

PRÉPA PASS & LAS

**MÉTHODE
COURS
EXERCICES**

SE PRÉPARER À RÉUSSIR DÈS LE LYCÉE

**BIOSTATISTIQUES
BIOLOGIE CELLULAIRE
PHARMACOLOGIE**

Théo Cossart (coord.)
Titouan Riffard, Valentin Mockelyn, Lucas Dessaux

ellipses

INTRODUCTION À LA BIOSTATISTIQUE

I LES BIOSTATISTIQUES

Les **biostatistiques** sont l'utilisation des statistiques pour la science médicale.

Elles ont une place prépondérante dans les sciences médicales, car elles permettent de décrire les phénomènes de santé dans une population ou bien de chercher des effets de facteurs sur la santé.

Les biostatistiques sont donc une spécialité mathématique. Son enseignement, souvent effectué par des professeurs de santé publique à l'aise avec ses notions, peut paraître abrupt et donne une image de complexité à cette matière.

Pour autant, les biostatistiques ne demandent pas de notions mathématiques complexes. **L'ensemble des notions nécessaires à la réussite dans cette matière sont vues pendant le lycée.** Les biostatistiques demandent surtout un apprentissage des méthodes et une compréhension des objectifs des différents tests réalisés.

Les biostatistiques sont formées par trois grandes catégories. Chacune de ses catégories permet de répondre un objectif précis :

- Les **statistiques descriptives** : Elles sont utilisées dans le but de **décrire** l'état de santé une population ou les performances d'un test.
- Les **statistiques analytiques** : Elles sont utilisées dans le but de **comparer** l'état de santé de plusieurs groupes ou les performances de tests différents.
- L'**épidémiologie** : Elle est utilisée dans le but de **trouver** des potentiels risques ou protecteurs de santé, pour ensuite agir pour améliorer la santé des populations.

II L'APPRENTISSAGE DES BIOSTATISTIQUES

Les biostatistiques nécessitent un apprentissage différent des matières de connaissance. En effet, il est demandé de savoir appliquer un raisonnement pour répondre à la question posée, plutôt que de savoir ressortir des connaissances par cœur.

Il est donc nécessaire de **comprendre le raisonnement** et savoir appliquer celui-ci.

C'est pour cela que l'entraînement en biostatistique passe par la **réalisation d'exercices régulièrement**, pour acquérir la méthode et la conserver. S'exercer régulièrement permet aussi de trouver son rythme pour avoir le temps le jour du concours.

Le jour du concours, vous aurez avec vous un formulaire, un livret regroupant l'ensemble des formules qui vous seront expliquées pendant le semestre.

C'est pour cela que **les formules ne sont pas à apprendre** mais à comprendre et savoir utiliser dans le cadre des exercices.

Enfin, les notions de cours peuvent aussi faire l'objet de question le jour du concours. Prenez donc le temps d'**apprendre les notions** qui semblent importantes pour le jour J !

III CE QUE VOUS ALLEZ TROUVER DANS CE LIVRE

L'ensemble des chapitres de biostatistiques sont construits sur le même modèle pour vous en faciliter la prise en main et la compréhension.

Pour commencer, vous trouverez tout d'abord une partie d'apprentissage.

Cette partie se déroule en deux temps : une explication des notions de cours, suivi immédiatement par une application directe des notions par un exercice commun à l'ensemble du chapitre.

L'exercice explicatif des notions se distingue des cours par son écriture en italique

Une fois la partie d'apprentissage terminée, vous retrouverez un exercice pour vous entraîner, reprenant une grande partie des notions vues dans le chapitre.

La correction de ce chapitre vous sera guidée pas à pas pour en comprendre le raisonnement.

Enfin, vous trouverez de nombreux schémas explicatifs pour rendre plus visuel les différentes notions abordées.

IV UN DERNIER MOT AVANT DE COMMENCER

Les biostatistiques sont une matière souvent délaissée par les étudiants.

Avoir une base solide et les bonnes méthodes vous permettra de vous en tirer dans les exercices classiques, vous permettant de faire la différence !

Je vous souhaite donc une bonne lecture et un bon apprentissage du merveilleux monde des biostatistiques !

LES STATISTIQUES DESCRIPTIVES

I INTRODUCTION ET DÉFINITIONS

1 INTRODUCTION

Les statistiques descriptives correspondent à l'utilisation des statistiques pour représenter de la manière la plus fidèle et pertinente possible l'organisation et la structure des données étudiées.

Cette branche des statistiques étudie des échantillons **uniques**. Elle n'effectue donc pas de comparaison.

2 DÉFINITIONS

Variable : C'est le nom donné à la donnée que l'on étudie.

Population générale : Elle représente l'ensemble des personnes que l'on veut étudier.



EXEMPLE

La population Française pour une étude en France, la population d'une ville pour une étude sur une ville particulière.

Cette population étant trop nombreuse pour être étudiée entièrement, on utilise des échantillons de celle-ci pour faciliter les recherches sur la variable.

Échantillon : C'est un petit groupe de personnes que l'on recrute au sein d'une population et au sein duquel on va étudier la variable d'intérêt.

Effectif : L'effectif représente le nombre de personnes incluses au sein de l'échantillon.

II L'ÉCHANTILLONNAGE

1 INTÉRÊT DES ÉCHANTILLONS

Les statistiques se veulent le reflet le plus précis possible de la population étudiée. Malheureusement, cette population est souvent trop grande pour que l'on puisse étudier notre variable d'intérêt chez toutes les personnes.

Pour rendre cette tâche possible, on construit à partir de notre population un échantillon de personnes chez qui on étudiera la variable d'intérêt. **L'échantillon** est donc un petit groupe de personnes d'une population. On peut construire plusieurs échantillons de personnes au sein d'une même population.

2 VARIABILITÉ D'ÉCHANTILLONNAGE

Si l'on construit deux échantillons différents au sein d'une même population, les personnes présentes au sein de cet échantillon ne seront pas identiques. On retrouvera donc des valeurs différentes pour la variable d'intérêt dans chaque échantillon. C'est ce qu'on appelle la **variabilité d'échantillonnage**.

Cette variabilité dépend en grande partie d'un seul facteur : le nombre de personnes au sein de l'échantillon, appelé **effectif** de l'échantillon. En effet, plus un échantillon est grand, plus la valeur de la variable d'intérêt sera proche de la réalité et moins elle variera d'un échantillon à un autre.

III LES DIFFÉRENTS TYPES DE VARIABLES

1 GÉNÉRALITÉ

- Il existe différents types de variables.
On distingue tout d'abord les **variables quantitatives** et les **variables qualitatives**.

Les variables quantitatives sont des variables que l'on peut **mesurer et quantifier par un nombre**. À l'inverse, les variables qualitatives sont des variables que l'on ne peut pas désigner par des nombres. On les **désigne alors par des noms**.

2 VARIABLES QUANTITATIVES

Au sein des **variables quantitatives**, on retrouve deux sous-types :

- Les variables dites discrètes qui ne peuvent prendre qu'un **nombre limité de valeurs**.



EXEMPLE

Le nombre d'enfants. On peut avoir 1 ; 2 ; 3 enfants, mais pas 2,4 ou 1,5 enfant.

- Les variables dites continues qui peuvent prendre **l'ensemble des valeurs** au sein d'un intervalle.



EXEMPLE

La taille. On peut faire n'importe quelle taille entre 25 cm et 250 cm, avec par exemple 55 cm ; 155,2 cm ou 178,8 cm.

3 VARIABLES QUALITATIVES

Au sein des **variables qualitatives**, on retrouve aussi deux sous-types :

- Elles sont dites nominales lorsqu'elles ne peuvent **pas être classées** entre elles dans un ordre logique.



EXEMPLE

La couleur des yeux. On peut avoir les yeux bleus, marrons ou verts. Mais on ne peut pas classer ces couleurs dans un ordre logique.

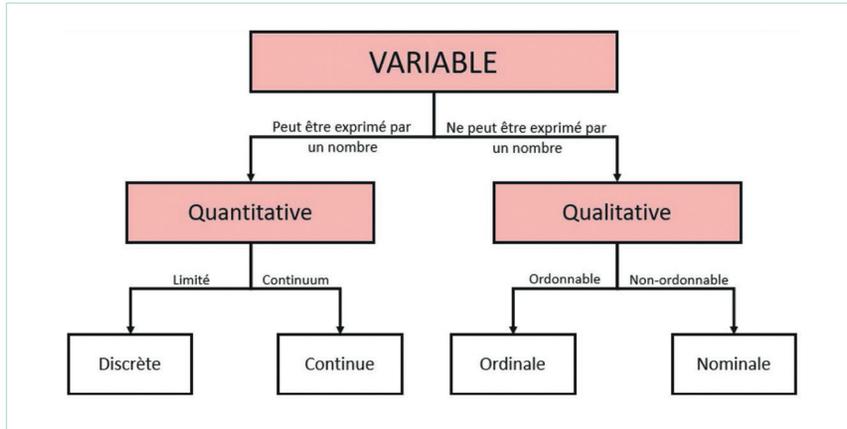
- Elles sont dites ordinales lorsqu'elles **peuvent être classées** dans un ordre logique.



EXEMPLE

La température ressentie. On peut trouver une pièce chaude, froide ou tiède. On peut classer ses variables dans un ordre logique, du plus froid au plus chaud par exemple.

4 SCHEMA RÉCAPITULATIF



IV LES PARAMÈTRES DESCRIPTIFS

1 LES 2 TYPES DE PARAMÈTRES DESCRIPTIFS

- Ce sont des paramètres mathématiques permettant de décrire un échantillon donné. Ils sont de plusieurs types :
 - Les **paramètres de position**, qui servent à décrire notre échantillon. On retrouve parmi eux : la Moyenne, la Médiane et les Quantiles.
 - Les **paramètres de dispersion**, qui servent à renseigner sur l'étalement des valeurs de notre échantillon. On retrouve parmi eux : l'Amplitude, l'Écart interquartile, la Variance, l'Écart type et le Coefficient de variation.



EXEMPLE

Tout au long de cette partie, nous allons illustrer le propos par un exemple commun.

On veut connaître la taille moyenne de la population française. Pour cela, on recrute un échantillon de 7 personnes que l'on mesure.

On mesure dans notre échantillon les tailles suivantes : 180,5 ; 155,4 ; 168,0 ; 143,8 ; 193,0 ; 161,7 ; 168,0 ; 189,4 ; 158,3.

On arrondira l'ensemble des résultats à 10^{-1} .

2 PARAMÈTRES DE POSITION

A. MOYENNE

Elle représente le **centre de gravité** de la distribution.

Elle est notée \bar{x} , appelé x barre.

Elle se calcule en faisant la somme de toutes les valeurs que l'on divise par le nombre de valeurs dans l'échantillon.

Elle se calcule avec la formule $\bar{x} = \frac{\sum x_i}{n}$.

Avec $\sum x_i$ la somme des valeurs de l'échantillon et n le nombre de valeurs dans l'échantillon.

C'est un paramètre assez représentatif mais très sensible aux valeurs extrêmes.

On peut seulement la calculer pour des variables quantitatives.



EXEMPLE

On veut calculer la moyenne de l'échantillon.

On prend toutes les valeurs, on en fait la somme et on divise par le nombre de valeurs :

$$\begin{aligned}\bar{x} &= \frac{\sum x}{n} \\ &= \frac{180,5 + 155,4 + 168,0 + 143,8 + 193,0 + 161,7 + 168,0 + 189,4 + 158,3}{9} \\ &= \frac{1518,1}{9} = 168,7\end{aligned}$$

B. MÉDIANE

Elle représente la **valeur pour laquelle la moitié de l'échantillon** lui est inférieure et l'autre moitié supérieure.

Elle est notée \tilde{x} , lu x tilde.

C'est un paramètre robuste, puisque peu sensible aux valeurs extrêmes.

Pour trouver la Médiane, on applique la méthode suivante :

Étape 1 : On commence par classer les valeurs dans un ordre croissant

Étape 2 : On cherche le rang de la valeur à choisir :

Il se calcule en fonction du nombre de valeurs dans l'échantillon.

- S'il est impair : On divise n par deux et on prend la valeur au-dessus.
- S'il est pair : On divise n par deux et on prend la valeur obtenue.