

INSTITUT DE FORMATION ET DE RECHERCHE
INTERDISCIPLINAIRES EN SCIENCES DE LA SANTÉ ET DE
L'ÉDUCATION
(IFRISSE)



Généralités en biostatistique

ILBOUDO Wendyam Fulbert
Statisticien démographe

Ce cours est adapté de celui de Ter Tiero Elias DAH MD, PhD

Plan

1. Introduction
2. Définitions et champs d'application de la statistique
3. Place de la statistique en épidémiologie
4. Concepts et notions courantes utilisés en biostatistique

Introduction

Pourquoi des statistiques en santé publique, en médecine ?

En santé publique : les statistiques servent à décrire les phénomènes morbides à partir de données diverses telles que les chiffres de mortalité, les informations issues des organismes de soins, les déclarations d'accidents, etc.

Elles sont aussi utilisées pour estimer les facteurs de risque d'une population et pour évaluer les succès diagnostiques, thérapeutiques et de prévention.

Elles permettent également d'établir l'incidence d'une maladie (nombre de nouveaux cas par an, dans une population déterminée)

Définition (1)

Statistique

- la discipline qui étudie des phénomènes à travers le **recueil ou la collecte de données**, leur **traitement** et **leur analyse**, l'**interprétation** des résultats et leur présentation afin de rendre ces données compréhensibles par tous. C'est à la fois une branche des mathématiques appliquées, une méthode et un ensemble de techniques. (Wkp)
- NB: Ne pas confondre la **S**tatistique (majuscule) et une **s**tatistique (minuscule) qui est le résultat d'un calcul obtenu à partir d'observations.
- Science de la **variabilité**
 - Permet la prise en compte des variations de mesure d'un phénomène

Traitement & analyse des données: statistique
descriptive

Interprétation des données: statistique inférentielle

Définition (2)

Biostatistique

- Application de la statistique (traitement statistique) à des données biologiques. Il s'agit d'un champ interdisciplinaire qui associe la biologie, l'informatique et la statistique.

NB: La biologie est caractérisée par une extrême variabilité

- Sources de variabilité :
 - Outil ou instrument de mesure (toise, balance, questionnaire de recueil de données)
 - Différence entre sujets (intra ou inter) pour une caractéristique donnée
 - Ex: la taille des étudiants de Master I en santé publique à l'IFRISSE
 - Le taux d'hémoglobine des femmes enceintes qui consultent pour leur CPN
 - Les différents symptômes de la Covid-19 chez les personnes infectées
- La variabilité pose plusieurs questions:
 - Sa compréhension. Quelle est la taille moyenne des étudiants de Master I en santé publique à l'IFRISSE?
 - Statistique descriptive
 - Son interprétation. Le taux d'hémoglobine des femmes enceintes et qui consultent pour leur CPN est-il plus bas que celui des femmes non enceintes?
 - Statistique inférentielle
 - Sa prédiction . Quel(s) symptôme(s) de la Covid-19 chez les africains est susceptible de conduire au décès?
 - Modélisation statistique

Champs d'applications

Très variés

- Sciences biologiques et médicales
- Démographie
- Sociologie
- Marketing
- Economie
- Physique
- Géophysique (météorologie, climatologie)
- Etc,

Certaines spécialités de la statistique

- Biostatistique
- Econométrie
- Actuariat
- Etc,

Place de la statistique en épidémiologie

Epidémiologie

- Discipline à cheval entre d'autres disciplines
 - Médecine, biologie humaine
 - Démographie
 - Sciences sociales
 - Statistique

La statistique dans l'épidémiologie (biostatistique) est l'outil qui permet de:

- Décrire les populations humaines (personnes, temps, lieu) et les phénomènes de santé
- D'établir le lien et la force de l'association entre exposition et événements d'intérêt (mesures d'association)
- Calculer les mesures d'impact des interventions de santé
- Modéliser l'évolution de phénomènes de santé dans un objectif d'explication ou de prédiction

Il n'y a pas d'épidémiologie sans biostatistique.

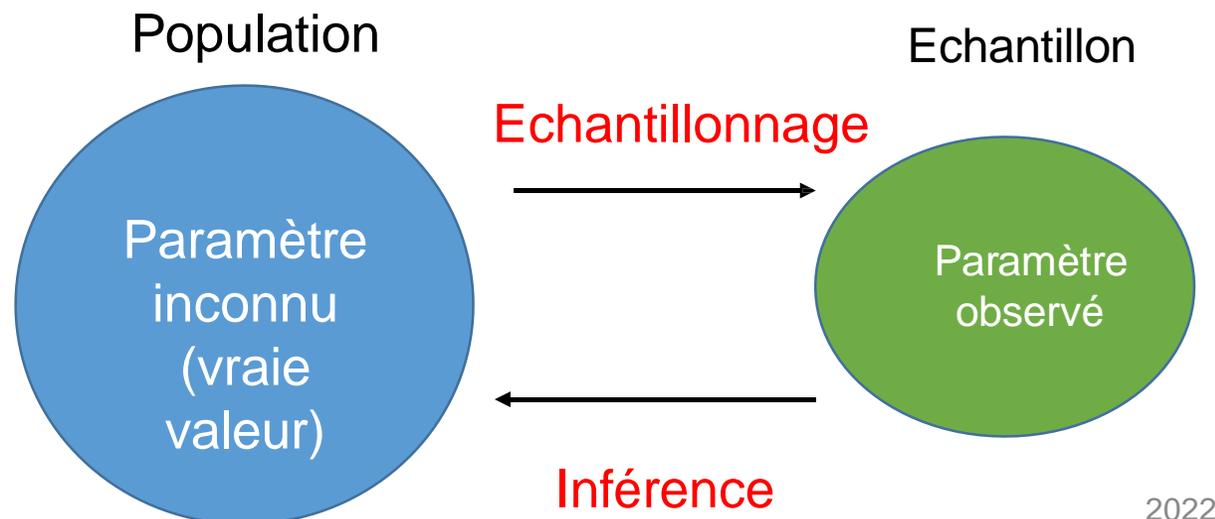
Concepts & vocabulaires courants en biostatistique (2)

Echantillonnage

- Processus par lequel on choisit ou extrait un échantillon d'une population source. Ce processus peut découler d'une méthode aléatoire, l'échantillonnage est dit aléatoire ou probabiliste. Ou alors, le processus n'est pas dû au hasard, l'échantillonnage est dit non probabiliste.

Estimation

- Tenter de définir les paramètres d'une population (inconnus) à partir des paramètres observés sur un échantillon. Il peut s'agir d'une grandeur quantitative (moyenne, médiane, écart-type) ou d'une grandeur qualitative (proportion).
- L'estimation est ponctuelle, et par intervalles.



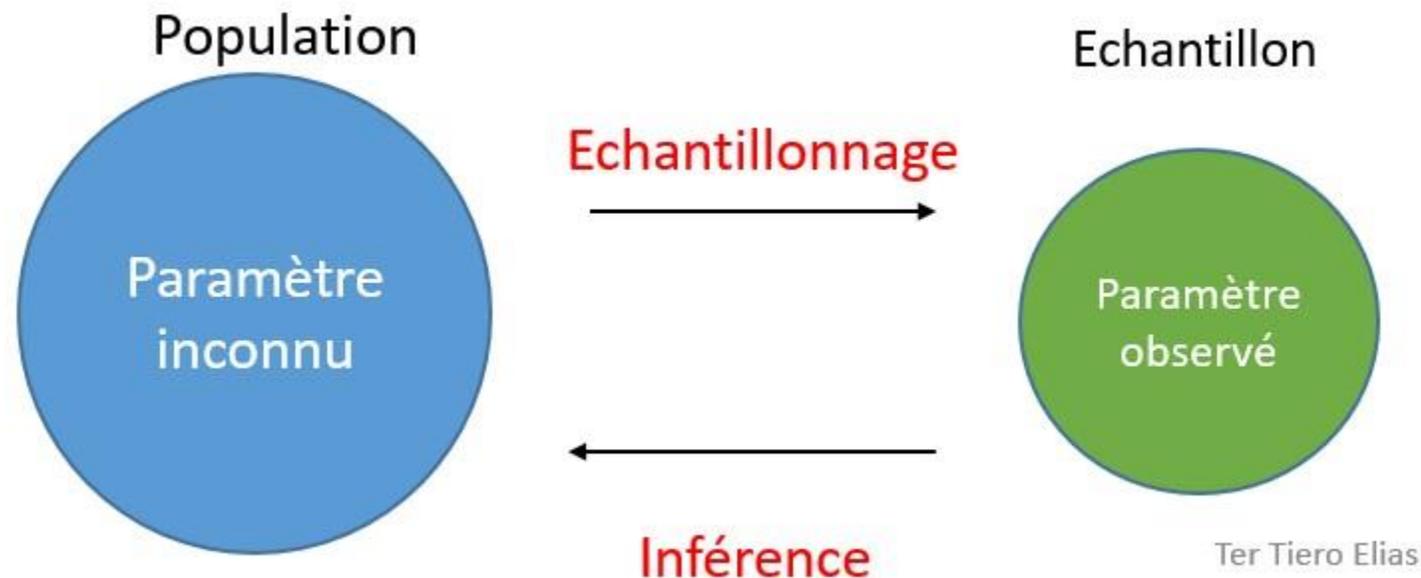
Notations

	Population P (valeur théorique)	Echantillon E (v. observée)
Variable quantitative		
Moyenne	M ou μ	m
Variance	S^2 ou σ^2	s^2
Ecart-type	S ou σ	s
Variable qualitative		
Fréquence	P ou π	p

Concepts & vocabulaires courants en biostatistique (3)

Inférence ou inférence statistique ou statistique inférentielle

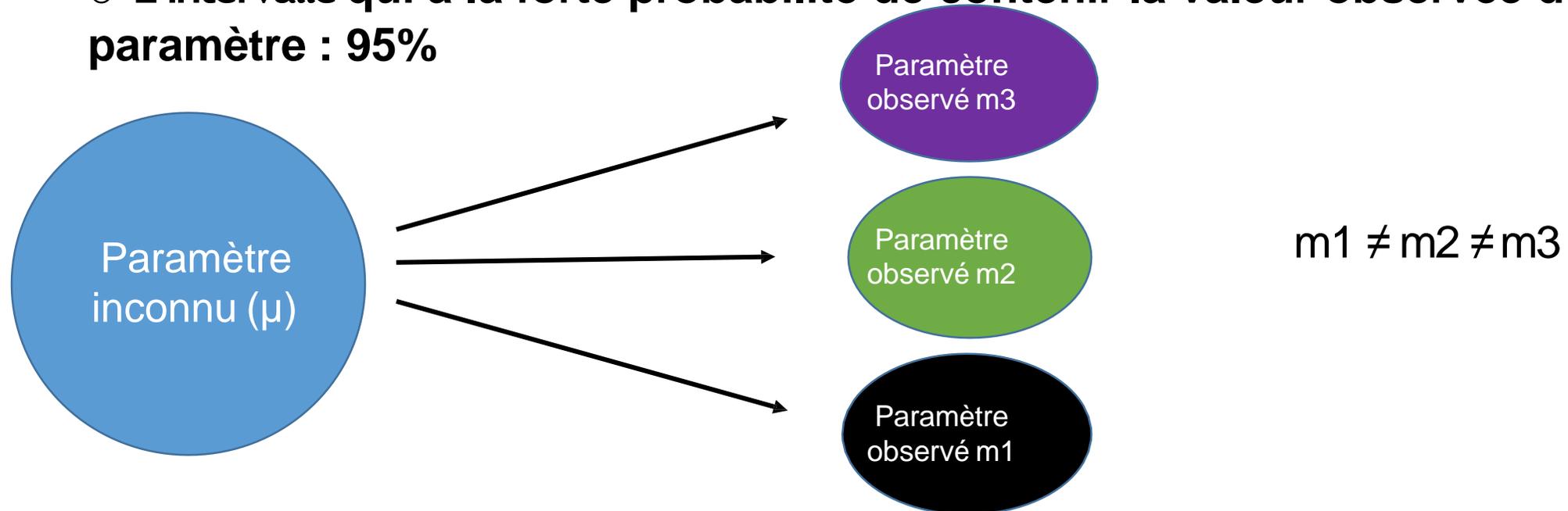
- La procédure qui consiste à utiliser les informations recueillies dans l'échantillon pour déduire des résultats concernant l'ensemble de la population.
- L'inférence statistique est basée sur les paramètres statistiques utilisés en statistique descriptive pour décrire une population (moyenne, écart-type, fréquence,...) et des calculs de probabilités.



Concepts & vocabulaires courants en biostatistique (4)

Fluctuations d'échantillonnage

- Si l'on effectue plusieurs échantillonnages pour estimer un paramètre inconnu on trouvera dans chaque échantillon sélectionné, un paramètre observé sensiblement différent.
- Implications
 - Les mesures d'estimation doivent tenir compte de cette fluctuation : **intervalle de confiance**
des valeurs ponctuelles observées
 - **L'intervalle qui a la forte probabilité de contenir la valeur observée du paramètre : 95%**



Concepts & vocabulaires courants en biostatistique (5)

Représentativité

- L'échantillon E est représentatif de la population P dont il est issu veut dire que sa composition est conforme à celle de P
- 2 conditions sont nécessaires pour cela. L'échantillon doit:
 - Avoir un effectif suffisamment grand
 - Etre extrait au hasard de P, c'est-à-dire tiré au sort.

Variable

- Une propriété d'un objet ou d'un évènement qui peut prendre différentes **modalités**. (Ex: le sexe est une variable. Il peut prendre les modalités femme ou homme. Nombre d'enfants par cour est une variable qui peut prendre les modalités 1,2,3,...10)

Caractérisations ou appellations d'une variable

Objet principal de l'étude	Objets secondaires
v. principale v. d'intérêt v. dépendante	v. secondaire v. exposition v. indépendante

Ex: On veut savoir la prévalence de l'HTA dans une population. En plus de mesurer la PA, on collecte également des données sur l'âge, le poids, les habitudes de vie.

- V. principale : la pression artérielle
- V. secondaires : âge, poids, habitudes de vie

Variable aléatoire

- Une variable dont la valeur est déterminée en fonction du résultat d'une expérience aléatoire. Elle permet d'associer une valeur à chaque sujet de population. On distingue
 - Variable aléatoire discrète: prend un ensemble fini ou infini **dénombrable** de valeurs (ex: sexe, statut matrimonial, nombre d'enfants dans une famille)
 - Variable aléatoire continue: prend un ensemble infini **non dénombrable** de valeurs (ex: taille, poids, pression artérielle)

Distribution d'une variable

- Répartition des différentes valeurs prises (**ou observations**) par cette variable dans une population ou un échantillon.
 - Ex: la distribution de la variable sexe des étudiants de Master I en santé publique à l'IFRISSE est la répartition du nombre total d'étudiants selon les modalités H/F.
- La distribution peut être résumée
 - A partir des paramètres de description :
 - Variable quantitative
 - Position ou de tendance centrale (moyenne, médiane, mode)
 - Dispersion (étendue, variance/écart-type, percentiles)
 - Variable qualitative : fréquence et intervalle de confiance
 - Dans un texte, un tableau, un graphique ou figure

Probabilité d'un évènement

- la proportion de fois où cet évènement se produirait si on répétait à l'infini les conditions où il peut se produire

Loi de probabilité (ou distribution de probabilité) d'une variable aléatoire

- Indique la donnée de la probabilité de survenue de chaque évènement possible
- A chaque variable aléatoire, l'on peut associer des paramètres ou nombres (moyenne, variance) qui permettent de la résumer.

Concepts & vocabulaires courants en biostatistique (9)

Loi de probabilité d'une variable aléatoire discrète

- Si X est discrète, elle prend un nombre dénombrable de valeurs x_i et on note p_i la probabilité que X soit égale à x_i , $p_i = P_r[X=x_i]$
- La somme (Σ) des valeurs p_i sur l'ensemble des valeurs x_i est $= 1$
- La loi de probabilité correspond à l'ensemble des couples $(x_i; p_i)$

Loi de probabilité d'une variable aléatoire continue

- Ensemble des valeurs est infini
- Considérer que les probabilités que X appartiennent à des intervalles de nombres réels : $]-\infty, x]$, $]x, +\infty[$ ou $[x_1, x_2]$
- La définition d'une variable aléatoire continue repose sur la notion de fonction de répartition ou fonction de densité de probabilité $F(X) = P_r[X < x]$

Documents ressources

1. François Dabis, Jean Claude Desenclos. Epidémiologie de terrain. Méthodes et applications 2017.
2. Jean Bouyer. Méthodes statistiques. Médecine-Biologie. 2017.
3. Thierry Ancelle. Statistique-Epidémiologie, 4^{ème} édition. 2017.

Merci pour votre attention