

LE SYSTÈME D'INFORMATION SANITAIRE

MODULE 3 : TRAITEMENT ET ANALYSE DES DONNEES

Objectifs du module

A l'issu de cette présentation vous devrez :

- ▶ connaître les notions de base en statistique ;
 - ▶ Savoir calculer les paramètres de position et de dispersion couramment utilisés ;
 - ▶ présenter les données de façon appropriée ;
 - ▶ présenter les indicateurs selon le mode approprié.
- 

Notions de base en statistique : La statistique

La statistique est une science qui a pour objet la collecte, l'analyse et l'interprétation des ensembles, observations relatives à un même phénomène et susceptible d'être caractérisées par un nombre »

Elle peut être descriptive ou inductive (inférentielle).

La statistique descriptive : branche des statistiques qui regroupe les nombreuses techniques utilisées pour décrire un ensemble relativement important de données.

Notions de base en statistique : La statistique

La statistique inductive (inférentielle) : Branche des statistiques permettant de tirer des conclusions sur une population à partir de l'étude d'un échantillon représentatif

Elle est fondamentale dans la recherche scientifique et procède toujours par « **induction** », c'est-à-dire par l'application du « **raisonnement inductif** ». A partir d'une étude portant sur un nombre limité de cas, on généralise les conclusions à toute la population

La statistique inductive définit les conditions sous lesquelles ces généralisations peuvent être considérées comme valables

Notions de base en statistique: Population

Population ou univers : On définit une population comme tout ensemble soumis à une étude statistique. On dit que la population est homogène lorsque les unités qui la composent sont de même nature.

L'échantillon : c'est un sous ensemble (portion) de la population. Le choix de l'échantillon peut être raisonné ou aléatoire.

L'unité statistique : c'est chaque élément constitutif de la population concernée par l'étude statistique

L'effectif : c'est le nombre d'unités composant la population ou de l'échantillon objet de l'étude.

Notions de base en statistique: Variables

Variable : Il est la propriété que l'on observe sur chaque unité statistique. C'est ce qui varie. Une variable est un facteur sujet à prendre des états différents suivant les personnes, le temps ou le lieu. Les caractères peuvent être soit quantitatifs, soit qualitatifs.

Exemple: *âge, sexe, taille, poids, etc.*

En épidémiologie ces variables sont regroupées en trois familles qui permettent de caractériser la maladie ou la santé. Il s'agit de la variable de personne, de temps et de lieux.

Notions de base en statistique: Variables quantitatives

Ce sont des variables qui sont susceptibles d'être dénombrées ou mesurées à l'aide d'instrument de mesure dont l'échelle varie selon la précision de l'instrument utilisé.

Exemple: le nombre d'enfants, la taille, le poids, la glycémie.

Les variables quantitatives sont subdivisées en variable **quantitative continue**, en variable **quantitative discrète** (isolée ou encore discontinue) et en variables temporelles. Une variable continue peut prendre plusieurs valeurs à l'intérieur d'un intervalle donné (taille, poids...).

Notions de base en statistique: Variables qualitatives

Elles désignent des variables dont les modalités ne peuvent être ni comptées ni mesurées, mais observables.

Exemple : le groupe sanguin, la nationalité, le sexe, etc.

Les variables qualitatives peuvent être **nominales** (provenance du patient) ou **ordinales** (niveau d'instruction des enquêtés).

Une autre catégorie de variable qualitative est constituée par les variables binaires ou dichotomiques (sexe).

Notions de base en statistique: Modalités d'une variable

Les modalités ou valeurs sont les manières dont le caractère peut se présenter ou encore les différentes formes que ce caractère peut prendre

Exemples:

Sexe : Masculin, féminin

Situation matrimoniale : Célibataire, marié, divorcé, veuf/veuve, ...

Méthodes de présentation des données statistiques

Il existe plusieurs méthodes de présentation des données en statistique. Les principales sont les suivantes :

- ▶ **les tableaux statistiques ;**
- ▶ **les représentations graphiques ;**
- ▶ **les cartes.**

Méthodes de présentation : Les tableaux statistiques

Il s'agit d'un ensemble de cellules constituées par l'entrecroisement de lignes et de colonnes. Un tableau statistique doit respecter certaines règles quant à sa présentation matérielle. Il doit comporter un titre, un corps et des indications complémentaires.

Le titre : doit répondre à 3 questions :

- **quoi** ? (éléments de données ou variables) ;
- **où** ? (Lieu des observations) ;
- **quand** ? (Période d'observation).

Méthodes de présentation : Les tableaux statistiques

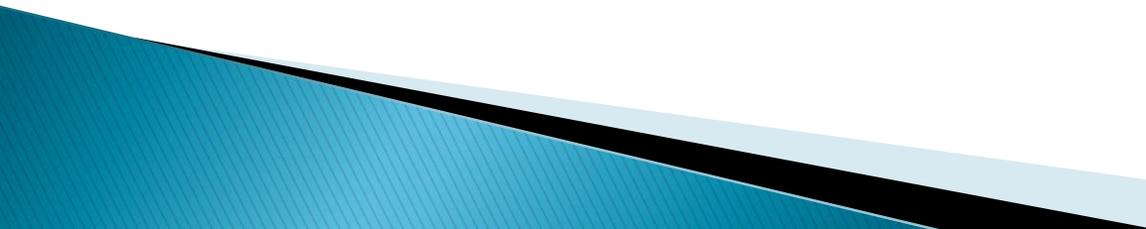
Tableau I : Situation des cas de paludisme enregistrés dans 6 FOSA de Kinshasa en 2021

Centres de santé	Nombre de cas
FOSA1	5 087
FOSA2	3 897
FOSA3	5 896
FOSA4	2 651
FOSA5	5 104
FOSA6	2 910
Total	25 545

Méthodes de présentation : Les représentations graphiques

La présentation des données sous forme de graphique est souvent plus parlante par rapport au tableau qui peut poser certaines difficultés au lecteur non habitué à utiliser des tableaux statistiques.

Le graphique, par définition, est une présentation géométrique des données statistiques dans un système d'axes. Un graphique comporte :

- ▶ **un titre ;**
 - ▶ **un corps ;**
 - ▶ **une légende ;**
 - ▶ **des indications complémentaires.**
- 

Méthodes de présentation : Les représentations graphiques

▶ Types de graphique

Il existe plusieurs types de graphiques parmi lesquels on peut citer :

- ▶ Les courbes de séries chronologiques ;
 - ▶ Les diagrammes en barres ;
 - ▶ Les histogrammes ;
 - ▶ Les diagrammes en cercles ou camembert ;
 - ▶ Les polygones de fréquence.
- 

Méthodes de présentation : Les représentations graphiques

Exemple de graphique

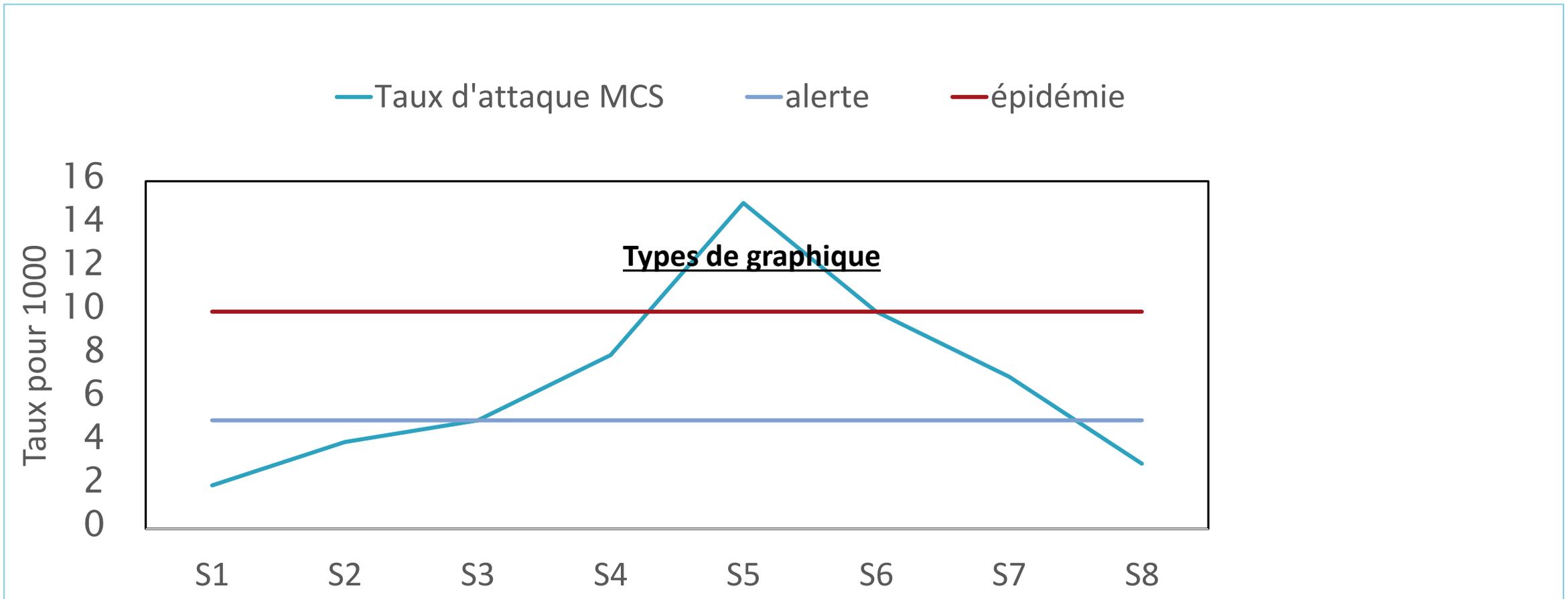


Figure 1 : Evolution du taux d'attaque MCS de S1 à S8 2022 dans le district sanitaire de Moundou

Calcul de quelques paramètres

Les paramètres sont des valeurs numériques permettant de résumer, de réduire une série statistique quantitative

On distingue deux types de paramètres les plus utilisés:

- ▶ Les paramètres de position (moyenne, médiane, le mode)
- ▶ Les paramètres de dispersion (étendue, variance, écart type, coefficient de variation)

Les paramètres de position : La moyenne arithmétique

La moyenne arithmétique est le paramètre le plus couramment utilisé en statistique. Elle sert à calculer la tendance centrale d'un ensemble de données numériques.

Elle est égale à la somme des valeurs divisée par le nombre de valeurs n :

$$\text{Moyenne} = \frac{\text{Somme des valeurs}}{\text{Nombre de valeurs}}$$

Exemple : Calculer la moyenne des tailles suivantes : 180 cm, 176 cm, 183 cm, 193 cm, 163 cm, 188 cm.

$$\text{Moyenne} = (180 + 176 + 183 + 193 + 163 + 188)/6 = 180,5 \text{ cm}$$

Les paramètres de position : Le mode

Le mode d'un ensemble de valeurs est celle qui possède la plus grande fréquence, autrement dit, le mode d'une série est la valeur (unité statistique) qui revient le plus fréquemment dans la série.

Exemple : soit la série : 8, 4, 4, 3, 4, 3, 8, 2,5

La valeur la plus fréquente dans cette série est 4, le mode est donc égal à 4.

Les paramètres de position: La médiane

La médiane est la valeur de la variable qui se trouve au milieu de la distribution quand les données sont rangées par ordre croissant ou décroissant. Autrement dit, la médiane est la valeur de la variable telle que le nombre des valeurs qui se trouvent en dessous est égal au nombre des valeurs qui se trouvent au-dessus. Pour la déterminer, on distingue le cas des données non groupées (variables discontinues) et le cas des données groupées (variables continues)

a) Médiane : cas des variables discontinues :

Exemples : Si n est impair on prend la valeur qui sépare l'ensemble en 2 parties

1-Données en nombre impair : 5 poids de nouveau-nés :

Mamadou : 3150 g

Bintou : 3200 g

Fanta : 3500 g ----- = médiane

Sékou : 3510 g

Alpha : 3720 g

Les paramètres de position: La médiane

a) Médiane : cas des variables discontinues :

Si n est pair on prend la valeur moyenne des 2 valeurs

2-Données en nombre pair : 6 poids nouveaux nés

Mamadou : 3150 g

Bintou : 3200 g

Fanta : 3500 g

Sékou : 3510 g

Alpha : 3720 g

→ = médiane

Ousmane : 3800 g

La valeur médiane se situe entre les poids de Fanta et de Sékou. Sa valeur est égale à la moyenne arithmétique des deux poids . $m=(3500+3510)/2=3505$.

Les paramètres de position: La médiane

▶ b). Médiane : cas des variables continues

Exemple: *150 malades d'un service d'après leur âge.*

La médiane est l'âge qui se situe au milieu de ces effectifs, en l'occurrence l'âge du 75ème ($150/2$) malade. En observant le tableau, on voit que ce 75ème malade se situe au niveau de la classe 35 à 39 ans et que par conséquent, l'âge médian sera compris entre 34 et 39 ans. L'âge médian est obtenu avec précision par interpolation linéaire.

Les paramètres de dispersion

▶ L'étendue

Pour un ensemble de valeurs, l'étendue est la différence entre la plus grande valeur d'une série statistique et la plus petite.

Pour les valeurs groupées, c'est la différence entre la valeur de la borne inférieure de la première classe et la valeur de la borne supérieure de la dernière classe.

▶ L'écart type et la variance

La variance est égale à la somme des carrés des écarts par rapport à la moyenne divisée par le nombre de valeurs. L'écart type est la racine carrée de la variance.

Soit $\sigma = \text{écart type}$

$$\text{▶ } \sigma = \sqrt{V(X)}$$

$$\text{▶ } V(X) = \frac{\sum_0^n n_i X_i^2}{N} - m^2 \text{ avec } \sum_0^n n_i = N \text{ et } m, \text{ la moyenne.}$$

Les paramètres de dispersion

▶ Le coefficient de variation

$$Cv = \frac{\text{Ecart type}}{\text{Moyenne}}$$

Le coefficient de variation traduit la forte ou la faible dispersion des données.

Méthode d'expression des indicateurs

- ▶ **Définition d'un indicateur** : C'est une variable qui mesure un changement d'état directement ou indirectement et qui évalue l'évolution vers les objectifs à atteindre.
- ▶ **Les qualités d'un bon indicateur** : Un bon indicateur doit être :
 - ✓ simple et compréhensible (facile à comprendre par les utilisateurs) ;
 - ✓ accessible (données disponibles) ;
 - ✓ sensible (aux variations du phénomène étudié) ;
 - ✓ pertinent (informations en relation avec le phénomène que l'on veut mesurer).
 - ✓ fiable (mesure ce qu'il est censé mesurer) ;
 - ✓ quantitatif (mesure quantitativement le phénomène étudié) ;
 - ✓ spécifique (concerne un seul phénomène et reflète des changements uniquement dans une situation précise) ;
 - ✓ reproductible (résultats obtenus identiques dans les circonstances similaires).

Méthode d'expression des indicateurs

Pour apprécier l'importance d'un phénomène dans deux localités, il faut faire une comparaison entre le nombre de cas et la population touchée. On y établit alors un rapport.

▶ Les rapports

Un rapport est l'expression la plus générale de la relation entre deux entités. Ces entités peuvent avoir un lien entre elles ou non. Un rapport peut se présenter sous forme de *proportion*, de *taux*, de *ratio* ou *d'indice*.

▶ La proportion

La proportion est un rapport dans lequel le numérateur est compris dans le dénominateur. Une proportion est une valeur qui se situe entre 0 et 1. Le résultat obtenu peut être multiplié par 100, 1000, 10000, etc.

***Exemple :** dans une classe ayant un effectif de 40 élèves dont 10 garçons, la proportion des garçons est : $10/40 = 0,25$ soit 25.*

Méthode d'expression des indicateurs

▶ Le ratio

Un ratio est un rapport dans lequel le numérateur et le dénominateur se réfèrent aux modalités d'une même variable, mais le numérateur n'est pas inclus dans le dénominateur. Le ratio peut prendre n'importe quelle valeur et n'a pas d'unité.

***Exemple** : dans une classe ayant un effectif de 40 élèves dont 10 garçons et 30 filles le ratio garçon/fille est : $10/30 = 1/3$*

Méthode d'expression des indicateurs

▶ Le taux

Le taux est un rapport qui mesure la vitesse de survenue ou de réalisation d'un événement. Le numérateur dénombre les événements qui sont survenus. Le dénominateur cumule le temps d'observation de chaque phénomène jusqu'à l'apparition de l'événement ou jusqu'à la fin de toute l'observation.

***Exemple:** sur un effectif de 300 personnes étudiées pendant 10 ans, on note :*

1 personne a eu le cancer en 7 ans ; 1 personne en 10 ans, et 2 personnes en 2 ans

*Le taux de survenue du cancer $T = \frac{4}{(1*7)+(1*10)+(2*2)+(296*10)} = \frac{4}{2981} = 0,0013$.*

Méthode d'expression des indicateurs

▶ L'indice

C'est un rapport utilisé lorsque le numérateur n'est pas compris dans le dénominateur et que l'un ou l'autre renvoie à des événements distincts.

L'indice est souvent utilisé lorsque le dénominateur n'est pas connu ou ne peut être calculé avec précision. Alors, on utilise un dénominateur qui s'approche de la réalité pour obtenir la mesure.

Exemple :

$$\text{Indice de mortalité maternelle} = \frac{\text{décès maternel}}{\text{Nombre de naissances vivantes}} \times 100\,000$$

Méthode d'expression des indicateurs

▶ L'incidence

L'incidence mesure la fréquence d'apparition de nouveaux cas d'une maladie ou d'un évènement au sein d'une population au cours d'une période donnée.

Taux d'incidence est la vitesse de survenue d'un évènement au cours d'une période donnée.

$$\text{Incidence} = \frac{\text{Nombre de nouveaux cas de la période}}{\text{Effectif de la population au cours de la période}} \times k$$

▶ Incidence cumulée

Il évalue le risque encouru par une personne de la population étudiée, pendant la période d'observation. C'est le nombre de nouveaux cas pendant une période donnée rapporté à la population à risque en début de la période d'observation.

La période peut être l'année, le mois, la semaine ou le jour etc.

Méthode d'expression des indicateurs

▶ La prévalence

La prévalence désigne le nombre de cas (nouveaux et anciens) existant à un moment donné. C'est une photographie instantanée d'un phénomène au sein d'une population en un temps « t ».

$$\text{Prévalence} = \frac{\text{Nombre total de cas à un moment donné}}{\text{Population totale}} \times 1000$$

▶ Le taux d'attaque

Le taux d'attaque est une forme particulière de taux d'incidence utilisé en cas d'épidémie. Il s'applique à des populations observées pendant une durée limitée, par exemple une population au sein de laquelle sévit une épidémie.

$$\text{Taux d'attaque} = \frac{\text{Nombre de nouveaux cas d'une maladie}}{\text{Effectif de la population exposée}} \times k$$

Méthode d'expression des indicateurs

▶ Distribution de fréquences

Une distribution de fréquence donne la ventilation du nombre total d'événements entre les diverses modalités ou classes étudiées. Les effectifs partiels (fréquences absolues), rapportés à l'effectif total, fournissent les fréquences relatives généralement exprimées en pourcentage.

$$\text{Fréquence} = \frac{\text{Effectif d'une modalité}}{\text{Effectif total de la série}} \times 100$$

▶ La létalité

La létalité mesure le risque de décéder par suite d'une maladie ou d'un phénomène donné. Elle peut être calculée pour l'ensemble des cas, ou spécifiquement dans certains sous-groupes de cas.

Exemple : dans une population de 968 718 malades de Sida on a enregistré 1 356 décès.

La létalité du VIH/SIDA est égale à: $1\ 356 * 100 / 968\ 718 = 1,4\%$.

Méthode d'expression des indicateurs

▶ Les taux de mortalité

Un taux de mortalité mesure la fréquence des décès survenus dans une population déterminée au cours d'une période spécifiée. Il existe plusieurs types de taux de mortalité :

Les taux bruts tiennent compte de la totalité des décès dans l'ensemble des groupes d'âges et quelle qu'en soit la cause.

Les taux spécifiques :

- ✓ les taux par cause se rapportent à une cause déterminée ;
- ✓ les taux par âge et/ou sexe tiennent uniquement compte des décès intervenus dans un groupe d'âge et/ou sexe spécifié.

MERCI POUR VOTRE ATTENTION