

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

**MINISTRE ENSEIGNEMENT SUPERIEUR RECHERCHE
SCIENTIFIQUE**

**UNIVERSITE FERHAT ABBAS DE SETIF
FACULTE DE MEDECINE DEPARTEMENT DE MEDECINE
SERVICE EPIDEMIOLOGIE MEDECINE PREVENTIVE**

*MODULE EPIDEMIOLOGIE
6eme ANNEE
ANNÉE UNIVERSITAIRE 2020/2021.*

TITRE :

STATISTIQUES

Pr Bouaoud Souad

I/ Généralités et notions de base :

1 Définition de la statistique :

La statistique est l'ensemble de méthodes scientifiques à partir desquelles on recueille, organise, mesure, présente, et analyse des données.

Les statistiques sont l'ensemble des données ou dénombrement que l'on recueille en effectuant des recensements ou des sondages, c'est une méthode de raisonnement scientifique mathématique et logique appliquée à des ensembles, et permettant d'interpréter les différents types de données qu'on rencontre dans les différents domaines de la vie et dont le caractère principal est la variabilité.

2 Population ou univers :

C'est une collection ou un ensemble d'objets ou de personnes étudiés : c'est un ensemble d'unités susceptibles d'être observées, définies par les caractéristiques de temps, de lieu, de sexe, de catégories socio-professionnelles ou maladie.

Exemple : population d'individus, d'animaux, d'événement (tremblement de terre...)

3 Unités d'observation :

C'est l'unité élémentaire constituant la population.

C'est un objet ou une personne sur lesquels porte l'étude.

4 L'échantillon : c'est une partie réduite d'une population.

C'est un sous-groupe faisant l'objet d'une observation ou d'une étude partielle. Pour pouvoir extrapoler les résultats de l'étude à l'ensemble de la population il faut que l'échantillon soit représentatif de cette dernière.

5 Les constantes et les variables :

Une constante est une caractéristique qui a la même valeur pour tous les individus, Exp : le nombre de mains.

Contrairement à une constante, une variable comporte nécessairement plus d'une modalité.

Les modalités sont les différentes catégories que peut présenter une variable, C'est ainsi que le sexe est une variable car il comprend deux modalités : homme et femme.

Même la maladie est une variable car on peut lui définir au moins deux modalités.

On distingue les variables quantitatives et les variables qualitatives.

Les Variables Qualitatives : Représentent des attribues non mesurables

Exp : le sexe, la race, la couleur des yeux, groupe sanguin...etc.

La variable qualitative est discrète et non numérique.

Les Variables Quantitatives : sont des variables qui prennent, des valeurs numériques.

On distingue : les variables quantitatives discrètes ou discontinues et les variables quantitatives continues.

Les variables quantitatives discrètes ou discontinues : sont des variables qui ne peuvent prendre que des valeurs numériques isolées dans un intervalle exp : le nombre d'enfants par famille prend des valeurs de 0 à 4-5-6...etc mais on ne peut pas avoir 3,1 enfant par exemple donc le nombre d'enfants /famille prend des valeurs isolées.

Les variables quantitatives continues : sont des variables qui peuvent prendre toutes les valeurs numériques d'un intervalle déterminé exp : TA°-t°-poids-taille...etc.

6 L'effectif absolu : (fréquence absolue) : c'est le nombre d'individus appartenant à une modalité donnée .exp : on a 60 malades (N) dont 25 hommes (n) et 35 femmes (n).

7 La fréquence effectif relatif : c'est le rapport entre l'effectif d'une modalité de la variables étudiée et l'effectif de la population ou plus généralement de l'échantillon sur lequel cette variable est mesurée, le numérateur fait obligatoirement partie du dénominateur, la fréquence relative s'exprime généralement par un pourcentage (%).

8 L'effectif cumulé : l'EC est égale à la somme des effectifs de ce même groupe et de tous les groupes antérieurs, il peut être absolu ou relatif.

9 Série statistique : « SS » : La série statistique est un ensemble de valeurs numériques résultantes d'une série d'observation.

-Chaque observation est une unité statistique.

-Les résultats de l'observation sont donnés à l'état brut dans un ordre quelconque d'où la mise en ordre de ces données.

-Les observations sont généralement résumées dans un tableau selon la nature des caractères étudiés : continus ou discontinus.

-On peut distinguer les distributions continues et les distributions discontinues.

-Les SS peuvent être classées en grandes ou en petites séries.

II-Statistiques Descriptives :

1-Buts :

- Décrire d'une façon simple et précise les données d'une distribution
- Rendre leur interprétation plus aisée.
- Comparer plusieurs SS entre elles

2-Traitement des données

Le traitement des données passe par plusieurs étapes :

- * la collecte de l'information et la mise en ordre des données.
- * la présentation tabulaire.
- * la présentation graphique
- * la réduction des données.

01- La collecte de l'information et la mise en ordre des données : la mise en ordre des données c'est la 1^{ère} étape du TRT après la collecte de l'information, elle consiste à donner les modalités recueillies de la variables étudiée par ordre.

02- La présentation tabulaire : consiste à présenter les informations recueillies sous forme de tableaux, permettant ainsi de donner une vision globale des résultats.

- le contenu du tableau : il doit contenir toutes les informations nécessaires à sa compréhension.

Un tableau doit avoir un numéro et un titre.

Le numéro est généralement en chiffre romain.

Le titre doit être précis explicite et informatif et doit répondre à 3 questions : quoi ? Où ? et quand ?.

Le tableau proprement dit comporte : l'entête-les têtes de lignes- les têtes de colonnes et un corps ou sont inscrits des données chiffrées.

03- La présentation graphique : a l'avantage d'offrir une meilleure vue d'ensemble de la SS que les tableaux. Elle permet par simple lecture de voir les caractéristiques essentielles de la série.

Le graphique doit contenir les informations suivantes :

- Un numéro et un titre qui doivent avoir de même contenu que ceux du tableau, des coordonnées, des axes et des échelles.
- Un tracé.

La présentation graphique diffère selon le caractère étudié.

-Pour les variables qualitatives la présentation graphique (PG) se fait par un graphique circulaire ou un diagramme à pilier ou un diagramme à bâtonnets.

-Pour les variables quantitatives discontinues la présentation graphique se fait par un diagramme à piliers ou à bâtonnets et pour les variables quantitatives continues, elle se fait sous forme d'un histogramme.

04 Réduction des données :

a) **Les paramètres de tendance centrale** : Ce sont des mesures qui localisent le centre de la SS, ils sont au nombre de 3 : la moyenne-le mode-et la médiane.

01- **La moyenne** : c'est la somme des valeurs rencontrées pour la variable étudiée par / au nombre total d'observation :(effectif total) :

$$m = \sum xi / N$$

m=moyenne arithmétique simple

xi= la valeur de la variable

N=l'effectif total (échantillon ou population)

Lorsqu'on a plusieurs modalités donc on parle d'une moyenne arithmétique pondérée, en multipliant la valeur par le nombre d'observation lui correspondant

$$m = \sum nixi / N$$

xi : devient le centre de la classe pour les variables groupées.

ni l'affectif correspondant à la modalité xi.

02 **Le mode (la valeur modale)** : c'est la valeur de la variable qui correspond à l'effectif absolu ou relatif le plus élevé.

03 **La médiane** : c'est la valeur de la variable qui divise l'effectif total de la distribution statistique en deux parties égales.

b) - **les paramètres de position** : les valeurs ou les modalités de la variable qui occupent un certain rang dans une série statistique lorsque les termes de celle-ci sont ordonnés par ordre croissant on décroissant.

On distingue :

***Le quartile** : qui divise l'effectif total de la distribution statistique en 4 parties égales

***Le décile**: qui divise l'effectif total de la distribution statistique en 10 parties égales

*Le **percentile**: qui divise l'effectif total de la distribution statistique en 100 parties égales

c) - **les paramètres de dispersion** :

01-La marge ou l'étendue : qui est = à la différence entre la plus grande et la petite valeur de la SS.

02-La variance (S^2) est égale à la somme des écarts par / à la moyenne rapportée à l'effectif total

$$S^2 = (\sum ni (xi-m)^2) / N$$

03-L'écart type c'est la racine carrée de la variance : $S = \sqrt{S^2}$

04-le coefficient de la variance : **CV** : C'est un paramètre qui estime le degré d'homogénéité de la Série Statistique : $CV = (S/m) \cdot 100$.