Université A- mira de Bejaia
Faculté des sciences économiques, commerciales et des sciences de gestion
Département SEGC(LMD)
Module stat I
Enseignant : Dr. Mousli

Chapitre 2 : Représentation des données statistiques

2.1. Présentation des données

La présentation des donnés d'une série statistique obtenue lors de l'étude d'un caractère quelconque peut être donnée soit sous forme d'un tableau ou sous forme d'un graphique**.**

2.1.1. Présentation des données sous forme de tableaux :

Un tableau statistique est un tableau constitué initialement de deux colonnes :

- La première colonne correspond aux classes, aux valeurs, aux modalités selon leurs caractères.
- La deuxième colonne correspond aux effectifs (n_i) des classes ou des valeurs, si non elle correspond aux fréquences (f_i)
- Le regroupement des informations chiffrées dans un tableau (ou graphe) exige les éléments suivants : Le titre du tableau, l'unité de mesure et la source des informations recueillies.

Il faut rappeler que le caractère peut être de type qualitatif ou quantitatif.

A. Présentation des données d'un caractère qualitatif

Pour présenter les résultats observés, nous distinguons d'abord les différentes modalités du caractère étudié (les différentes valeurs de la variable considérée). On les désigne par $x_1, x_2, x_3, \dots, x_k$. Il s'agit donc d'un caractère présentant k modalités. Nous déterminons ensuite le nombre d'individus associés à chacune des modalités. On parlera alors d'effectifs absolus(n_i). Ils sont identifiés par $n_1, n_2, n_3, \dots, n_k$.

Nous pouvons aussi évaluer l'effectif relatif de chacune des modalités. On parlera alors de fréquences relatives. Ces dernières sont représentées par $f_1, f_2, \dots, f_k$et obtenues grâce au rapport suivant :

$f_i = \frac{n_i}{N}$ Avec Neffectif total $\left(N = \sum_{i=1}^{k} n_i = n_1 + n_2 + n_3 + \dots + n_k\right)$

Nous pouvons identiquement exprimer l'effectif relatif en pourcentage. Il suffit de multiplier l'effectif relatif de chacune des valeurs de la variable par **100.**

Exemple : une courte enquête auprès de 10 étudiants pour connaître leur nationalité permet d'obtenir la série statistique suivante :

Algérien, étranger, algérien, algérien, étranger, algérien, algérien, étranger, algérien, étranger,

Pour présenter les résultats observés, nous distinguons d'abord les différentes modalités du caractère considéré. Le caractère étudié ici est la nationalité représentée par $\boldsymbol{X_i}$. Nous déterminons ensuite le nombre d'individus associés à chacune des valeurs de cette variable. Ils sont identifiés par $\boldsymbol{n_i}$. Enfin, un tableau est construit pour résumer toutes ces informations.

Tableau 1 : La répartition des 10 étudiants selon leur nationalité

Nationalité X_i	**Effectifs Absolus** $\boldsymbol{n_i}$	**Effectifs Relatifs** $\boldsymbol{f_i = n_i/N}$	**Effectifs Relatifs** $\boldsymbol{f_i}(\%)$
Algérien	6	6/10=0,6	60
Étranger	4	4/10=0,4	40
Total	10	1,00	100

Source : enquête personnelle

B. Présentation des données d'un caractère quantitatif

En raison de la nature des valeurs que peut prendre un caractère quantitatif dans une série statistique, nous distinguons 2 catégories : le caractère quantitatif discret et le caractère quantitatif continu.

B1. Caractère quantitatif discontinu

Un caractère quantitatif est discontinu si l'ensemble des valeurs qu'il peut prendre est fini ou dénombrable. Lorsque ces valeurs sont entières, on dira que le caractère est discret.

Exemple 1 : Les données portant sur le nombre de filles dans une famille ou le nombre de voitures par foyer sont des caractères discrets.

Exemple 2 : la distribution de fréquences suivante représente la répartition de 100 familles selon le nombre de garçons ($0 \leq X_i \leq 5$).

Tableau 2 : La répartition de 100 familles selon le nombre de garçons

Le nombre de garçons $\boldsymbol{X_i}$	$\boldsymbol{n_i}$	$f_i \left(\frac{n_i}{N}\right)$
0	15	15/100=0,15
1	25	25/100=0,25
2	35	35/100=0,35
3	10	10/100=0,10
4	10	10/100=0,10
5	05	05/100=0,05
Total	100	1,00

Exemple 3 :

La série suivante résulte d'une enquête auprès de 30 étudiants pour connaître leur âge :

20 18 21 19 19 18 18 22 20 19 17 18 23 19 21

18 18 19 17 18 19 22 21 21 20 18 17 23 20 22

Tableau 3 : La répartition d'un groupe de 30 étudiants selon leur âge (en années)

Age (en années)	n_i	f_i	f_i (%)
17	3	0,1	10
18	8	**0,26**	**27**
19	6	0,2	20
20	4	0,13	13
21	4	0,13	13
22	3	0,1	10
23	2	**0,06**	**7**
Total	30	0,98≈ 1	100

Remarque :

Notons que la somme des effectifs relatifs ne donne pas un total de 1 (ou de 100 en %). Cela s'explique par les résultats des opérations effectuées lors du calcul de l'effectif relatif. Dans la pratique et comme dans le tableau précédent, il suffit de modifier la valeur de l'effectif relatif d'une valeur ou de plusieurs valeurs en ne perdant pas de vue l'objectif d'obtenir un total de 100.

Reprenons les effectifs relatifs concernant le tableau précédent. Nous obtiendrons alors :

10% 27% 20% 13% 13% 10% 7%

La somme de ces fréquences relatives est bien égale à 100 (voir la colonne 4).

B2. Caractère quantitatif continu

Un caractère quantitatif est continu s'il peut prendre n'importe quelle valeur dans intervalle donné de nombres réels.

Par convention, l'intervalle est fermé à gauche et ouvert à droite.

Désignons par e l'extrémité de classe. La ième classe est alors notée « e_{i-1}, e_i » correspondant à l'intervalle [e_{i-1} $-$ e_i [.

L'amplitude a_i d'une classe est la largeur de l'intervalle de cette classe. La ième classe aura comme amplitude : $a_i = e_i - e_{i-1}$.

Exemple : l'amplitude de la classe [2 – 6 [est égale à : $a_i = 6 - 2 = 4$

Les amplitudes des classes peuvent être égales ou inégale sur l'intervalle total de variation de la variable statistique. Il peut également arriver que les classes extrêmes ne soient délimitées.

Par convention, on prend comme amplitude de la classe non délimitée, celle de la classe adjacente.

Comme la variable statistique ne prend plus une valeur préciseX_i, mais un ensemble de valeurs possibles comprises dans un intervalle appelé classe, nous sommes tenus de calculer le centre de classe, noté C_i qui est la valeur de la variable statistique égale à la moyenne arithmétique des valeurs des extrémités de la classe.

D'où :$c_i = \frac{e_{i-1}+e_i}{2}$

Exemple :

Le centre de classe de : [2 – 6 [est égale à : $c_i = \frac{2+6}{2} = 4$

Exemple1 :

Le système de pointage dans une entreprise publique a permis d'enregistrer les temps de retard (exprimés en minutes) de la part des travailleurs :

Tableau 4 : Répartition des 140 travailleurs selon leur temps de retard

e_i	n_i	Centre de classe C_i	Fréquences relatives f_i%
[04-08[	5	(04+08)/2= 6	3,57
[08-12[	20	(08+12)/2=10	14,29
[12-16[	15	(12+16)/2=14	10,71
[16-20[	60	(16+20)/2=18	42,86
[20-24[	25	(20+24)/2=22	17,86
[24-28[	15	(24+28)/2=26	10,71
Total	140	-	100,00

Exemple2 :

Soit le caractère « durée de vie d'un certain type d'appareils électriques ». Les modalités retenues sont résumées dans le tableau ci-dessous :

Tableau 5 : Tableau de répartition

e_i	Amplitude (a_i)	Intervalle utilisé	Centre de classe C_i
- 5	**5**	[00-05[	2,5
[05-10[	5	[05-10[	7,5
[10-20[	10	[10-20[	15
[20-25[	5	[20-25[	22,5
[25-30[	5	[25-30[	27,5
[30-50[	20	[30-50[	40
+ 50	**20**	[50-70[	60

Nous donnons à la classe « moins de 5 » une amplitude de 5 car la classe adjacente est la classe « 5-10 » d'amplitude égale à 5.

De même, la classe « plus de 50 » une amplitude de 20 car la classe adjacente est la classe « 30-50 » d'amplitude 20.

2.1.2. Représentation graphique des séries statistiques

Soit une population d'un effectif total N que nous voulons étudier selon un certain caractère K. Nous désignons par n_i l'effectif correspondant à chaque modalité du caractère. Comme nous l'avons vu antérieurement, il est possible de présenter les résultats obtenus sous forme de tableaux. Mais il est identiquement possible de recourir aux méthodes de représentation graphique. Celles-ci diffèrent selon le caractère qualitatif ou quantitatif de la variable étudiée.

A. Cas d'un caractère qualitatif

Dans ce cas, deux types de représentation graphique sont utilisés :

-Représentation cartésienne avec les **diagrammes en colonnes (tuyaux d'orgue)**

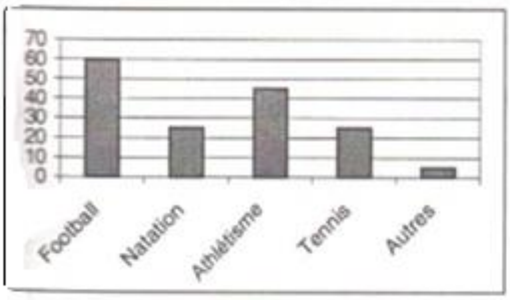
-Représentation avec le **diagrammes circulaire (représentation par secteurs)**

Exemple1 : Une enquête auprès de 160 étudiants de l'université de Bejaia au sujet de leur sport préféré a fourni les résultats suivants :

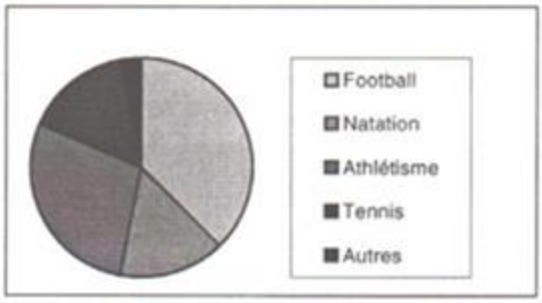
Tableau 6 : La répartition des 160 étudiants selon leur sport préféré

Sport préféré $\boldsymbol{X_i}$	Nombre de jeunes $\boldsymbol{n_i}$	f_i	$f_i \times 360^0$
Football	60	0,375	135
Natation	25	0,156	56.25
Athlétisme	45	0,281	101.25
Tennis	25	0,156	56.25
Autres	5	0,031	11.25
Total	160	1	360

Représentation en tuyaux d'orgue



Représentation par secteurs



B. Cas d'un caractère quantitatif

Le comptage des unités ou éléments observés se fait dans ce cas, suivant des valeurs d'une variable discontinue ou des intervalles de classes d'une variable continue.

B1. Cas d'un caractère quantitatif discret (discontinu) :

Une variable quantitative discontinue se définit comme une variable prenant un nombre limité. Elle est dite discrète si elle prend seulement des valeurs entières.

Exemple : la répartition d'un échantillon de N=100 familles de 5 enfants en fonction du nombre de filles ($0 \leq X_i \leq 5$).

Tableau 7 : Répartition des 100 familles selon le nombre de filles

Nombre de filles	Effectifs (n_i)	Fréquences relatives f_i (n_i/N)	Effectifs cumulés croissants $n_i \uparrow$	Effectifs cumulés décroissants $n_i \downarrow$
0	15	0.15	15	100
1	35	0.35	50	85
2	25	0.25	75	50
3	10	0.10	85	25
4	5	0.05	90	15
5	10	0.10	100	10
Total	100	1.00	-	-

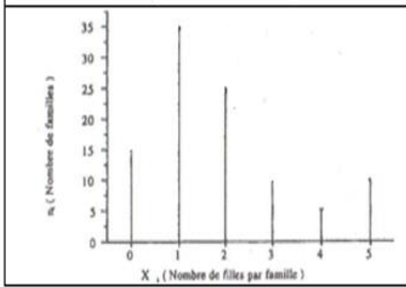
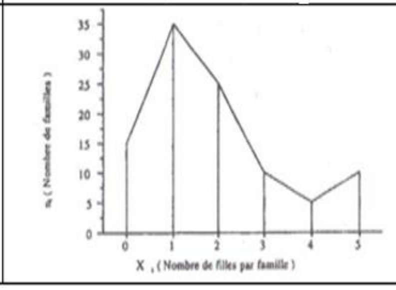
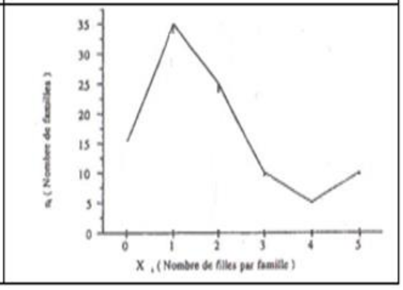
Signification :

Les effectifs cumulés indiquent le nombre d'individus pour lesquels la variable statistique est inférieure à X_{i+1}

Exemple :

15 représente le nombre de familles pour lesquelles il y a eu moins d'une fille, c'est-à-dire aucune fille. De la même façon, 50 indique le nombre de familles pour lesquelles il y a eu moins de 2 filles.

Dans le cas d'une série discontinue, il existe plusieurs façons de représenter graphiquement les résultats. Nous rappellerons ici les formes les plus utilisées :

• Diagramme en bâtons	• Polygone d'effectifs	• Courbe de fréquences
		

A chaque valeur de la variable X_i portée en abscisse, on fait correspondre un bâton de longueur proportionnelle à l'effectif n_i correspondant à cette valeur.

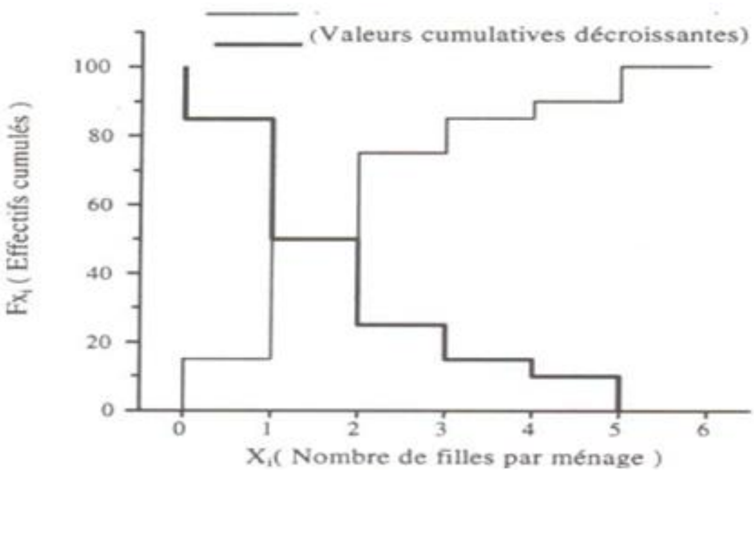
- **Courbe cumulative croissante et décroissante**

Nous pouvons également construire des diagrammes en **« escaliers »** représentant les fréquences cumulées d'une série statistique en fonction des valeurs de la variable.

Diagramme ascendant : on monte une marche de l'escalier à chaque valeur de X_i et cette marche a pour une hauteur l'effectif $n_i$correspondant à X_i. Chaque palier compris entre deux valeurs consécutives de la variable statistique X_i et X_{i+1}

Ce diagramme représente en effet l'effectif cumulé F(x) pour lequel la valeur prise par la variable statistique est inférieure.

Diagramme descendant : le principe est le même que ci-dessus



B2. Cas d'un caractère quantitatif continu :

Remarque :

Dans la réalité, une variable n'est pas toujours continue. Le degré de précision de la mesure impose souvent une discontinuité. Or, lorsque la précision de la mesure introduit une série trop grande de valeurs possibles, la présentation dans un tableau de ces valeurs possibles de X rendrait celui-ci très difficile à lire. Il convient donc de grouper ces valeurs dans des classes convenablement choisies.

- **Histogramme :** deux cas sont distingués selon les amplitudes des classes

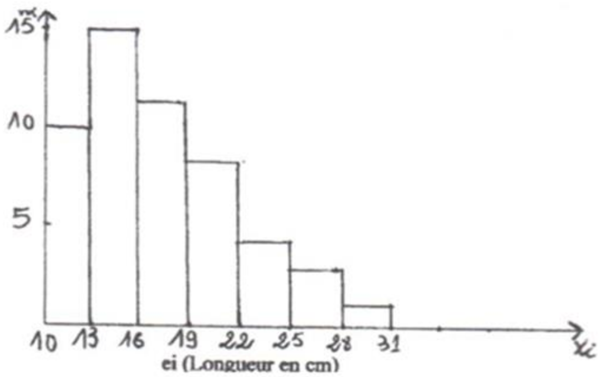
➢ **Cas d'amplitudes égales :**

Si toutes les amplitudes des classes sont égales, il n'y a pas lieu de rectifier les effectifs puisque les hauteurs des rectangles de l'histogramme sont directement proportionnelles aux effectifs.

Exemple : Soit la série suivante sur la distribution des salaires horaires des travailleurs.

Tableau 8 : Distribution des salaires horaires des travailleurs **Histogramme**

salaires horaires	n_i
[10-13[	10
[13-16[	15
[16-19[	12
[19-22[	07
[22-25[	03
[25-28[	02
[28-31[	01
Total	50



Si les classes sont d'amplitudes inégales, il faut d'abord rectifier l'effectif avant de tracer l'histogramme. Pour cela, on doit multiplier l'effectif de chaque classe par le rapport :

$$n_{ic} = n_i \times \frac{\text{amplitude la plus petite}}{\text{amplitude de la classse considérée}}$$

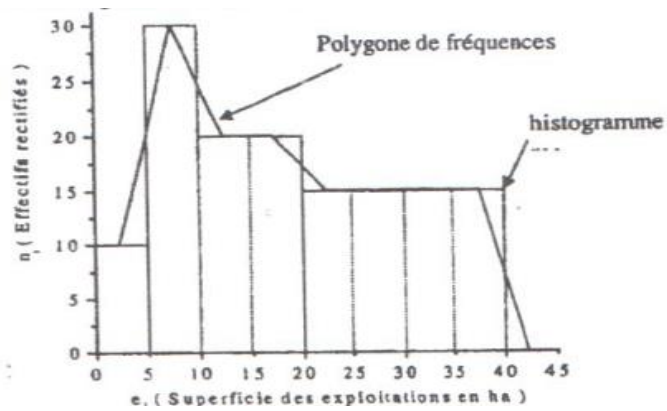
Exemple 2 :

Le tableau ci-dessous donne la répartition des exploitations agricoles selon leur superficies :

Tableau 9 : Distribution de 140 exploitations agricoles selon leur superficie (en hectare)

Superficie (Hectare)	n_i	a_i	n_i **Corrigés**	$n_i \uparrow$	$n_i \downarrow$
[0 – 5[	10	5	10	10	140
[5 –10[	30	5	30	40	130
[10-20[	40	10	$\frac{40 \times 5}{10} = 20$	80	100
[20-40[	60	20	$\frac{60 \times 5}{20} = 15$	140	60
Total	140	-	-	-	-

Comme les amplitudes des classes sont inégales, on doit procéder à la correction des effectifs. Il ne s'agit que d'une correction partielle puisque la 3ème classe n'est que le double de la première et de la deuxième classe, celles ayant la plus petite amplitude. Il en est de même pour la dernière classe, qui est le quadruple de la première et deuxième classe. L'histogramme sera alors le suivant :



Attention : les effectifs rectifiés ne servent que pour la construction de l'histogramme.

- **Courbe cumulative croissante et décroissante**

Dans ce cas, nous traçons une courbe dont seuls quelques points sont connus.

*** Pour la courbe croissante,** on joint les points de coordonnées (ei, $n_i \uparrow$) c'est-à-dire :

- En abscisse, la valeur de l'extrémité droite eide la classe i

- En ordonnée, la valeur de l'effectif cumulé $n_i \uparrow$ correspond cette classe.

*** Pour la courbe décroissante,** on joint les points de coordonnées ($e_{i-1}, n_i \downarrow$) c'est-à-dire :

- En abscisse, la valeur de l'extrémité gauche e_{i-1} la classe i ;

- En ordonnée, la valeur de l'effectif cumulé $n_i \downarrow$ correspond cette classe.

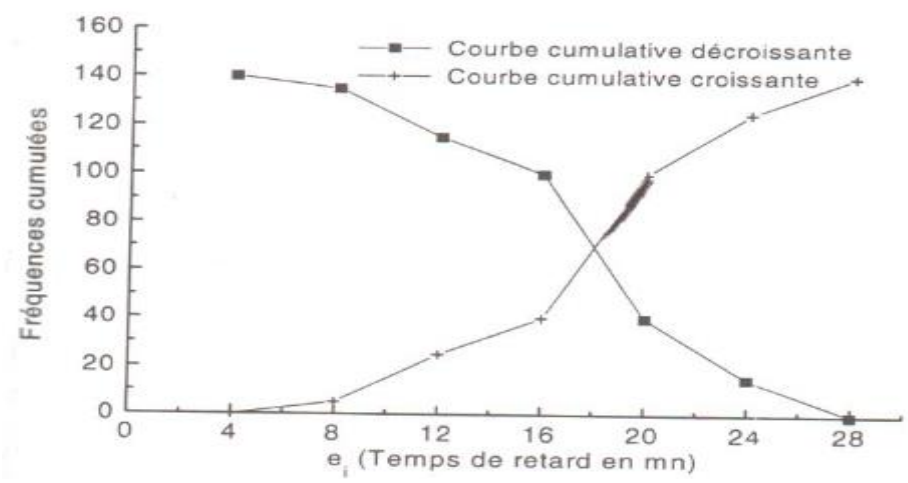
Exemple1 : Cas d'une variable dont les classes sont égales.

Reprenons les données du tableau 4 sur la répartition des 140 travailleurs selon leur temps de retard exprimé en minutes.

Tableau 10 : La répartition des 140 travailleurs selon leur temps de retard

Temps de retard	n_i	$n_i \uparrow$	$n_i \downarrow$
[04-08[	5	5	140
[08-12[	20	25	135
[12-16[	15	40	115
[16-20[	60	100	100
[20-24[	25	125	40
[24-28[	15	140	15
Total	140	-	-

La courbe cumulative croissante et décroissante



Remarque :

La même chose pour le cas d'une variable dont les classes sont inégales on trace les courbes de fréquences cumulées sans corriger l'effectif.

Attention :

C'est une courbe que l'on trace et non une succession de segments de droite. Par ailleurs, on remarquera que les courbes croissantes et décroissantes sont symétriques par rapport à un axe parallèle à l'axe des abscisses et ayant comme ordonnée N/2