.9*	*1*
Fréquence cumulé décroiss	*1*	*0.9*	*0.8*	*0.4*	*0.2*	*0.1*

1.2.2 Représentation Graphiques:

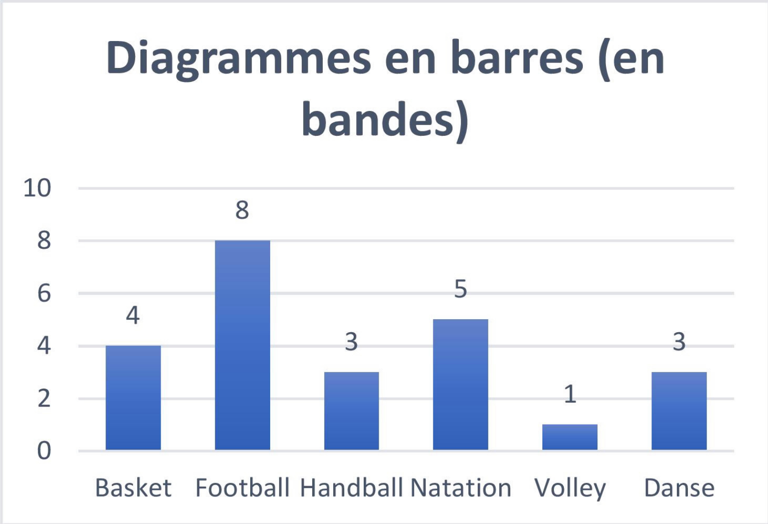
A- Pour une variable statistique qualitative:

Exemple 1.2.2 *On a demandé à 24 élèves d'une même classe de collège quel était leur préféré, les réponses sont présentées dans le tableau:*

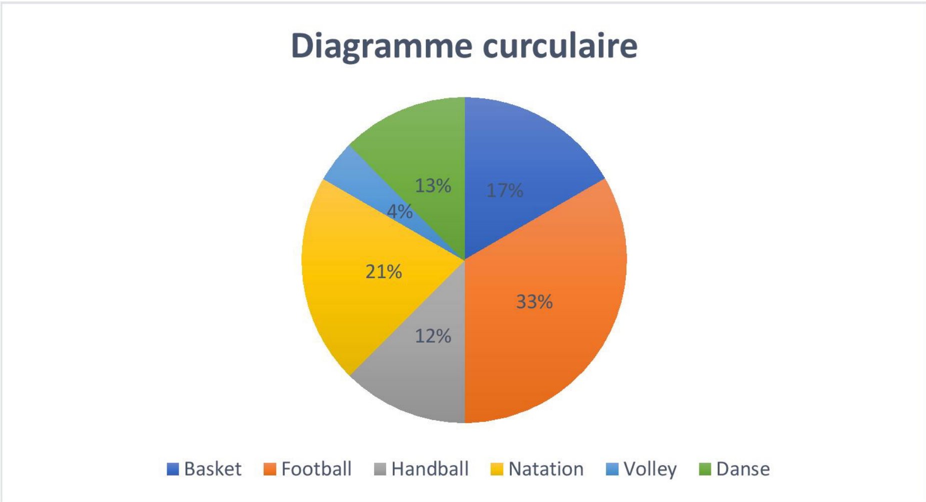
$N = 24$

Sport	*Basket*	*Football*	*Handball*	*Natation*	*Volley*	*Danse*
Effectifs	4	8	3	5	1	3
Angles(en°)	60°	120°	45°	75°	15°	45°

1- Diagrammes en barres (en bandes)



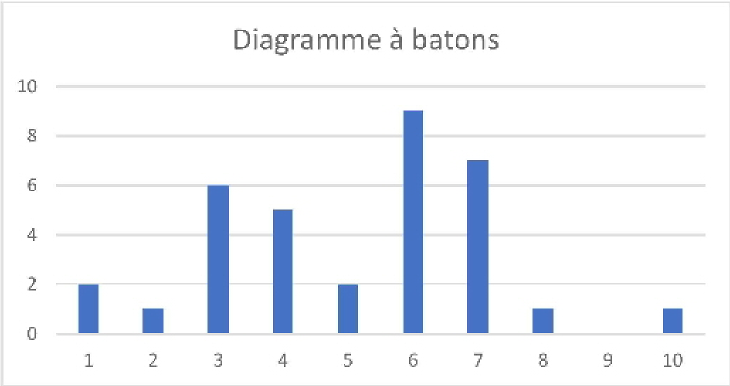
2- Diagrammes circulaire ou semi-circulaire



B- Pour une variable statistique discrète:

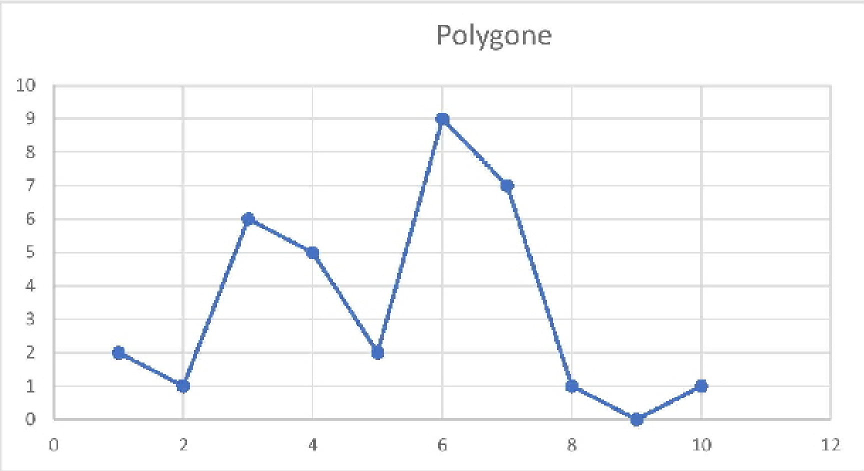
1- Diagrammes à batons

Note	1	2	3	4	5	6	7	8	9	10
n_i	2	1	6	5	2	9	7	1	0	1



On remplace parfois l'effectif par la fréquence, ce qui donne bien sure le même aspect au diagramme.

2-Polygône des effectifs (et des fréquences) Le polygône des effectifs (et des fréquences) est obtenue en reliant les extrémités des batons du diagramme précédent;



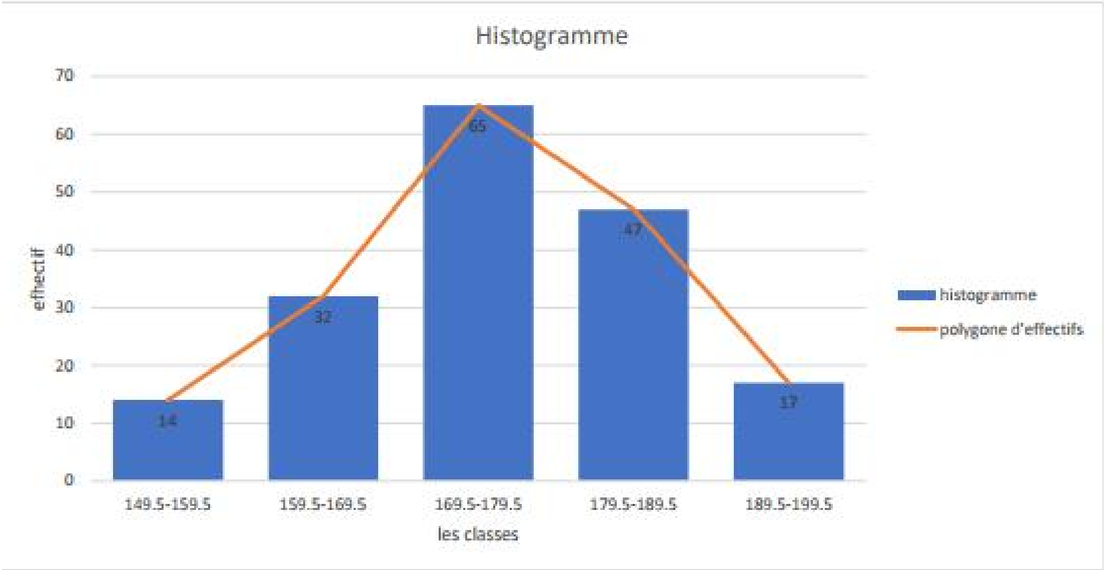
C- Pour une variable statistique continue:

1-Histogramme:

Exemple 1.2.3 *Reprenons la distribution groupée des tailles des 175 étudiants:*

Classe de taille	$[149.5; 159.5[$	$[159.5; 169.5[$	$[169.5; 179.5[$	$[179.5; 189.5[$	$[189.5; 199.5[$
Eff n_i	*14*	*32*	*65*	*47*	*17*
Eff cumulé	*14*	*49*	*111*	*158*	*175*
Fréq f_i	*0.08*	*0.18*	*0.37*	*0.27*	*0.10*
F_i	*0.08*	*0.26*	*0.63*	*0.90*	*1*

L'histrogramme correspondant est donc:



2-Les courbes cumulatives (polygone des fréquences cumulées) Est obtenue en portant les points dont les abscisses représentent la borne supérieure de chaque classe et ordonnées les fréquences cumulées correspondantes, puis en reliant ces points par des segments de droite son équivalent dans la Théorie probabiliste est la fonction de répartition.

1.2.3 Caractéristiques de position

A-Pour une variable statistique discrète:

Exemple 1.2.4 *Notes à l'examen de statistique:* $N = 14$

Note x_i	*2*	*4*	*6*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*
Eff n_i	*1*	*1*	*1*	*1*	*1*	*1*	*2*	*2*	*1*	*1*	*1*	*1*

1-Le mode: Mo : la valeur la plus fréquente, $Mo = 11, 12$.

2-La médiane: Med : qui partage la série statistique en deux groupe de meme fréquence:

-Si $N = 2n$ paire: $Med = \frac{a_n + a_{n+1}}{2}$

-Si $N = 2n + 1$ impaire: $Med = a_{n+1}$

Donc $\frac{N}{2} = 1$,et $Med = 11$

3- La moyenne:

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \sum_{i=1}^{k} f_i x_i, \overline{x} = 10.21$$

B- Pour une variable statistique continue:

Exemple 1.2.5 *Les notes dans les classes:*

Les classes	$[0;4[$	$[4;8[$	$[8;12[$	$[12;16[$	$[16;20]$
eff n_i	*1*	*2*	*5*	*5*	*1*
Les centres c_i	*2*	*6*	*10*	*14*	*18*
f_i	*0.07*	*0.14*	*0.36*	*0.36*	*0.07*
F_i	*0.07*	*0.21*	*0.57*	*0.93*	*1*

1-Le mode: Est $[8, 12[$ et $[12, 16[$.

2- La classe médianne: On a $\frac{N}{2} = 7$; Classe.Med=$[8, 12[$.

Med : c'est la valeur correspondant à un effectif cumulé de 50% sur le polygône des fréquences cumulées.

3-La moyenne $\overline{x}$:

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i c_i}{N} = \sum_{i=1}^{k} f_i c_i, \overline{x} = 10.88$$

1.2.4 Caractréristiques de dispersion

1-Etendue:

l'étendue d'un série statistique est égale à la différence entre la plus grande et la plus petite valeur de la série.

$$e = X_{\max} - X_{\min}$$

Exemple 1.2.6 *24,16,18,22,16,26,35,25,15,76.*

$$e = 76 - 16 = 50 \ ans$$

2-La variance:

$V(x)$ est la moyenne des carrées des des écarts des valeurs x_i à la moyenne $\overline{x}$; c'est à dire:

$$V(x) = \frac{1}{N} \sum_{i=1}^{p} n_i \left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{p} f_i \left(x_i - \overline{x}\right)^2$$

3-Ecart-type:

$$\sigma_x = \sqrt{V\left(x\right)}$$

Remarque 1.2.1
- *L'écart type est un paramètre plus fin que l'etendue, car il tient compte de la répartition des valeur.*
- *L'écart type à la meme unité que les valeurs de la série étudiée.*
- *L'écart type mesure la dispersion des valeurs de la série autour de la moyenne. Plus les valeur de la série sont concentrées autour de la moyenne.*
- *Plus σ_x est grand et plus la série observée est dispersée.*

Autre (bien plus pratique)

$$V\left(x\right) = \frac{1}{N} \sum_{i=1}^{p} n_i x_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2$$

Preuve. (Exercice) ■

Remarque 1.2.2 *Tout comme la moyenne, pour calculer une variance* (*ou* σ_x) *pour une variable continue on remplace les x_i par c_i les centre de classe:*

$$V\left(x\right) = \frac{1}{N} \sum_{i=1}^{p} n_i c_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2$$

Exercice 1.2.1 *Calculez les variance et écart-type de la série suivante:*

x_i (*en* €)	$[0, 1600[$	$[1600, 2400[$	$[2400, 3200[$
c_i	*800*	*2000*	*2800*
n_i	*9*	*7*	*4*

On rappelle que $\overline{x} = 1620€, N = 20$

Méthode 01:

$$V(X) = \frac{1}{20} \sum n_i (x_i - 1620)^2 = 631600 \text{ €}^2$$

Méthode 02:

$$\begin{aligned} \overline{x^2} &= \frac{1}{20}\left(9 \times 800^2 + 7 \times 2000^2 + ...\right) = 3256000 \text{ €}^2 \\ V(X) &= \overline{x^2} - \overline{x}^2 = 3256000 - 1620^2 = 631600 \text{ €}^2 \end{aligned}$$

Ecart-type:

$$\sigma_x = \sqrt{631600} = 794.7€$$

1.2.5 Coefficient de variation

-Pour comparer la dispersion de deux séries qui ne sont pas exprimées dans les mêmes unités, on utilise le coefficient de variation

$$Cv = \frac{\sigma_x}{\overline{x}} = \text{c'est le \% de variation par rapport à la moyenne, sant unité}$$

-Plus grand est le coefficient de variation; plus grand est la dispersion.

Exemple 1.2.7 (comparaison de série) *1- Soit x la série statistique de 4 produits en Francs:100F, 200F,300F,400F.*

2-Soit y la série statistique de 4 produits en €: 15€,30€,45€,60€.

Série	N	$\overline{x}$	$V(X)$	σ_x	Cv
1	4	$250F$	$12500F^2$	111.80F	0.45
2	4	$37,5$€	218.25€	16.77€	0.45

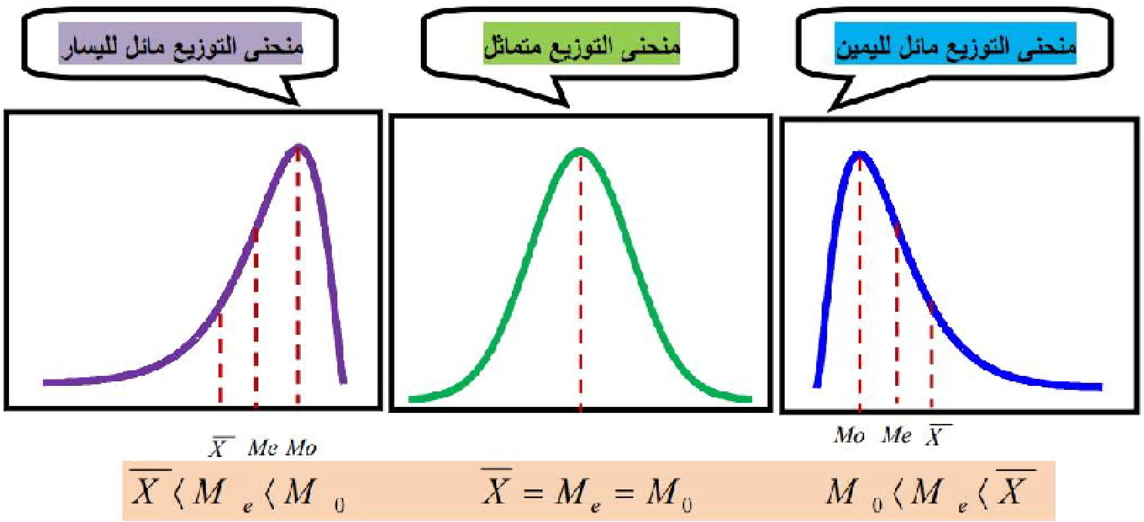
On a $Cv_1 = Cv_2$, les deux série de même dispersion.

1.2.6 Caractristiques de forme:

Les statisticiens ont explicité deux critères pour décrire la forme d'une telle courbe: **sa symétrie et son aplatissement.**

1- Symètrie d'une distribution

- Si la courbe d'une distribution de fréquence est symètrique. alors les valeurs sont réparties dans les mêmes proportions autour du mode; et donc $\overline{x} = Mo = Me$.

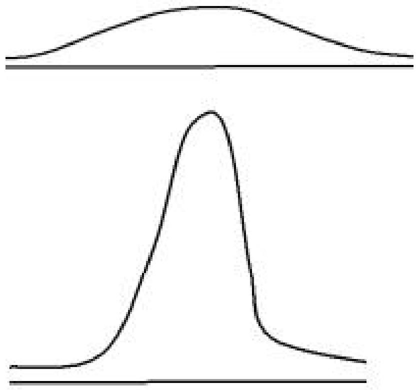


2- Aplatissement d'une distribution

Dans le cas d'une variable unimodale,

- Plus les valeurs sont dispersées, plus la courbe de distribution apparaît aplatie.

$\Downarrow\Downarrow\Downarrow$



⇑⇑⇑

- Plus les valeurs sont concentrées autor du mode, moins la courbe de distribution apparaît aplatie.

Le responsable de la matière : Samiha aichouche.