

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Yahia Fares de Médéa

Statistiques descriptives

Mekdache Noureddine

2020 - 2021

Chapitre 1

Généralités sur la statistique

La statistique est l'étude de la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques.

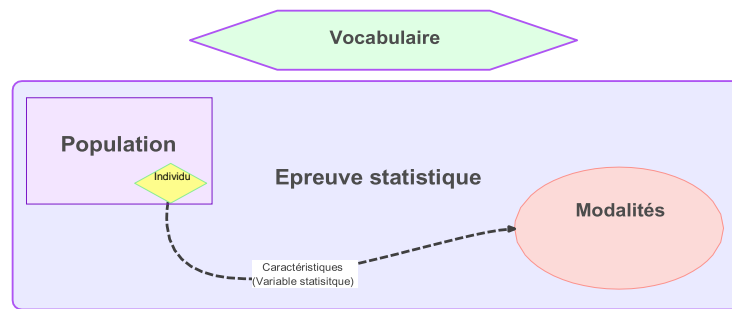
L'analyse des données est utilisée pour décrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les filières universitaires, de l'économie à la biologie en passant par la psychologie et bien sûr les sciences de l'ingénieur. La statistique consiste à :

- Recueillir des données.
- Présenter et résumer ces données.
- Tirer des conclusions sur la population étudiée et d'aider à la prise de décision.
- En présence de données dépendant du temps, nous essayons de faire de la prévision.

1.1 Vocabulaire

Les statistiques consistent en diverses méthodes de classement des données tels que les tableaux, les histogrammes et les graphiques, permettant d'organiser un grand nombre de données. Les statistiques se sont développées dans la deuxième moitié du XIX^e siècle dans le domaine des sciences humaines (sociologie, économie, anthropologie, ...). Elles se sont dotées d'un vocabulaire particulier.



1.1.1 Épreuve statistique

Les statistiques descriptives visent à étudier les caractéristiques d'un ensemble d'observations comme les mesures obtenues lors d'une expérience. L'expérience est l'étape préliminaire à toute étude statistique. Il s'agit de prendre "contact" avec les observations. De manière générale, la méthode statistique est basée sur le concept suivant.

Définition 1

L'épreuve statistique est une expérience que l'on provoque.

Exemple 1 (La durée de vie des lampes)

Imaginons le cas suivant : un fabricant d'ampoules électriques ayant le choix entre 4 types de filaments se propose d'étudier l'influence de la nature du filament sur la durée de vie des ampoules fabriquées. Pour ce faire, il va faire fabriquer 4 échantillons d'ampoules identiques, sauf en ce qui concerne le filament, faire brûler les ampoules jusqu'à extinction, puis comparer les résultats obtenus.

1.1.2 Population

En statistique, on travaille sur des populations. Ce terme vient du fait que la démographie, étude des populations humaines, a occupé une place centrale aux débuts de la statistique, notamment au travers des recensements de population. Mais, en statistique, le terme de population s'applique à tout objet statistique étudié, qu'il s'agisse d'étudiants (d'une université ou d'un pays), de ménages ou de n'importe quel autre ensemble sur lequel on fait des observations statistiques. Nous définissons la notion de population.

Définition 2

On appelle population l'ensemble sur lequel porte notre étude statistique. Cet ensemble est noté Ω .

Exemple 2

On considère l'ensemble des étudiants de la section A. On s'intéresse au nombre de frères et sœurs de chaque étudiant. Dans ce cas

Ω = ensemble des étudiants.

Si l'on s'intéresse maintenant à la circulation automobile dans une ville, la population est alors constituée de l'ensemble des véhicules susceptibles de circuler dans cette ville à une date donnée. Dans ce cas

Ω = ensemble des véhicules.

1.1.3 Individu (unité statistique)

Une population est composée d'individus. Les individus qui composent une population statistique sont appelés unités statistiques.

Définition 3

On appelle individu tout élément de la population Ω , il est noté ω (ω dans Ω).

Remarque 1

L'ensemble Ω peut être un ensemble de personnes, de choses ou d'animaux...

L'unité statistique est un objet pour lequel nous sommes intéressés à recueillir de l'information.

Exemple 3

– Dans l'exemple indiqué ci-dessus, un individu est tout étudiant de la section.

– Si on étudie la production annuelle d'une usine de boîtes de boisson en métal (canettes).

La population est l'ensemble des boîtes produites durant l'année et

une boîte constitue un individu.

1.1.4 Caractère (variable statistique)

La statistique « descriptive », comme son nom l'indique cherche à décrire une population donnée. Nous nous intéressons aux caractéristiques des unités qui peuvent prendre différentes valeurs.

Définition 4

On appelle caractère (ou variable statistique, dénotée $V.S$) toute application

$$X : \Omega \rightarrow C.$$

L'ensemble C est dit : ensemble des valeurs du caractère X (c'est ce qui est mesuré ou observé sur les individus)

Exemple 4

Taille, température, nationalité, couleur des yeux, catégorie socioprofessionnelle...

Remarque 2

Soit Ω un ensemble. On appelle et on note $\text{Card}(\Omega)$, le nombre d'éléments de Ω .

$$\text{Card}(\Omega) := \text{nombre d'éléments de } \Omega = N.$$

1.1.5 Modalités

Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci.

Exemple 5

–Variable est " situation familiale "

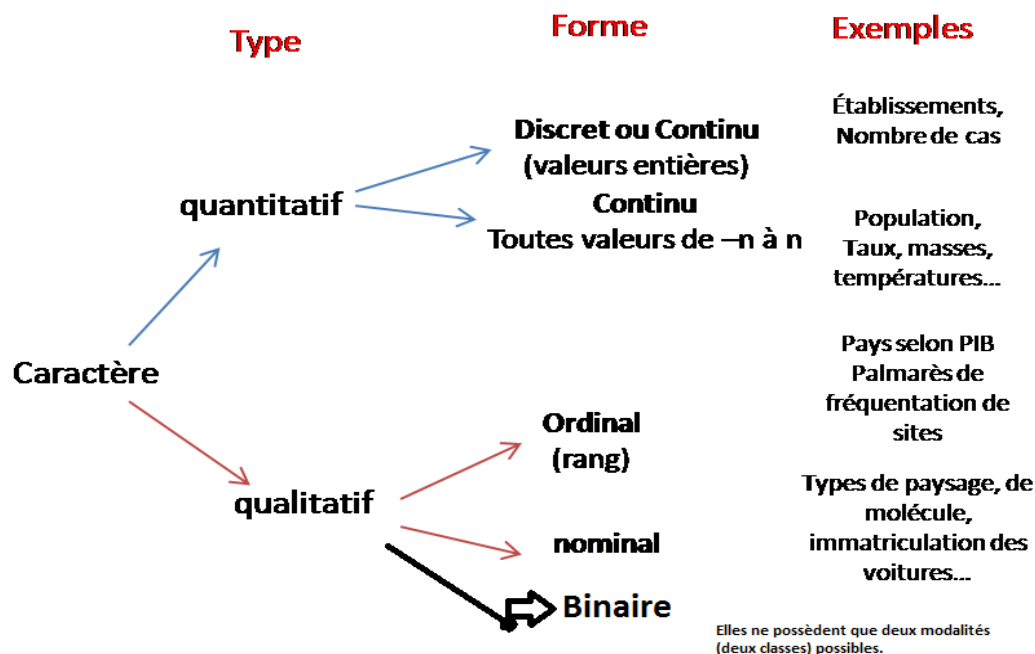
Modalités sont " célibataire, marié, divorcé "

- Variable est "statut d'interrupteur "
Modalités sont " 0 et 1 ".
- Variable est "catégories socio-professionnelles "
Modalités sont " Employés, ouvriers, retraités,... ".

Les modalités sont les différentes situations dans lesquelles les individus peuvent se trouver à l'égard du caractère considéré.

1.2 Types des caractères

Nous distinguons deux catégories de caractères : les caractères qualitatifs et les caractères quantitatifs.



1.2.1 Caractère qualitatif

C'est un caractère non mesurable. Les modalités ne sont pas des valeurs numériques.

Définition 5

Les éléments de C sont représentés par autre chose que des chiffres.

1.1.2 Caractère qualitatif ordinal :

Les modalités peuvent être ordonnées selon une certaine hiérarchie.

Exemple 6 : Niveau d'étude : primaire, secondaire, supérieur.

Etat mécanique d'une Voiture: mauvais, moyen, bon, excellent.

1.1.3 Variable qualitative nominale :

Exemple 7

On a demandé aux élèves d'une classe de terminale composée de 10 élèves quelle est leur chaîne de télévision préférée.

Population : classe de terminale.

Effectif total $N=10$

Individu : un élève de la classe.

Variable : chaîne de TV préférée.

Modalités: TF1, F2, F3, C+, M6.

C'est une variable qualitative nominale.

1.1.4 Variable qualitative binaire :

Exemple 8 :

Variable	Classes
État de santé	Malade / sain
Survie	Vivant /décédé
Sexe	Homme /femme
Tabagisme	Fumeur / non-fumeur

Représentations graphiques variables qualitatives

◦ Diagramme en bâtons

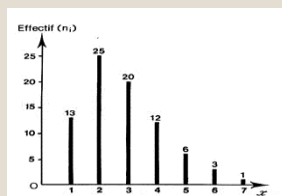
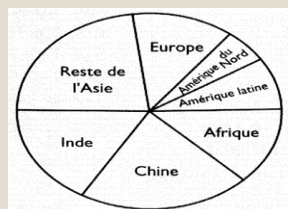


Diagramme en secteurs



2.1.1 Caractère quantitatif

Un caractère est quantitatif si les modalités sont des nombres.

2.1.2 Caractère quantitatif discret :

- Les notes représentent un caractère quantitatif discret. (Les valeurs prises sont isolées 4 , 5 , ... , 16 , 17)

2.2.2 Caractère quantitatif continu :

- Le temps de révision est un caractère quantitatif continu. (Le temps de révision pourrait être n'importe quel nombre t , tel que $21 \leq t < 33$ par exemple. Les valeurs de ce caractère sont regroupées en classes ($[0 ; 1 [$; $[1 ; 2 [$...) L'amplitude des classes n'est pas forcément la même . En général, on fait l'hypothèse d'une répartition uniforme à l'intérieur de chaque classe , nous avons la définition suivante.

Définition 6

L'ensemble des valeurs est représenté par des chiffres. De même, il est partagé en deux sortes de caractères, discret et continu (voir l'exemple).

Exemple 9

–Le salaire d'employés d'une usine. Modalités :

30000 da , 40000da... Type : Discret.

– Le salaire d'employés d'une usine.

Modalités: $[30000 ; 40000]$ da

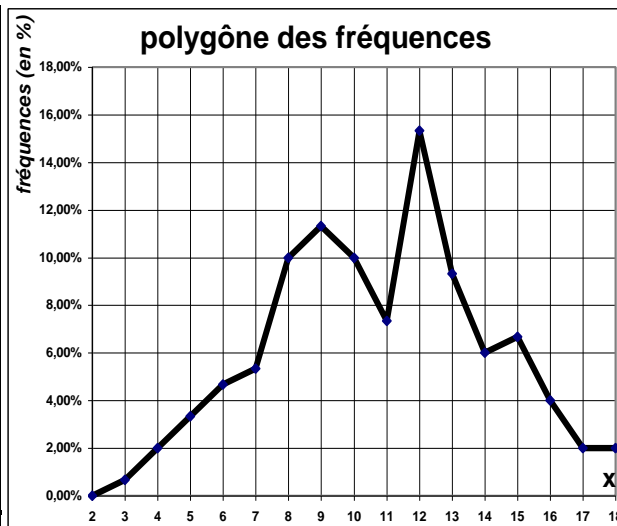
Type : continu.

Représentations graphiques variables quantitatives :

Exemple 10 :

X_i	f_i	%	$X_i f_i$
2	0	0,00%	0
3	1	0,67%	3
4	3	2,00%	12
5	5	3,33%	25
6	7	4,67%	42
7	8	5,33%	56
8	15	10,00%	120
9	17	11,33%	153
10	15	10,00%	150
11	11	7,33%	121
12	23	15,33%	276
13	14	9,33%	182
14	9	6,00%	126
15	10	6,67%	150
16	6	4,00%	96
17	3	2,00%	51
18	3	2,00%	54
Σ	150	100,00%	1617

c) **moyenne= 10.78**



d) **12 (le mode)**

e) **47,33%**

classes	C_i	f_i	%	$C_i f_i$
[1,3]	2	1	0,7%	2
[4,6]	5	15	10,0%	75
[7,9]	8	40	26,7%	320
[10,12]	11	49	32,7%	539
[13,15]	14	33	22,0%	462
[16,18]	17	12	8,0%	204
		150	100%	1602

moyenne= 10.68

