



# Méthode statistique en médecine



# Introduction et définitions

Biostatistiques : statistiques appliquées au domaine de la santé publique

→ But : décider si une observation est due au hasard ou explicable

→ 3 objectifs : description, évaluation et mise en place des observations



# Introduction et définitions

Statistiques : art de collecter, analyser et interpréter des données

2 types de statistiques :

- Descriptives : description grâce à des paramètres
- Déductives : observation due au hasard ?

# Introduction et définitions

Données : résultats d'une observation

→ N'est intéressante que si on l'observe ou la compare à d'autres individus

→ On parle alors de variable (inter ou intra sujet)

# Introduction et définitions

Paramètre : grandeur apportant une information résumée sur la variable étudiée (*ex : moyenne, médiane...*)

Série statistique : collection d'objets de même nature avec des caractéristiques différentes (*ex : étudiants en LAS à Nice*)

Population : série exhaustive de tous les individus étudiés (*ex : population française*)

Echantillon : sous-ensemble fini et d'effectif limité (*ex : 100 LAS tirés au sort*)

# Point important +++

Echantillon : connu

Population : inconnue

La population étant inaccessible dans son entièreté,  
on réalise l'étude sur un échantillon connu.

# Variables

| Quantitative  | Qualitative   |
|---|---|
| <p>Mesurée ou dénombrée</p> <p>Peut être :</p> <ul style="list-style-type: none"><li>• Discrète (sans virgule → <i>âge</i>)</li><li>• Continue (avec virgule → <i>glycémie</i>)</li></ul> | <p>Non mesurable</p> <p>Peut être :</p> <ul style="list-style-type: none"><li>• Binaire (<i>homme/femme</i>)</li><li>• Nominale (<i>couleur des cheveux</i>)</li><li>• Ordinale (<i>échelle de douleur, stade d'une maladie</i>)<br/>→ pièges</li></ul> |

# Variable quantitative ordinaire

Une variable qualitative ordinaire peut être approximée en une variable pseudo quantitative

*Une échelle de douleur, on classe de 1 à 10 (d'où le "pseudo quantitatif"), mais cela reste des scores subjectifs → non mesurés*

→ UNE VARIABLE PSEUDO QUANTITATIVE RESTE QUALITATIVE +++

# Paramètres

## I. Moyenne

→ Pour une variable quantitative discrète

→ Pour une variable quantitative continue

$$m = \frac{\sum_{i=1}^n x_i}{n}$$

$$m = \frac{\sum_{i=1}^n n_i x_i}{n}$$

2. Variance : Indique la dispersion des valeurs autour de la moyenne

# Paramètres

3. Médiane : Valeur centrale si données rangées dans un ordre croissant

*Ex : {3,4,6,8,10} → La médiane est 6.*

Dit autrement, avec  $n$  l'effectif étudié, la médiane sera :

Si  $n$  pair : moyenne des  $\frac{n}{2}$  et  $\frac{n+1}{2}$  valeurs

Si  $n$  impair :  $\frac{n+1}{2}$  eme valeur

# Paramètres

4. Quartiles : partagent la série ordonnée en 4 groupes de même effectif

*Ex : Premier quartile  $\rightarrow$  25% de la série est inférieur à cette valeur*

|         | Avantages   | Inconvénients   |
|---------|---|---|
| Moyenne | <ul style="list-style-type: none"><li>• Simple à calculer</li><li>• Facile à manipuler dans des tests statistiques donc adaptée aux calculs statistiques</li><li>• Très significative si la répartition des données est assez symétrique avec une faible dispersion</li></ul> | <ul style="list-style-type: none"><li>• Sensible aux valeurs anormales (maximum et minimum)</li></ul> |
| Médiane | <ul style="list-style-type: none"><li>• Calcul facile</li><li>• Peu sensible aux valeurs anormales</li><li>• Utilisable pour des valeurs ordinales, des classes</li></ul>   | <ul style="list-style-type: none"><li>• Se prête moins aux calculs statistiques</li></ul>             |

# Estimation d'un paramètre au niveau d'une population

2 types d'estimation :

- Ponctuelle : valeur, jugée meilleure à un instant  $t$  (peu fiable)
- Par intervalle : Intervalle de valeurs contenant la valeur recherchée → Intervalle de confiance (IC)

→ L'estimation par intervalle est moins précise mais plus fiable +++

# Méthodologie pour estimer des données quantitatives

1. Détermination précise de la population étudiée (=population cible)
2. Tirage au sort (TAS) d'un échantillon représentatif (n sujets)
3. Calcul de l'intervalle de confiance

# Définitions (encore)

Ecart-type : Mesure la dispersion d'un ensemble de données autour de la moyenne +++

Degré de liberté / DDL : Le nombre de valeurs nécessaires à connaître pour résoudre une équation et connaître toutes les valeurs de la série (voir stats déductives)

# Intervalle de confiance

Intervalle de confiance : Estimation de la moyenne vraie  $\mu$  à partir de la moyenne  $m$  calculée sur l'échantillon

$$\mu \in \left[ m \pm \frac{\varepsilon s}{\sqrt{n}} \right]$$

Aussi appelée intervalle au risque  $\alpha$   
L'IC correspond à ce qui est entre les crochets dans la formule

# Risque $\alpha$

C'est la probabilité de se tromper dans l'estimation de  $\mu$  (donc que  $\mu$  ne se trouve pas dans notre intervalle de confiance)

Le plus souvent,  $\alpha = 5\%$

# Ecart réduit $\varepsilon$

C'est une valeur qui dépend du risque  $\alpha$ , ils varient en sens inverse

Pour  $\alpha = 5\%$   $\varepsilon = 1,96$

Pour  $\alpha = 1\%$   $\varepsilon = 2,60$

Un écart réduit mesure de combien d'écart-types une observation particulière est éloignée de la population

# Précision de l'estimation

Le risque  $\alpha$  conditionne la précision de l'estimation. Si on prend moins de risque, on a un intervalle de confiance plus grand, et donc plus de chances que la moyenne soit dedans.

| IC Large   | IC Resserré   |
|--|---|
| Si $\alpha \searrow$ alors $\epsilon \nearrow$ donc l'IC $\nearrow$  | Si $\alpha \nearrow$ alors $\epsilon \searrow$ donc l'IC $\searrow$   |
| <ul style="list-style-type: none"> <li>→ On a plus de chances que <math>\mu</math> soit comprise dans l'IC</li> <li>→ Par contre on perd en précision</li> </ul> | <ul style="list-style-type: none"> <li>→ On a moins de chance que <math>\mu</math> soit dans l'IC</li> <li>→ Mais on diminue l'IC, on gagne en précision</li> </ul> |



# Indice de précision $i$

Il permet de calculer la précision de l'estimation de  $\mu$ .  
Cette valeur représente la largeur de l'IC.

$$i = \frac{\varepsilon S}{\sqrt{n}}$$

Donc  $\mu = [m \pm i]$

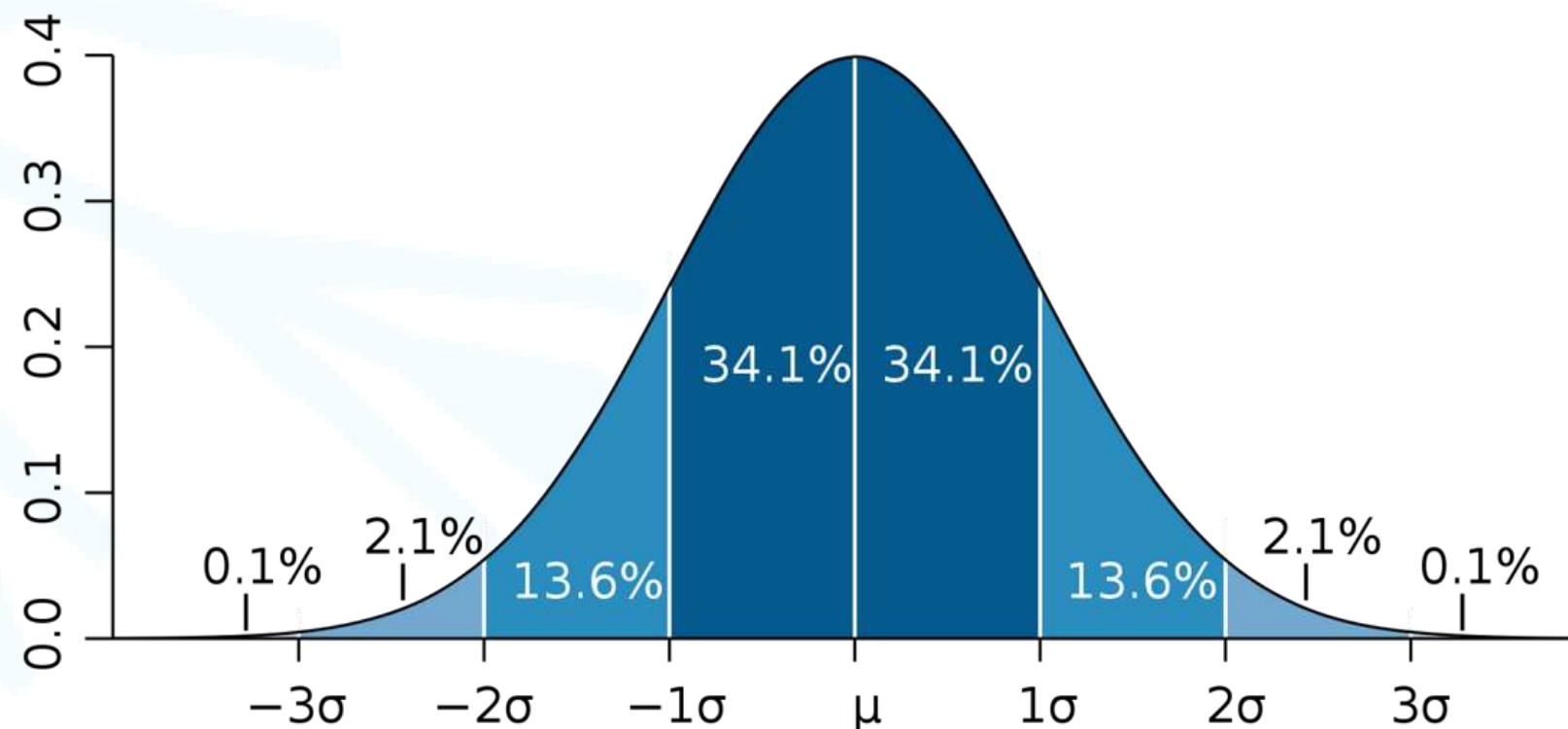
Si  $n \nearrow$ ,  $i \searrow$  donc l'IC  $\searrow$  donc la précision  $\nearrow$

# Loi de Gauss ou loi normale

Distribution en forme de cloche plutôt symétrique par rapport à la moyenne avec

- l'IC en abscisse
- n en ordonnée

L'aire sous la courbe correspond au % de la population concernée



# Sondage

Le sondage est une application directe de l'IC calculée sur des données qualitatives. Tout résultat d'un sondage doit être accompagné d'un IC.

Pour une bonne estimation il faut :

- Un échantillon constitué par TAS
- Pas de biais de sélection
- Un IC qui accompagne l'estimation
- Une taille importante de l'échantillon

# Introduction et définitions

Biostatistiques : statistiques appliquées au domaine de la santé publique

→ But : décider si une observation est due au hasard ou explicable

→ 3 objectifs : description, évaluation et mise en place des observations

FIN