

STATISTIQUE DESCRIPTIVE

BOUALEM BENSEBAA
FACULTÉ DE MATHÉMATIQUES-USTHB

1. LA STATISTIQUE

Definition 1. La statistique est la science qui permet d'obtenir des informations sur un caractère en vue de les traiter.

Definition. Les informations recueillies sur un caractère (ou variable) en vue de les étudier et d'en tirer des conséquences sont appelées des statistiques.

Les caractères étudiés sont de deux natures : qualitatives et quantitatives.

Example. .

- (1) Caractère qualitatif : sexe du nouveau-né, mention de réussite, météo, qualification professionnelle..
- (2) Caractère quantitatif : elles sont de deux nature
 - (a) Discrètes : valeurs finies et entières : âge, nombre d'enfants,..
 - (b) Continues : toutes les valeurs peuvent être prises : salaires, taille, poids, température

2. QUELQUES DÉFINITIONS

- Unité statistique (ou individu) : élément de base de toute analyse statistique. Ces éléments doivent être distincts mais de même nature (personnes, objets, etc. . .).
- Population : l'ensemble des unités statistiques ou individus sur lequel portent les conclusions de l'étude.
- Échantillon : sous-ensemble de la population. Idéalement, on voudrait que ce sous-ensemble représente bien la population.

Remark. Note : Une enquête statistique effectuée à l'aide de toutes les unités statistiques d'une population s'appelle un recensement. Pour des raisons économiques, les enquêtes statistiques sont habituellement effectuées sur un échantillon seulement.

3. QUELQUES DÉFINITIONS

- Caractère : caractéristique ou propriété que l'on souhaite étudier. Cette caractéristique doit être définie et mesurée de la même manière pour chaque unité statistique.
- Modalités : valeurs distinctes prises par le caractère. La nature des modalités définit le « type » du caractère.

Example. .

- (1) On a enregistré le nombre d'enfants par ménage pour chacun des ménages d'une certaine ville d'Algérie
- (2) On veut connaître la proportion des familles d'une ville donnée possédant plus d'un véhicule

4. VARIABLE STATISTIQUE

Definition. Les caractères étudiés sont de nature qualitative (sexe du nouveau né, couleur des yeux, appréciation d'un professeur par les étudiants etc..) ou quantitative (taille en cm, le poids en kg, le nombre d'enfant d'une famille, le salaire annuel en DA, etc...).

Les caractères sont appelés variables statistiques

Definition. Une variable statistique est discrète si elle ne peut prendre qu'un nombre fini ou dénombrable de valeurs.

Une variable statistique est continue si elle peut prendre toutes les valeurs réelles dans un intervalle.

Example. .

- Letemps passé sur le web, la taille d'un individu, sont des variables continues
- le nombre d'enfants, le nombre d'années de scolarité, sont des variables discrètes.

5. CLASSEMENT DES DONNÉES

Definition. On appelle effectif (dans le cas d'une série discrète), le nombre de fois que l'on trouve la même valeur. Si l'effectif est exprimé en pourcentage on parle alors de fréquence de cette valeur Dans le cas d'une série continue, on répartit les données en classes

On rassemble les informations dans un tableau à deux colonnes où sont indiquées les valeurs du caractère et leurs effectifs.

Il faut prendre soin de classer par ordre croissant les observations d'un caractère quantitatif.

X (Modalités)	Effectifs n_i	Effectifs f_i	Effectifs F_i
x_1	n_1	f_1	$F_1 = f_1$
x_2	n_2	f_2	$F_2 = f_1 + f_2$
...
x_j	n_j	f_j	$F_j = F_{j-1} + f_j$

6. EXEMPLES

Considérons les prix de 105 maisons vendues par une agence immobilière :

1502000 1511000 1512000 1569000 1584000 1624000 1649000 1676000 1691000 1692000 1713000
1717000 1720000 1721000 1776000 1778000

1784000 1789000 1792000 1805000 1814000 1848000 1862000 1867000 1895000 1901000 1914000
1918000 1924000 1929000 1938000 1939000 1947000 1948000 1950000 1958000 1962000 1962000
1967000 1974000

1986000 2018000 2020000 2022000 2022000 2037000 2041000 2047000

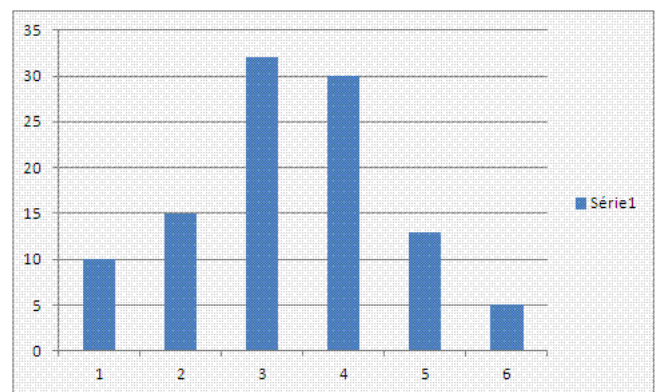
2048000 2063000 2063000 2065000 2066000 2077000 2078000 2079000 2097000 2106000 2108000
2109000 2110000 2117000 2117000 2121000 2128000 2131000 2132000 2140000 2158000 2172000
2178000 2182000

2183000 2199000 2204000 2219000 2226000 2235000 2236000 2238000 2260000 2268000 2273000
 2276000 2287000 2295000 2297000 2310000 2323000 2345000 2357000 2374000 2401000 2435000
 2440000 2441000

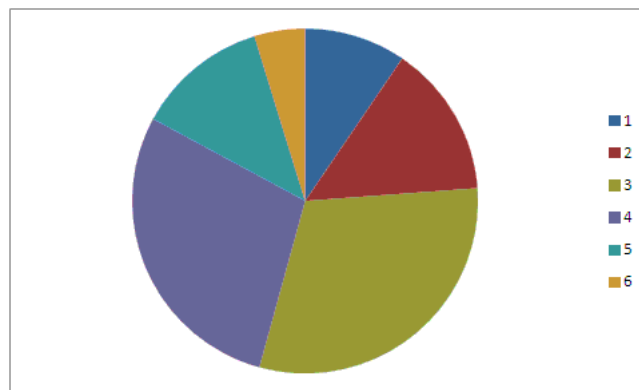
2470000 2476000 2478000 2486000 2522000 2546000 2569000 2587000 2669000

On utilise des classes d'intervalles pour regrouper les données

Prix	n_i	f_i	F_i
[1500000, 1700000)	10	10/105	10/105
[1700000, 1900000)	15	15/105	25/105
[1900000, 2100000)	32	32/105	57/105
[2100000, 2300000)	30	30/105	87/105
[2300000, 2500000)	13	13/105	100/105
[2500000, 2700000)	5	5/105	1
Total	105	1	



Représentation graphique : Graphique en Bâtonnet.



Graphique Secteurs.

Types.

(1) Graphique des effectif

(a) Série discrète : en bâtonnets

(b) Série continues : histogrammes de rectangles

Dans ce cas, on définit le polygone des effectifs reliant le milieu des sommets des rectangles.

Remark. Il est important que lors du classement des effectifs en classe, il faut avoir des classes de même amplitude. Le choix d'une amplitude d'une classe se fait sur celle qui revient le plus souvent.

- (1) Si une classe est d'amplitude t fois l'amplitude de la classe choisie, on opère une correction sur l'effectif de cette classe en le divisant par t

7. EFFECTIFS ET FRÉQUENCES

- (1) L'effectif n_i d'une valeur est le nombre d'apparition de cette valeur. Si N est l'effectif total des observations, alors

$$\sum n_i = N$$

- (a) La fréquence f_i associée à un effectif n_i est donnée par

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

et on a $\sum f_i = 1$.

- (2) Les effectifs cumulés croissants (ECC)
- (a) Dans le cas d'une série continue, l'effectif cumulé croissant correspondant à une classe $[a, b[$ est le nombre d'individus de valeurs strictement plus petites que b .
- (i) Dans le cas d'une série discrète, l'effectif cumulé croissant correspondant à une valeur, est le nombre d'individus de valeurs strictement plus petites que celle-ci.
- (b) Les fréquences cumulées croissantes sont définies de la même façon que les effectifs cumulés croissants en remplaçant les définitions précédentes le mot effectif par le mot fréquence.

8. CARACTÉRISTIQUES DE TENDANCES

- (1) LA MOYENNE

Definition. La moyenne se calcule tout simplement en divisant la somme des observations par leur nombre. Au niveau de la notation, on distinguera les moyennes d'une population et d'un échantillon comme suit :

Notation : μ : moyenne d'une population :

\bar{x} : moyenne d'un échantillon

- - Données brutes : $\sum \frac{x_i}{N}$, où les x_i représentent les N observations.

- Distribution : $\frac{\sum n_i x_i}{N} = \sum f_i x_i$

- Distribution en classe : $\sum f_i C_i$, où les C_i sont les centres des classes

- Distribution :

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

où les x_i représentent les modalités de la variable.

- Distribution pour données regroupées en classes : (approximation)

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

= où les C_i représentent les centres de classes

Propriétés de la moyenne.

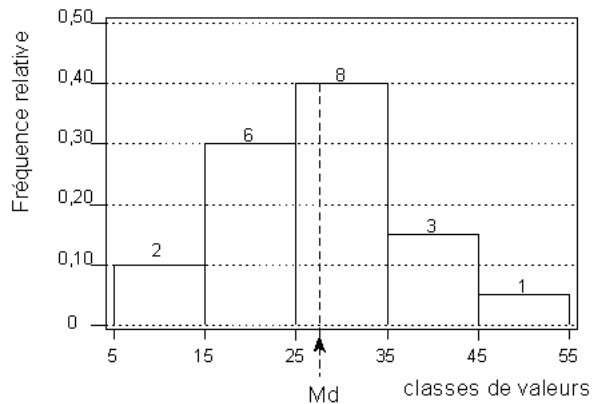
- (1) C'est la mesure la plus largement utilisée.
- (2) La moyenne correspond au centre de gravité de la distribution.
- (3) Elle n'a de sens que pour un caractère quantitatif (mesuré sur une échelle d'intervalles ou de rapport)
- (4) Elle peut être influencée par les valeurs extrêmes (observations très grandes ou très petites).
En pareil cas, elle jouerait mal son rôle pour mesurer la tendance « centrale » de la série d'observations.

Definition. Le Mode, ou classe modale, est la valeur, ou la classe, correspondant au plus grand effectif

Definition. La Médiane est la valeur partageant l'effectif en deux parties égales.

Concrètement, la médiane se calcule de la façon suivante

- (1) Dans le cas des séries discrètes,
Si l'effectif N est impair, c'est la valeur correspondant à l'Effectif Cummulé Croissant (ECC) : $(N + 1) / 2$
Si N est pair, c'est la valeur comprise entre deux observations $\frac{N}{2}$ et $\left(\frac{N}{2} + 1\right)$ d'effectif n



Déterminer la classe médiane (celle qui contient la $(N/2)$ e observation.

$$M_d = \text{Borne Inf} + \frac{(0,5 - F_{i-1})}{f_i} A_i$$

où f_i est la fréquence relative de la classe médiane, F_{i-1} la fréquence cummulée de la classe précédente et A_i est l'amplitude de la classe médiane .

Dans notre cas

$$M_d = 25 + \frac{0,5 - 0,4}{0,3} 10 = 25,33$$

Mesures de dispersion. Ce sont des informations sur la dispersion et l'éparpillement de

(1) L'étendue est la différence entre la plus grande et la plus petite valeur.

(2) Variance

$$(a) \text{ Données brutes } v = \frac{\sum (X_i - \mu)^2}{N} = \frac{\sum X_i^2}{N} - \mu^2$$

$$(b) \text{ Distribution } V = \frac{\sum N_i (X_i - \mu)^2}{N} = \frac{\sum N_i X_i^2}{N} - \mu^2$$

$$(c) \text{ Distribution en classe : } V = \frac{\sum (C_i - \mu)^2}{N} = \frac{\sum f_i C_i^2}{N} - \mu^2$$

Plus la variance est grande, plus grand est l'éparpillement.

9. QUANTILES

Definition. On appelle quantile d'ordre α , la valeur de la variable statistique qui est supérieur à une proportion de $\alpha\%$ d'observations et inférieur à une proportion $(1 - \alpha)\%$ d'observations.

- Si $\alpha = 10$, le quantile est appelé décile
- Si $\alpha = 25$, le quantile est appelé quartile
- Si $\alpha = 50$, le quantile est appelé 2ème quartile, c'est aussi la médiane
- Si $\alpha = 75$, le quantile est appelé 3ème quartile