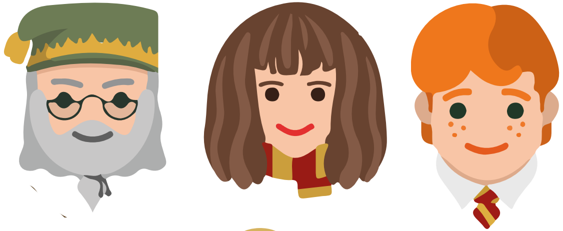
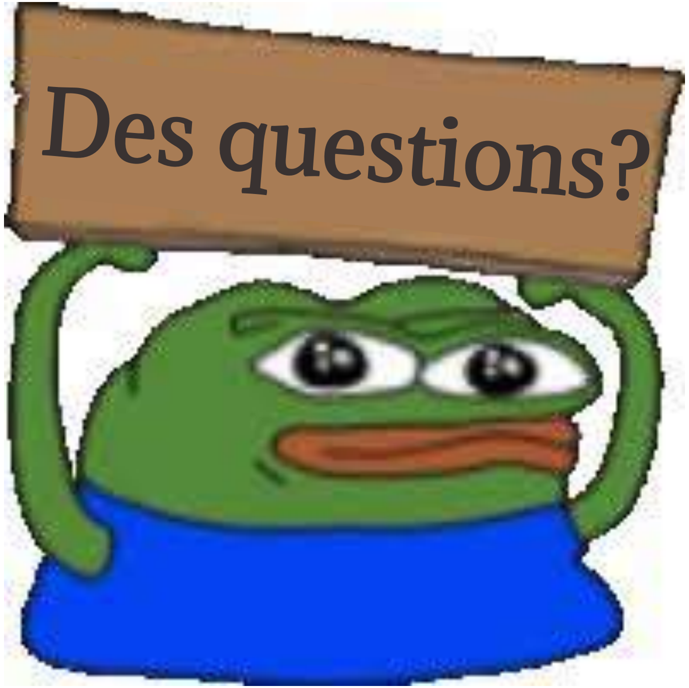
Méthodes



statistiques

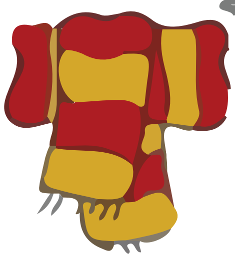
TTR Harrypot’tut

Camilyatomic, Madelyme, Eva’zectomie



Salle Socrative : BIOSTATREINE

1. Introduction & Definitions

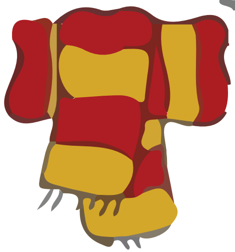


Biostatistiques : statistiques appliquées au domaine de la santé publique

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

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

1. Introduction & Definitions



Statistiques : art de collecter analyser et interpréter des données

→ 2 types de statistiques :

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

Données : résultat d'une observation

-> observation ou comparaison à d'autres individus -> variable

Une variable prend une valeur pour un individu

Grande variabilité inter-sujet ou intra-sujet

Définitions (oui encore…)

- Paramètre : grandeur apportant une information résumée sur la variable
- Série statistique : collection d'objets de même nature avec des caractéristiques différentes
- Population : série exhaustive de tous les individus étudiés
- Échantillon : sous-ensemble fini et d'effectif limité. Représentatif : TAS

Points ++

L'échantillon est **connu**

La population est **inconnue**

Variables :

Quantitative :	Qualitative :
Mesurée ou **dénombrée** Peut être : ❖ **Continue** ❖ **Discrète** *Ex : Taille, poids, nombre de cigarettes fumées par jour...*	Ne peut être mesuré mais susceptible d'etre **classée** ❖ Binaire : oui/non ❖ Multiple - **Nominale** (statut marital) - **Ordonnées** (gravité maladie)

Variable qualitative ordinale :

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

➡ Une variable pseudo quantitative reste **Qualitative**

3. Les Paramètres :

I- Moyenne :

Variable quantitative discrète :

$$\boldsymbol{m} = \frac{\Sigma x_i}{\boldsymbol{n}}$$

Variable quantitative continue :

$$\boldsymbol{m} = \frac{\Sigma x_i n_i}{\boldsymbol{n}}$$

3. Les Paramètres :

II- Variance :

Variance : la dispersion des données autour de la moyenne

3. Les Paramètres :

III- Mediane :

N Pair :

moyenne des valeurs
$\frac{n}{2}$ et $(\frac{n}{2}) + 1$

N impair :

$$\frac{n+1}{\mathbf{2}}$$

3. Les Paramètres :

IV- Quartiles :

Quartiles : séparent la série en 4 sous-séries de même effectif

Comparaison :

	Avantages	Inconvénients
Moyenne	+ Simple à calculer + Facile à manipuler dans des tests stats donc adaptée aux calculs statistiques + Très significative si la répartition des données est assez symétrique avec une faible dispersion	- Sensible aux valeurs anormales (max et min)
Médiane	+ Calcul facile + Peu sensible aux valeurs anormales + Utilisable pour des valeurs ordinales, des classes	- Se prête moins aux calculs statistiques

Estimation :

- ❖ 2 types d'estimation :

- Ponctuelle : meilleure à un instant t
- Par intervalle : Intervalle de Confiance

➡ **L'estimation par intervalle est moins précise mais plus fiable**

Données quantitatives :

- Écart-type : dispersion des données entre elles et autour de la moyenne
- Degré de liberté ou DDL : nombre de valeur à connaître pour résoudre une équation et connaître toutes les valeurs de la série

Intervalle de confiance :

- Intervalle de confiance : estimation de la moyenne vraie μ à partir de la moyenne m calculée sur l'échantillon

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

- Aussi appelé intervalle au risque α

Risque α

- C'est le risque d'erreur dans l'estimation de la moyenne vraie : le risque que μ ne soit pas dans l'intervalle

- En général α=5 %

Écart réduit ε

❖ Il varie en sens inverse du risque α

- Pour α = 5% ε=1,96
- Pour α = 1% ε=2,60

Précision de l'estimation

- Le risque α va conditionner la précision de l'estimation

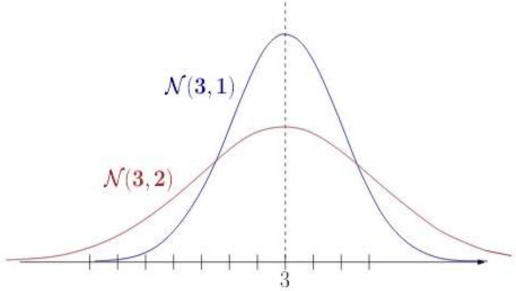
- Moins de risque -> intervalle augmente donc ε augmente : la moyenne a plus de chance d'être dedans mais la précision diminue

Indice de précision

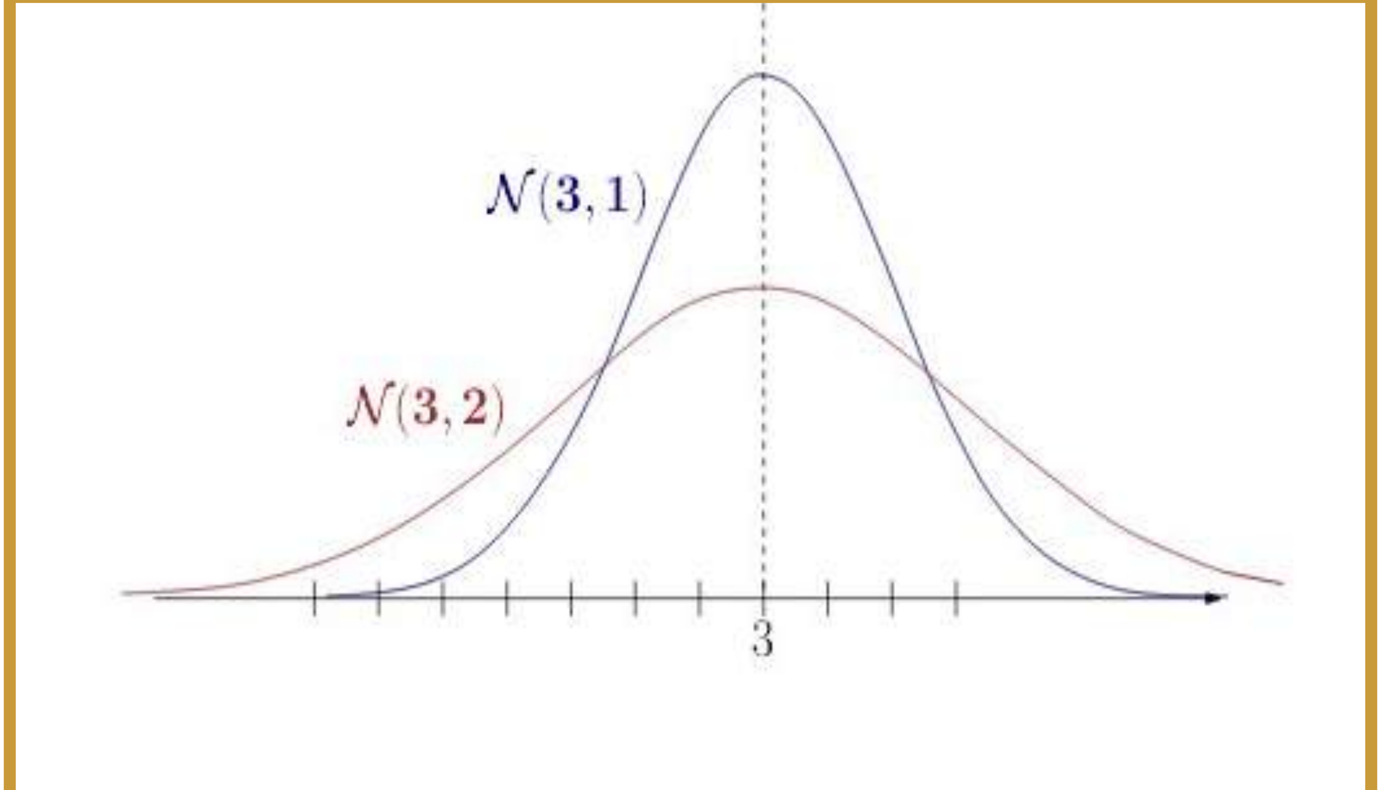
- Permet de calculer la précision de l'estimation de µ = largeur de l'IC
- $i = \varepsilon s/\sqrt{n}$
- Donc $\mu \in [m \pm i]$
- D'après la formule : si n↑, i↓ donc l'IC se resserre donc la précision ↑

Loi de Gauss ou loi normale

- Distribution en forme de cloche plutôt symétrique par rapport à la moyenne
- En abscisse : l’IC
- En ordonnée : n



- L’aire sous la courbe correspond au % de la population concernée
- Cette loi de Gauss est approximable par une loi normale N(μ, σ)


$\mathcal{N}(3,1)$
$\mathcal{N}(3,2)$
3

Données qualitatives

- On estime un pourcentage p à partir du pourcentage observé P_{obs}

- L'estimation assure une correspondance entre échantillon et population

Intervalle de confiance

- Estimation du pourcentage vrai p à partir du pourcentage observé p_obs sur l'échantillon

- $p \in [p_{obs} \pm \varepsilon s]$