
CHPITRE

1

Chapitre I : Rappel générale à la statistique

I.1- Généralités

I.1.1- Définitions

La statistique peut être défini comme étant un outil scientifique qui permet de recueillir, organiser, classer, résumer et présenter des informations statistiques qualitatives ou quantitatifs en vue de les analyser pour en prendre décisions judicieuses.

La statistique peut être aussi défini comme l'ensemble des méthodes (ou encore des techniques) généralement mathématiques permettant de décrire et d'analyser ou de traiter de façon quantifiée, des phénomènes repérés par des éléments nombreux, de même nature, susceptibles d'être dénombrés et classés (Bernard PY, 2007). Il est a noté aussi, qu'il faut distinguer cette dernière (**la statistique**) par rapport (**une statistique**) un terme parfois utilisé pour désigner « des données statistiques ».

De manière générale, on distingue deux groupes de méthodes statistiques : celui des méthodes descriptives et celui des méthodes inférentielles.

– **La statistique descriptive.**

Le terme de statistique descriptive, regroupe les méthodes dont l'objectif principal est la description des données étudiées, sans tenir compte en considération d'aucunes des hypothèses de type probabiliste sur ces données. Cette description des données se fait à travers leur présentation (la plus commode et la plus synthétique possible), leur représentation graphique et le calcul de résumés numériques (ou caractéristiques numériques).

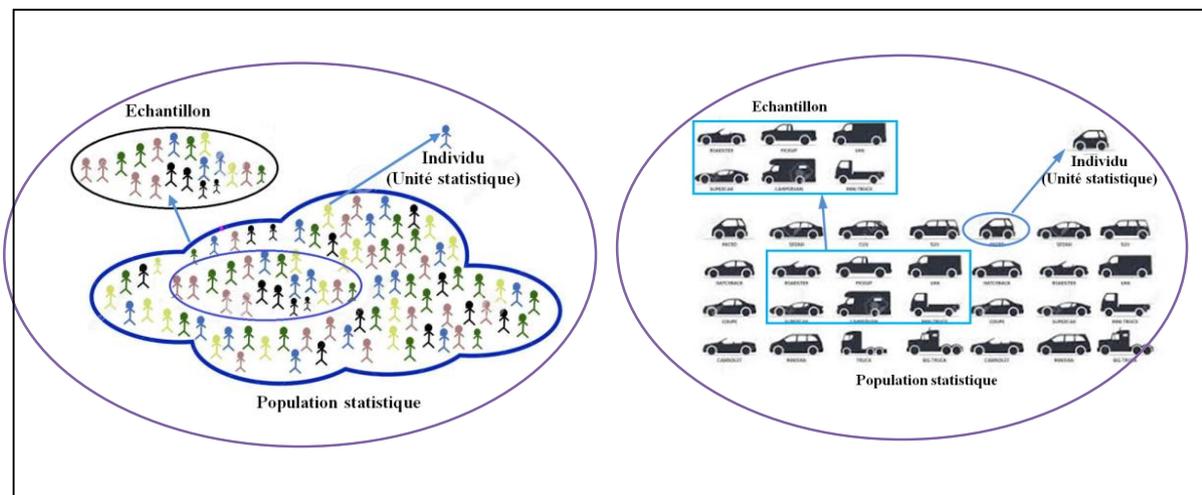
– **La statistique inférentielle.**

Le terme de statistique inférentielle, statistique mathématique ou statistique inductive regroupe les méthodes dont l'objectif principal est de préciser un phénomène sur une population globale, à partir de son observation sur une partie restreinte de cette population (échantillon). D'une certaine manière, il s'agit donc d'inférer du particulier au général. Le plus souvent, le passage d'échantillon à la population se fait moyennant des hypothèses de type probabiliste.

I.1.2- Vocabulaire de la statistique

Les termes généralement utilisés en statistiques sont :

– **Population (ou population statistique)** : c'est l'ensemble d'éléments ou d'individus appelés unités statistique concerné par une étude statistique. On parle aussi de champ de l'étude (Par exemple : notes d'un **groupe d'étudiants**, l'ensemble des **véhicules fabriqués par une usine**...).



a) Ensemble des étudiants

b) Ensemble des véhicules fabriqués par une usine

Figure 1 : Exemples de Population statistique, Echantillon et unité statistique

– **Individu (ou unité statistique)** : désigne tout élément de la population considérée. Dans le premier exemple (figure 1, a) : un individu est tout étudiant du groupe, et dans le deuxième exemple (figure 1, b) : un individu est tout véhicule susceptible d'être produit par une usine.

– **Echantillon** : On appelle échantillon le sous-ensemble de la population sur lequel sont effectivement réalisées les observations.

Généralement, dans une étude statistique, il est fréquent que l'on n'observe pas la population tout entière (par exemple, on n'observe pas tous les véhicules ayant circulé un jour donné dans une ville, mais seulement ceux étant passés dans certains points particuliers). Les observations du phénomène considéré sont donc réalisées sur une partie restreinte de la population, appelée échantillon.

– **Taille de l'échantillon** : noté généralement (N), c'est le nombre d'individus (unités statistiques) formant l'échantillon (échantillon de taille 50, de taille 200...).

– **Enquête (statistique)** : c'est l'opération consistant à observer (ou mesurer, ou questionner...) l'ensemble des individus d'un échantillon (ou, éventuellement, de la population complète).

– **Recensement** : enquête dans laquelle l'échantillon observé est en fait la population tout entière (on parle aussi d'enquête exhaustive). Par exemple le Recensement Général de la Population et de l'Habitat (RGPH, le recensement général agricole...)

– **Sondage** : c'est, au contraire, une enquête dans laquelle l'échantillon observé est un sous-ensemble strict de la population (on parle, dans ce cas, d'enquête non exhaustive).

– **Variable (statistique)** : c'est une caractéristique (âge, salaire, genre. . .), définie sur la population et observée sur l'échantillon.

D'un point de vue mathématique, une variable est une application définie sur l'échantillon.

Si cette application est à valeurs dans \mathbb{R} (ensemble des nombres réels), ou dans une partie de \mathbb{R} , elle est dite quantitative (âge, salaire, taille. . .) ; sinon elle est dite qualitative (grade ou fonction, catégorie socio-professionnelle ...).

On admettra alors que les variables quantitatives sont celles prenant des valeurs numériques et que les variables qualitatives sont celles prenant des valeurs non numériques.

Afin de différencier les deux types de critères, les critères qualitatifs sont appelés des caractères et les critères quantitatifs des variables.

– **Modalités** : On désigne par celles-ci différentes catégories d'un caractère qualitatif et on qualifie de valeurs les différents chiffres d'une variable

– **Modalités ordinales, modalités nominales** : Les modalités d'un caractère qualitatif, si elles ne peuvent pas être mesurées quantitativement, sont parfois susceptibles d'être classées. Ce sont des modalités ordinales.

– **Valeurs discrètes, valeurs continues** : Une variable quantitative peut-être discrète ou continue. Lorsque le nombre de valeurs possibles est fini (exemple : le nombre d'enfants, le nombre de pièces d'un logement, etc.), la variable est discrète. Lorsque le nombre de valeurs possibles de la variable est infini (exemple : la taille, le poids ou le revenu des ménages), la variable est continue.

– **Unités individuelles et unités groupées** : Les unités d'une population, que le critère soit qualitatif ou quantitatif (discret ou continu), peuvent être présentées individuellement (c'est généralement le cas lorsque les données sont saisies) ou regroupées. Le regroupement peut être effectué par modalités, par valeurs ou par classes de modalités ou de valeurs.

– **Données (statistiques)** : ce terme, désigne l'ensemble des individus observés (ceux de l'échantillon), l'ensemble des variables considérées et les observations de ces variables sur ces individus.

Les données sont en général présentées sous forme de tableaux (individus en lignes et variables en colonnes).

Tableau 01 : Exemple de données.

Individus	Genre	Age	Revenu mensuel net
individu 1	H	55	2068
individu 2	H	41	4687
individu 3	H	28	1235
individu 4	F	64	1941
individu 5	F	32	2456

Exemple 1 : soit une population de 600 étudiants, avec un effectif féminin de 230 et un effectif masculin de 370.

Les informations de cet exemple se traduisent dans le vocabulaire de la statistique descriptive comme suivant :

Tableau 02 : Exemple d'un critère qualitatif

P	Population	Effectif total : $n = 600$
i	unités statistiques	Chaque étudiant $i = 1, 2, \dots, n$
X	Caractère	Le genre
X_F X_M	Modalités	Féminin ou Masculin
n_M n_F	Effectifs associés à chaque modalité	370 hommes, 230 femmes

Exemple 2 : soit un échantillon de 10 étudiants ayant passé un examen. Ils ont obtenu les notes suivantes (sur 20) : {16, 8, 6, 14, 10, 18, 13, 9, 10, 15}.

Tableau 2 : Exemple d'un critère quantitatif

E	Échantillon	Effectif de l'échantillon : $n=10$
i	Unités statistiques	Chaque étudiant $i = 1, 2, \dots, n$
X	Variable	Notes
{ x_1, x_2, \dots, x_n }	Valeurs (*)	{6, 8, 9, 10, 13, 14, 15, 16, 18}
{ n_1, n_2, \dots, n_h }	Effectifs associés à chaque valeur	{1, 1, 1, 2, 1, 1, 1, 1, 1}

I.2- Statistique descriptive

Cette partie de la statistique concerne l'analyse exploratoire ou analyse des données, elle regroupe deux points importants :

- 1) Ensemble de méthodes : la statistique descriptive ne contient aucune théorie, mais seulement des outils d'investigation et de mesure des données chiffrées.
- 2) Décrire et analyser, de façon quantifiée, des phénomènes repérés par des éléments nombreux : décrire, c'est-à-dire faire des tableaux, des graphiques, calculer des moyennes afin de faire ressortir la signification.

I.2.1- Valeurs discrètes, valeurs continues

Une variable quantitative peut-être discrète ou continue. Lorsque le nombre de valeurs possibles est fini (le nombre d'enfants, le nombre de pièces d'un logement, etc.), la variable est discrète. Lorsque le nombre de valeurs possibles de la variable est infini (la taille, le poids ou le revenu des ménages), la variable statistique est dite continue.

I.2.2- Unités individuelles et unités groupées

Les unités d'une population, que le critère soit qualitatif ou quantitatif (discret ou continu), peuvent être présentées individuellement ou regroupées. Le regroupement peut être effectué par modalités, par valeurs ou par classes de modalités ou de valeurs.

Exemple 1 : Un questionnaire de satisfaction réalisé sur un échantillon de 10 consommateurs afin d'évaluer une prestation en cochant l'une des six catégories suivantes :

Questionnaire de satisfaction	
a) nulle,	<input type="checkbox"/>
b) médiocre,	<input type="checkbox"/>
c) moyenne,	<input type="checkbox"/>
d) assez bonne,	<input type="checkbox"/>
e) très bonne,	<input type="checkbox"/>
f) Excellente.	<input type="checkbox"/>

Ces données peuvent être représentées individuellement (tableau 3), groupées par modalités (tableau 4) ou par classes de modalités (tableau 5).

Tableau 3 : Données présentées individuellement

consommateur	1	2	3	4	5	6	7	8	9	10
Évaluation	a	e	e	c	e	f	a	f	e	b

Tableau 4 : Données groupées par modalités

Identificateur (*)	a	b	c	d	e	f
Évaluation	2	1	1	0	4	2

Tableau 5 : Données groupées par classes de modalités

Classes	De nulle à assez bonne (a – b – c – d)	De très bonne à excellente (e – f)
Effectif	4	6

Exemple 2 : On a mesuré 20 personnes et les résultats sont (en cm) :

{148, 165, 145, 173, 148, 145, 152, 180, 135, 170, 170, 170, 142, 148, 165, 175, 180, 180, 180, 180}

Il s'agit d'une variable continue (la taille), mais dont les valeurs sont ici connues individuellement. On peut aussi effectuer un regroupement par taille car certaines tailles, comme 170 ou 180, apparaissent plusieurs fois (tableau 6).

Tableau 6 : Données groupées par valeurs

Taille	135	142	145	148	152	165	170	173	175	180
Effectifs	1	1	2	3	1	2	3	1	1	5

Il est également possible d'effectuer un regroupement par classes de valeurs, soit un regroupement par classes d'amplitudes égales (tableau 7), ou un regroupement par classes d'amplitudes inégales (tableau 8). On désigne généralement par a_i , l'amplitude d'une classe.

Tableau 7 : Groupement par classes
(amplitudes égales)

Classes	Effectifs
[130-140[1
[140-150[6
[150-160[1
[160-170[2
[170-180]	10

Tableau 8 : Données groupées par
valeurs (amplitudes inégales)

Classes	Effectifs
[130 -150[7
[150 -170[3
[170 -180]	10

Lorsque les unités statistiques sont groupées par classes, on calcule un centre de classe, désigné par c_i , qui est égal à la moyenne des extrémités de classes (voir le tableau 9 pour le calcul des centres de classe du tableau 8).

Tableau 9 : Calcul des centres de classe des données du tableau 8

Classes	Centres de classe (c_i)
[130-150[$(130+150)/2 = 140$
[150-170[$(150+170)/2 = 160$
[170-180]	$(170+180)/2 = 175$

Exemple 3 : On a questionné 100 ménages sur le nombre d'ampoules électriques utilisées dans leur domicile. Dans le premier tableau, les données sont regroupées par nombre d'ampoules. Dans le second tableau, elles sont regroupées par classes.

Tableau 10 : Regroupement par nombre d'ampoules

d'ampoules	2	3	4	5	6	7	8	9	11	12	13	15
Effectifs	5	8	8	10	18	16	10	9	6	5	3	2

Tableau 11 : Regroupement par classes

Classes	Effectifs
[2-5[21
[5-10[63
[10-15[16

I.2.3- Effectifs, fréquences, pourcentages

Une fois les unités statistiques d'une population répertoriées, celles-ci sont présentées dans des tableaux sous diverses manières : effectifs ou fréquences absolues, fréquences relatives,. Il convient de définir ces termes avec précision :

1) Effectifs ou fréquences absolues

Il s'agit de la répartition brute des données. Lorsque les données sont présentées individuellement, chaque donnée a la même fréquence unitaire d'apparition, leur effectif ou fréquence absolue est égal à 1. Lorsque les données sont regroupées par valeurs ou modalités, les effectifs ou fréquences absolues correspondent au nombre de données qui ont la valeur ou modalité, ou encore qui sont groupées dans une classe donnée.

Symboliquement, les effectifs ou fréquences absolues s'écrivent n_i . Et la somme des effectifs est égale à N . Ainsi, dans le cas du tableau 11, les effectifs ou fréquences absolues sont respectivement égaux à $n_1=21$, $n_2=63$ et $n_3=16$. De plus, on a :

$$n_1 + n_2 + n_3 = 21 + 63 + 15 = 100 = N$$

2) Fréquences relatives et pourcentages

La fréquence relative est égale à la fréquence absolue divisée par l'effectif total : $f_i = \frac{n_i}{N}$

On a donc :

$$f_1 + f_2 + f_3 \dots \dots + f_h = \frac{n_1}{N} + \frac{n_2}{N} + \frac{n_3}{N} \dots \dots + \frac{n_h}{N} = \frac{n_1 + n_2 + n_3 \dots n_h}{N} = \frac{N}{N} = 1$$

Le pourcentage des données qui correspondent à une modalité, à une valeur ou à une classe s'obtient en multipliant la fréquence relative correspondante par 100. C'est-à-dire :

Pourcentage de la valeur (i) = $f_i \times 100$

Le tableau 12 reprend l'exemple de la répartition des ménages en fonction du nombre d'ampoules utilisées à leur domicile, en ajoutant la colonne des fréquences relatives à côté de celle des fréquences absolues. La dernière ligne correspond aux totaux.

Tableau 12 : Répartition des ménages en fonction du nombre d'ampoules à leur domicile

Classes	Effectifs ou fréquences absolues	Fréquences relatives	Pourcentages
[2-5[21	0,21	21
[5-10[63	0,63	63
[10-15]	16	0,16	16
Total	100	1	100

Les colonnes 2 (fréquences absolues) et 4 (pourcentages) contiennent les mêmes valeurs car l'effectif total est égal à 100. Si celui-ci était différent de 100, les valeurs contenues dans les deux colonnes seraient différentes.

I.3- Les séries statistiques à une dimension

I.3.1- Tableaux et graphiques

Les tableaux et les graphiques sont les principaux moyens de présentation des données statistiques.

I.3.1.1- Les Tableaux

Un tableau statistique est juste une mise en forme des données relative au caractère de la population que l'on souhaite étudier, présentée de façon la plus compréhensible possible.

A – Tableaux de données qualitatives

Le tableau ci-dessous présente la répartition par continent des utilisateurs d'Internet en 2021. Le caractère étudié – la répartition continentale des utilisateurs d'Internet – est qualitatif. Il a sept modalités, listées dans la première colonne. La seconde colonne indique les effectifs, c'est-à-dire ici le nombre d'utilisateurs d'internet dans chacune des zones. La dernière ligne, en caractères gras, indique le total mondial.

Tableau 1 : Utilisateurs d'Internet (Effectifs en Décembre 2021/ www.internetworldstats.com/stats)

Continent	Effectifs en millions
Africa	601,94
Asia	2 916,89
Europe	747,21
Latin America / Carib.	534,52
North America	347,91
Middle East	206,76
Oceania / Australia	30,55
Total	5 385,80

On peut aussi présenter ces données sous forme de pourcentage en ajoutant d'autres colonnes, afin d'en faciliter l'analyse et de mieux apprécier la part de chaque zone dans le total des utilisateurs (tableau 2).

Tableau 2 : Utilisateurs d'Internet (Effectifs en pourcentages en Décembre 2021)

Continent	Effectifs en millions	Pourcentages	Pourcentages cumulés
Africa	601,94	11.18%	11.18%
Asia	2 916,89	54.16%	65.34%
Europe	747,21	13.87%	79.21%
Latin America / Carib.	534,52	9.92%	89.13%
North America	347,91	6.46%	95.59%
Middle East	206,76	3.84%	99.43%
Oceania / Australia	30,55	0.57%	100.00%
Total	5 385,80	100	

B –Tableaux de données quantitatives**1) Variable quantitative discrète, valeurs connues individuellement**

Soit 100 ménages interrogés sur le nombre de pièces de leur logement. La variable « nombre de pièces » est quantitative et discrète (les valeurs sont dénombrables). En outre, les valeurs, n'ayant pas été groupées, sont connues individuellement. On obtient le tableau ci-dessous, où x_i représente le nombre de pièces et n_i les effectifs correspondants :

Tableau 4 : Nombre de pièces du logement (x_i , n_i)

Nombre de pièces (x_i)	Effectifs (n_i)
1	5
2	30
3	40
4	20
5	5

2) Variable quantitative discrète, valeurs regroupées

Considérons l'exemple précédent (le nombre de pièces sur des logements de 100 ménages). La variable « nombre de pièces » est quantitative et discrète (les valeurs sont dénombrables). Cette fois, les valeurs ont été groupées. On obtient le tableau ci- dessous :

Tableau 5 : Nombre de pièces du logement (x_i) Groupement par classes

Nombre de pièces (x_i)	Effectifs (n_i)
[1-3[35
[3-5]	65

3) Variable quantitative continue, valeurs connues individuellement

Soit un échantillon de 122 réponses d'étudiants à la question « À quel âge avez-vous obtenu votre bac ? ». Bien qu'il s'agisse d'une variable quantitative continue, les données sont présentées par âge et non par groupe d'âge. On obtient alors le tableau ci-après :

Tableau 6 : Âge d'obtention du bac (x_i) Groupement par valeurs

Age (x_i)	Effectifs (n_i)
16	5
17	25
18	45
19	20
20	15
21	8
22	4

4) Variable quantitative continue, données groupées

Considérons les données de l'exemple précédent (122 réponses d'étudiants à la question « À quel âge avez-vous obtenu votre bac ? »). Cette fois, les données sont présentées par groupe d'âge.

Tableau 7 : Âge d'obtention du bac (x_i) Groupement par classes

Age (x_i)	Effectifs (n_i)
[16-18[30
[18-20[80
[20-22]	12

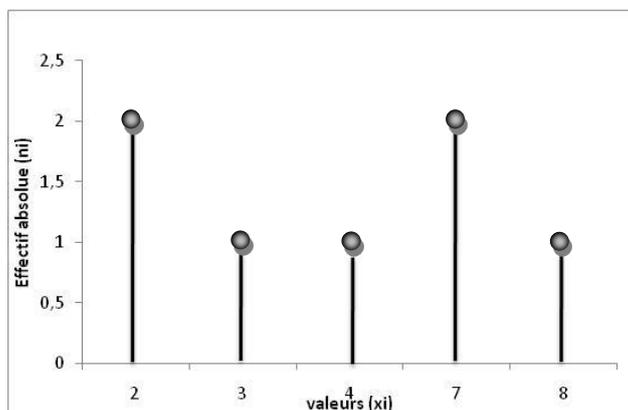
I.3.1.2- Graphiques

Les graphiques sont un autre moyen de représentation des données statistiques, parfois indispensable, surtout lorsqu'on veut mettre en valeur une comparaison saisissante. Ainsi, La présentation des données statistiques sous forme graphique joue un rôle essentiel pour permettre à un auditoire ou à des lecteurs de suivre une explication.

Nous nous allons voir que quelques exemples de graphiques les plus connus et les mieux adaptés aux données qu'il s'agit de représenter.

A – Données individuelles

Soit la série de chiffre où le 7 et le 2 sont répétés 2 fois : {8, 2, 3, 7, 4, 7, 2} : ces données peuvent être représentés à l'aide d'un diagramme en point individuels.



B – Données groupées par modalités ou valeurs

Que les données soient regroupées par modalité, comme c'est le cas pour les groupements qualitatifs, ou par valeurs, comme c'est le cas pour les groupements quantitatifs, on dispose de nombreuses représentations graphiques. Nous limiterons notre présentation aux plus connues, à savoir : le diagramme en bâtons et le diagramme en barres.

1) Diagramme en bâtons

Le diagramme en bâtons repose sur le principe que les valeurs ou modalités sont représentées par des bâtons, identiquement espacées les unes des autres.

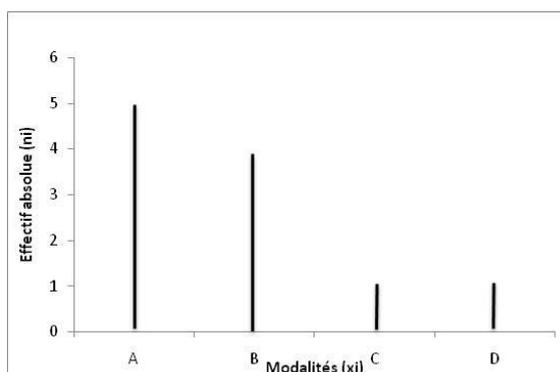


Diagramme en bâtons

2) Diagramme en barres

Le diagramme en barres repose sur le même principe que le diagramme en bâtons, sauf qu'au lieu de bâtons, on a des barres rectangulaires de base identique et identiquement espacées les unes des autres.

Soit un échantillon de 11 pour lesquels on cherche leurs préférences concernant 4 produits A,B,C,D. On obtient les résultats groupés suivants :

$$\{\{A, 4\}, \{B, 4\}, \{C, 1\}, \{D, 1\}\}$$

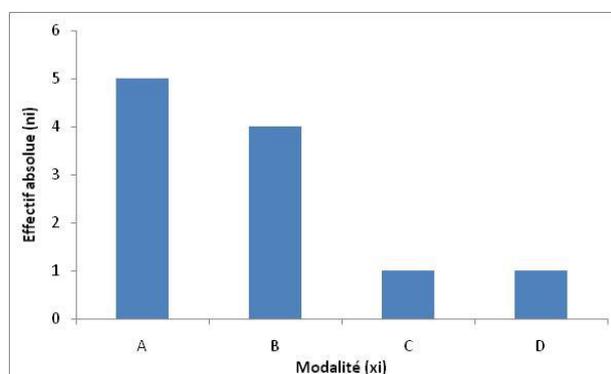


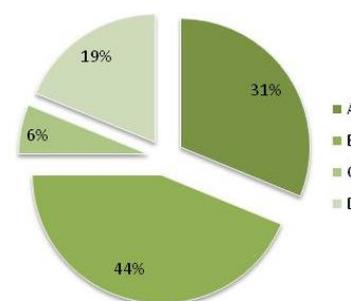
Diagramme en barres

3) Camembert ou graphique « en tarte » ?

Les Camembert ou « Pie Chart » c'est-à-dire, « graphique en tarte » est une représentation graphique universel convient à toutes les données, dès l'instant où il s'agit d'exprimer des parts ou des pourcentages.

Soit les chiffres d'affaires en millions d'euros des quatre principales entreprises du marché d'un produit (pour simplifier, on suppose que ces entreprises contrôlent la totalité du marché). ainsi, la part de marché n'est en fait qu'un pourcentage :

Entreprise	Chiffre d'affaires	Part de marché (%)
A	50	31,25
B	70	43,75
C	10	6,25
D	30	18,75
Total	160	100



Graphique en camembert

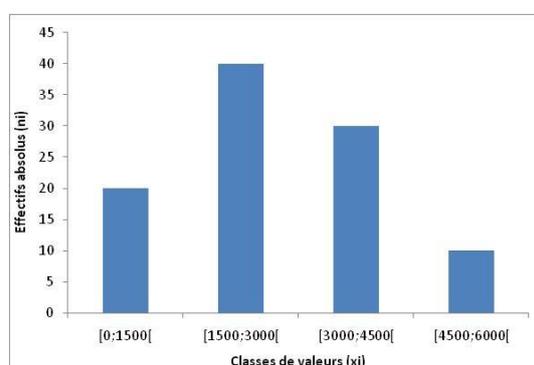
4) L'histogramme

Ce type de graphique convient particulièrement aux variables quantitatives regroupées par classes. Parfois les classes ont des amplitudes égales. C'est le cas le plus évident. Parfois non, cependant, les amplitudes des classes sont différentes. Cette dernière nécessite une correction des effectifs ou fréquences (Il faut alors opérer une correction en suivant la méthode indiquée ci-après. L'histogramme peut être construit à partir des effectifs (les n_i) ou à partir des fréquences (et d'ailleurs aussi en prenant les pourcentages).

Soit un échantillon de 100 ménages distribués selon leur revenu mensuel en euros. On définit des classes d'amplitudes égales à 1500 euros.

Tableau 9 : Répartition d'un échantillon de 100 ménages par classe de revenu mensuel (amplitude de classe = 1500 euros)

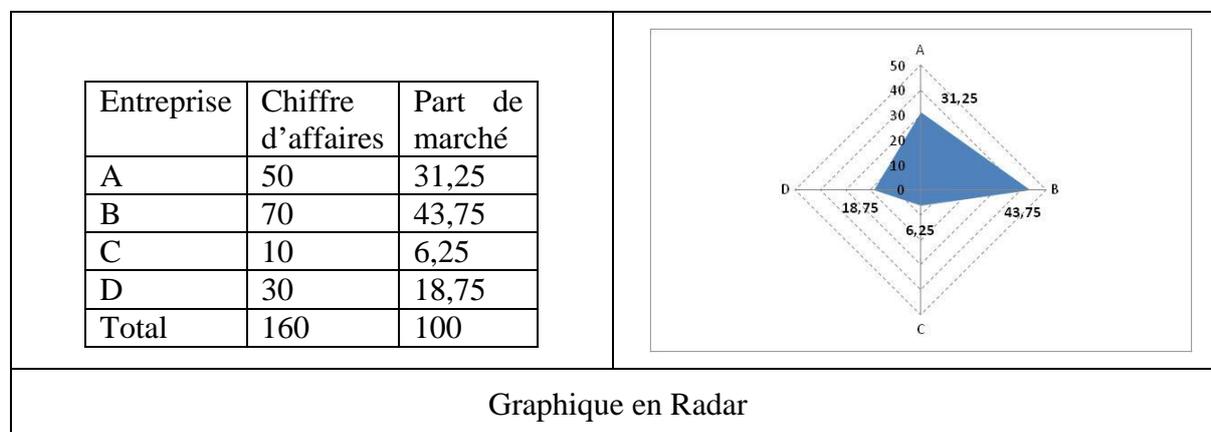
Classe de revenu	n_i	f_i
[0; 1500[20	0,2
[1500; 3000[40	0,4
[3000; 4500[30	0,3
[4500; 6000[10	0,1



Graphique en histogramme

5) Graphique en Radar

C'est une excellente représentation graphique pour faire des comparaisons visuelles (figure 11), ce type de graphique est utilisé afin de comparer la répartition réelle des parts de marché des 4 entreprises A, B, C et D avec une répartition égalitaire où chacune aurait 25% du marché (cette répartition égalitaire est représentée par les losanges en pointillé). Les parts de marché réelles sont indiquées sur chaque axe. On voit ainsi immédiatement que A et B ont une part de marché supérieure à la répartition égalitaire et B et C une part de marché inférieure. On peut à partir de là calculer combien il faut retrancher à A et à B (et combien par conséquent il faut redistribuer à C et D) pour revenir à une répartition égalitaire).



6) Les graphiques de séries chronologiques

Ce type de graphique est utilisé généralement pour faire des comparaisons dans le temps, les données sont organisées en fonction de temps appelées aussi séries chronologiques. Le résultat de cette représentation est une courbe de séries chronologiques, par exemple la figure (13), représente l'évolution de la population de l'Algérie pour la période 1961-2021, tiré d'un site web (données mondiales).

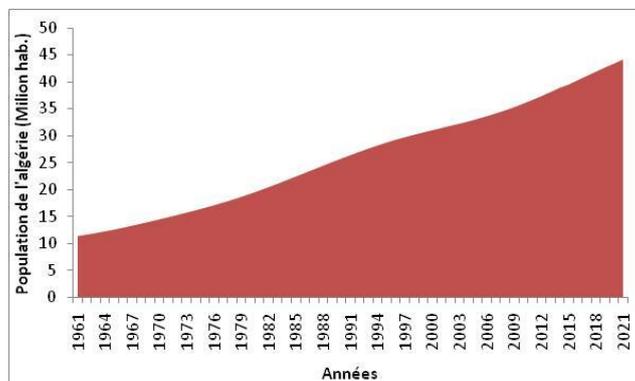


Figure : Evolution de la population de l'Algérie entre 1961 et 2021.