Méthodes statistiques

Biostatistiques - camiléon

Introduction

Biostatistiques = stats appliquées au domaine de la SP

3 objectifs :

- Description pop
- Évaluation ttt, techniques, coûts
- Mise en place observations épidémiologiques

Biostats → décider si observation explicable ou due au hasard?

Statistiques

= art de collecter, analyser et interpréter des données

Biostatistiques

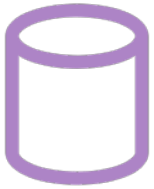
stats appliquées à la médecine

2 types de biostatistiques

- **<u>Descriptives</u>** = on décrit une situation avec <u>paramètre</u> (taille, âge pop française)
- **<u>Déductives</u>** = observation due au <u>hasard</u> ? (personnes 1m65 brunes, hasard?)

Données

= résultat de l'observation d'un individu, grâce à un instrument de mesure / sens



Définitions

- But d'une donnée → **observer** / **comparer** plusieurs individus
- Une donnée est une variable

VARIABLE PREND UNE VALEUR POUR UN INDIVIDU

- **Grande variabilité** due au hasard ou physiologique :
  - Inter sujet (entre 2 sujets)
  - Intra sujet (1 sujet)

Paramètres

Grandeur apportant une **information résumée/synthétisée** sur la variable étudiée

→ Moyenne, écart-type

Série statistique

Collection d’objets de **même nature** mais de **caractéristiques différentes**

→ Hommes et femmes

Variable quantitative

Mesurable, obtenue par appareil de mesure

→ taille, poids

Variable qualitative

Non mesurable

→ Couleur yeux, douleur

Population

Série **exhaustive** de tous les individus étudiés sur lesquels on veut **inférer des décisions**

→ Population française

Echantillon

Sous ensemble fini et **d’effectif limité**, extrait de la population

Il doit être **représentatif** d’où la nécessité d’un TAS

→ 100 personnes prises dans la pop française

/!\ échantillon connu // population inconnue /!\

Types de variables

- 2 types
  - Qualitatives
    - ***Binaires*** (H/F)
    - ***Nominales*** (couleur yeux)
    - ***Ordinales*** (douleur)
  - Quantitatives
    - ***Discrètes*** (âge)
    - ***Continues*** (poids)

    → La différence être les deux c'est que la variable continue possède un nombre infini de valeurs entre deux valeurs

Variable qualitative ordinale

- Une **variable qualitative ordinale** peut être approximée en **variable pseudo quantitative**

/!\ une variable pseudo quantitative reste qualitative /!

→ Les classements en médecine

Paramètres

- **Moyenne**
  - Variable quantitative discrète $m = \frac{\Sigma xi}{n}$
  - Variable quantitative continue $m = \frac{\Sigma xini}{n}$
- **Variance**
  - Indique la **dispersion** des données autour de la moyenne

Calcul de la moyenne :

- Moyenne de 12/16/5 ?
  - J’ai 3 valeurs donc n=3
  - $m = \frac{\sum xi}{n} = \frac{12+16+5}{3} = \frac{33}{3} = 11$

Paramètres

- **Médiane**
  - Valeur de l'observation centrale, divise la série d'effectif **en 2 sous-séries de même effectif**
    - Pour n pair = la médiane est la moyenne des deux valeurs correspondantes à $\frac{n}{2}$ et $\left(\frac{n}{2}\right)+1$
    - Pour n impair = la médiane est donnée par $\frac{n+1}{2}$
- **Quartiles**
  - Valeur de la variable qui partagent la série d'effectif en **4 sous-séries de même effectif**

Calcul de la médiane :

- Médiane de 12/16/5
  - On classe nos valeurs → 5/12/16
  - On a 3 valeurs donc n=3
  - Comme n est impair on applique :
  - $med = \frac{n+1}{2} = \frac{4}{2} = 2$
  - Ma 2e valeur est ma médiane → 12

Calcul du quartile :

- Quartile 1 de 12/16/5/8/14
  - On classe les valeurs → 5/8/12/14/16
  - On a 3 valeurs donc n=5
  - On applique la formule du quartile :
  - $Q1 = \frac{1}{4} * n = \frac{5}{4} = 1,25$
  - Donc Q1 entre valeur 1 et valeur 2
  - Moyenne des deux valeurs $m = \frac{5+8}{2} = 6,5$
  - Q1 =6,5

Avantages / Inconvénients moyenne/médiane

	Avantages	**Inconvénients**
Moyenne	- simple à calculer - facile à manipuler dans les test stats donc adaptées aux calculs statistiques - très significative si la répartition des données est assez symétrique et avec une faible dispersion	- sensible aux valeurs anormales (max et min)
Médiane	- calcul facile - peu sensible aux valeurs anormales - utilisable pour les valeurs ordinales, les classes…	- se prête moins aux calculs statistiques

Estimations en statistiques

- Les études biostats sont réalisées sur des **échantillons représentatifs** → **échantillonnage**
- Après l'étude on doit être sur que les résultats sont légitimes et qu'on peut les extrapoler à la population
- On va réaliser une **estimation** du résultat vrai à partir des données obtenues avec l'échantillon

Estimations en statistiques

- 2 types d'estimations
  - **Ponctuelles** : meilleure à instant t
  - **Par intervalle** : notion **d'intervalle de confiance**, +++fiable

Ecart-type

Dispersion d’un ensemble de données autour de la moyenne

= variabilité des mesures entre elles et // moyenne

+ faible + homogène

Degré de liberté

Nombre de valeurs nécessaires à connaitre pour pouvoir résoudre l'équation et **connaitre toutes les valeurs de la série**

Intervalle de confiance

Estimation de la moyenne vraie µ à partir de la <u>moyenne m</u> calculée sur l’échantillon

Intervalle de confiance

- On donne un intervalle auquel μ appartient :

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

- **IC = intervalle au risque α**

Risque α

Risque d'erreur dans <u>l'estimation de la moyenne vraie</u>

Risque α

- Risque que notre IC ne comprenne pas la valeur vraie de μ
- En général on prend un **risque α = 5%**

Ecart-réduit ε

Valeur qui **dépend du risque α**, ils varient en sens inverse

Ecart-réduit ε

Mesure de combien d'écart-types une **observation particulière est éloignée** de la population

α = 5%	ε = 1,96
α = 1%	ε = 2,60

Précision de l'estimation

- Les <u>variation du risque α </u>vont **conditionner la précision** de l'estimation et la **largeur de l'IC**
- \- de risque α → IC + grand donc ε↑ → + de chances que la moyenne soit dedans

Indice de précision i

Permet de calculer la **précision de l’estimation** de μ

Indice de précision i

- Permet de calculer la précision de l'estimation de μ
- <u>= largeur de l'IC</u>

$$\boldsymbol{i} = \frac{\varepsilon \boldsymbol{s}}{\sqrt{\boldsymbol{n}}}$$

Précision / IC

- IC est compris

  entre $[\boldsymbol{m} - \frac{\varepsilon \boldsymbol{s}}{\sqrt{\boldsymbol{n}}}]$ et $[\boldsymbol{m} + \frac{\varepsilon \boldsymbol{s}}{\sqrt{\boldsymbol{n}}}]$ donc entre $[\boldsymbol{m} - \boldsymbol{i}]$ et $[\boldsymbol{m} + \boldsymbol{i}]$.

- Donc d'après la formule de i

  si n↑ alors i↓ donc l'IC↓ donc la précision ↑

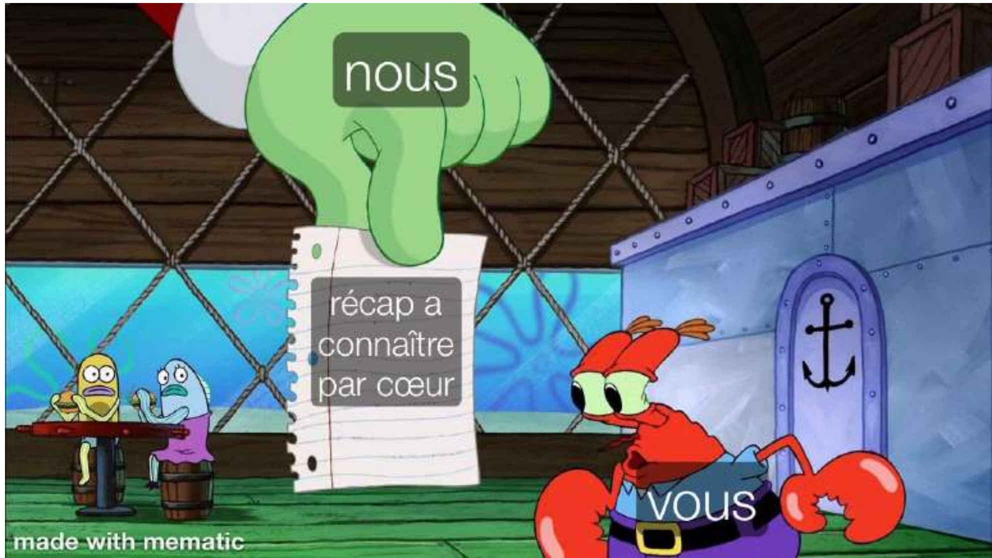
- On peut conclure que + n↑ → + précision ↑

RECAP

- L'IC = **estimation de la moyenne vraie µ** à partir de la **moyenne m** calculée sur l'échantillon. Il est aussi appelé **"intervalle au risque α"**.
- Risque α = **risque d'erreur** dans l'estimation de **µ**.
- ε = écart-réduit **distribution des données** autour de la **moyenne**.
- Les variations du risque α déterminent la précision de l'estimation.
- i représente la **largeur de l'IC** : $\boldsymbol{i} = \frac{\boldsymbol{\varepsilon s}}{\sqrt{\boldsymbol{n}}}$

IC = [m ∓ i]

- Si **n**↑ alors **i**↓ donc l'**IC**↓ donc la **précision**↑
- Si **α**↑ alors **ε**↓ donc **i**↓ donc l'**IC**↓ donc la **précision**↑



Loi de Gauss = loi Normale

- Distribution des variables plutôt symétriques autour de la moyenne avec forme de cloche = **courbe gaussienne = courbe de Gauss**
- Courbe en cloche avec
  - Abscisses = IC $[\boldsymbol{m} \pm \boldsymbol{\varepsilon s}]$
  - Ordonnée = n
  - Aire sous la courbe = % de la pop concernée

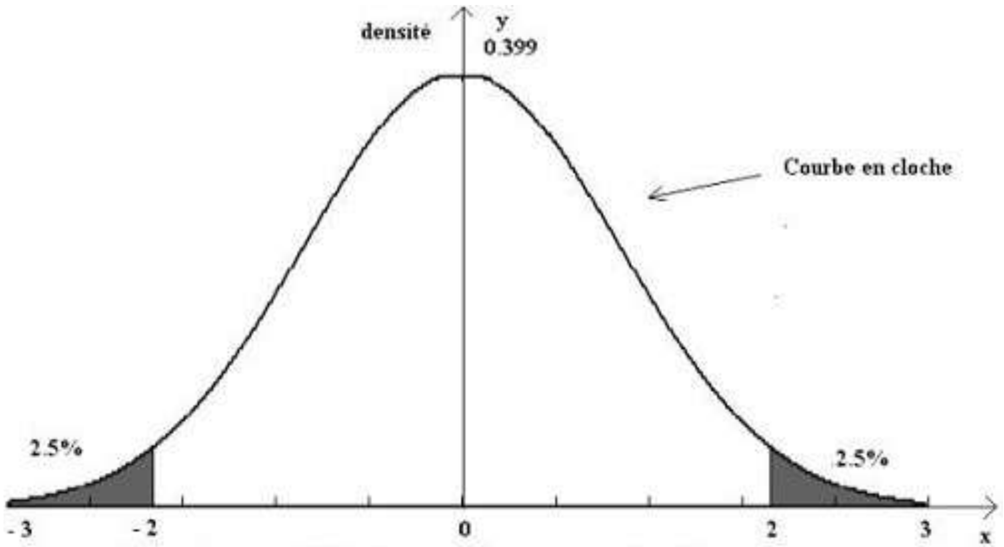


Figure 9.4 : densité de la loi normale de moyenne nulle et de variance un (courbe en cloche)

Loi de Gauss = loi Normale

- Cette loi permet de visualiser l'IC autour de la moyenne, l'écart-type, la dispersion autour de cette valeur moyenne et la moyenne.
- On utilise la **loi Normale** : pour chaque (μ,σ) il existe une loi normale de moyenne μ et d'écart type σ : on la note **N (μ, σ)**

Estimation données qualitatives

- On estime un **%**
- Estimation assurant **correspondance entre échantillon – population**

IC

= **estimation de la moyenne vraie µ** à partir de la moyenne m calculée sur échantillon

IC

- On donne un intervalle :

$$\boldsymbol{p \in [pobs \pm \varepsilon s]}$$

FIN

Socrative : BIOSTATSDUFEU

Reposez vous bien, on revient vite pour votre plus grand bonheur ☺