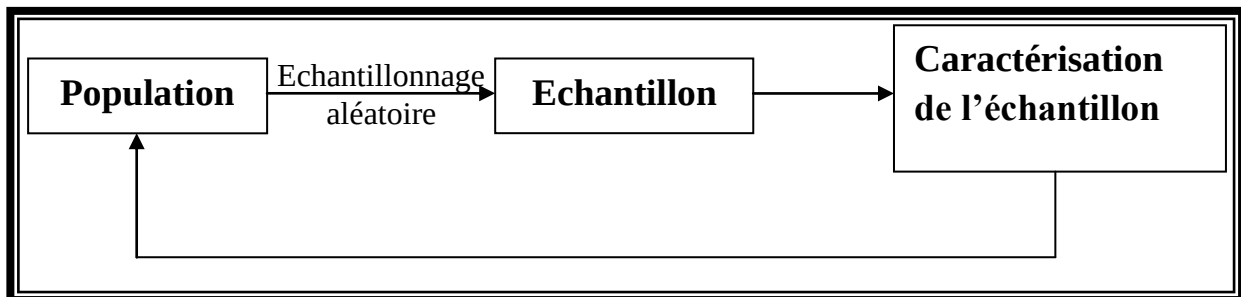


Chapitre 1 : Statistiques Descriptives

I. Introduction :

- La statistique est une méthode scientifique qui consiste à réunir des données chiffrées sur des ensembles nombreux, à les analyser et à commenter et critiquer ces données.
- On peut résumer les statistiques descriptives par le schéma suivant :



- Les ensembles étudiés sont appelés « population ».
- Les éléments de la population sont appelés « individus » ou « unités statistiques ».
- La population est étudiée selon un ou plusieurs caractères.

II. L'échantillonnage statistique :

Pour recueillir des informations sur une population statistique l'on dispose de deux méthodes :

1/ méthode de recensement :

Chaque individu de la population est étudié selon les caractères définit pour l'étude.

2/Méthode d'échantillonnage :

Qui conduit à examiner une fraction de la population appelé **échantillon**.

L'échantillonnage représente l'ensemble des opérations qui ont pour objet de prélever un certain nombre d'individu dans une population donnée.

L'échantillon doit être représentatif de la population et seul l'*échantillonnage aléatoire* assure la représentation de la composition et de la complexité.

L'échantillonnage aléatoire simple est une méthode qui consiste à prélever au hasard et d'une façon indépendante « n » individus d'une population à « N » individus.

III. Les caractères statistiques :

1/les caractères qualitatifs :

Exemple : la couleur, les groupes sanguin, les nucléotides d'ADN et la présence ou l'absence d'un caractèreetc.

2/les caractères quantitatifs :

Le caractère quantitatif est **discret** s'il peut prendre uniquement certaines valeurs dans un intervalle donné. En général, il en résulte d'un comptage ou un dénombrement.

Exemple : - Le nombre de petit par ponte.

- Le nombre de cellule dans une culture.....etc.

Le caractère quantitatif est continue s'il peut prendre théoriquement n'importe quelle valeur dans un intervalle en générale il résulte d'une mesure.

Exemple : - Le poids, la taille, taux de glycémie, le rendementetc.

Remarque : *Le nombre de valeurs possibles dans un caractère donné dépend de la précision de la mesure.*

IV. La représentation des données :

Il existe plusieurs niveaux de descriptions statistiques :

- La présentation brute des données.
- Des présentations par tableau numérique.
- Des représentations graphiques et des résumés numériques fournit par un petit nombre de paramètres caractéristiques.

1. Les séries statistiques :

Une série statistique correspond aux différentes modalités d'un caractère sur un échantillon. Le nombre d'individu qui constitue l'échantillon s'appelle **la taille de l'échantillon**.

2. Tableaux statistiques :

Le tableau de distribution de fréquence est un mode synthétique de présentation des donnés.

Sa constitution est immédiate dans le cas d'un caractère discret mais nécessite une transformation des donnés dans le cas d'un caractère continue.

3. Les présentations graphiques :

Ont l'avantage de renseigner immédiatement sur l'allure générale de la distribution. Elles facilitent l'interprétation des donnés.

a) *Les caractères quantitatifs discrets :*

La présentation graphique est le diagramme en battons ou la hauteur des battons correspond à l'effectif « n_i » associé à chaque modalité du caractère « X_i ».

b) *Les caractères quantitatifs continus :*

La représentation est un histogramme ou la hauteur du rectangle est proportionnelle à l'effectif « n_i ».

V. Les indicateurs numériques :

Le dernier niveau de la description statistique et le résumé numérique par des indicateurs ou des paramètres caractéristiques.

V.1. Indicateurs de Position :

1. La moyenne :

Soit un échantillon de « n » valeur observé ($x_1, x_2, x_3, \dots, x_n$) d'un caractère quantitatif « x ». On définit sa moyenne observée $\bar{X}$ comme la moyenne arithmétique de « n » valeurs.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

2. La médiane :

C'est une valeur du caractère pour laquelle la fréquence cumulée est égale à 0,5 ou 50%. Elle correspond donc au centre de la série statistique classée par ordre croissant ou à la valeur pour laquelle 50% des valeurs observées sont supérieures, et 50% sont inférieures.

* Quand le nombre est *pair* de données, la médiane est donnée par la formule suivante :

$$\text{Médiane} = \text{Me}_0 = [X_{n/2} + X_{(n/2+1)}]/2$$

* Quand le nombre est *impair* de données, la médiane est donnée par la formule suivante :

$$\text{Médiane} = \text{Me}_0 = X_{(n+1)/2}$$

Remarque 1 : La médiane ne s'applique qu'aux échelles ordinales d'intervalles et de rapport car elle nécessite un ordre linéaire entre les variables.

Remarque 2 : La moyenne et la médiane sont considérées comme des indicateurs de position et elles ont pour objectifs de caractériser l'ordre des observations.

V.1. Indicateurs de Variation :

1. La variance :

Soit un échantillon de N valeurs observées ($x_1, x_2, x_3, \dots, x_n$) d'un caractère quantitatif X et d'une moyenne $\bar{X}$.

On définit la variance « δ^2 » comme la moyenne arithmétique des carrés des écarts à la moyenne et elle est donnée par la formule suivante :

$$\text{Variance} = \delta^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

2. L'écart type :

L'écart type observé d'un échantillon correspond à la racine carré de la variance :

$$\text{Écart type} = S = \sqrt{\delta^2} = \sqrt{var}$$

Remarques :

- * La variance est toujours un nombre **positif**.
- * Elle n'a pas de sens biologique direct contrairement à l'écart type qui s'exprime dans les mêmes unités de la moyenne.

3. Le coefficient de variation :

Il est noté « **C.V** » c'est un indice de dispersion relative, et il est donné par la formule suivante :

$$\text{C.V} = \frac{S}{\bar{X}} * 100\%$$

Ce coefficient est indépendant du choix des unités et permet la comparaison des distributions de fréquence d'unités différentes.