

# Biostatistiques

*Oussama Essahili*



■ Pr Bennani

## Résumé basé sur QROCS

### Applications de la biostatistique

- Santé publique y compris l'épidémiologie
- Conception et analyse des essais cliniques
- Informatique médicale
- Santé environnementale
- Génétique et biologie moléculaire
- Écologie (prévision écologique)
- Bio-informatique

### Rôle de la biostatistique :

1. **Organiser** les données provenant des observations
2. **Décrire** les phénomènes par des paramètres résumant les observations
3. **Estimer** les valeurs de ces paramètres dans les populations d'où proviennent les échantillons observés
4. **Comparer** ces paramètres entre plusieurs populations
5. **Prédire** la probabilité de survenue des événements

### Définition de la variabilité biologique :

- Variabilité pris en compte par la biostatistique, entraîne une variabilité dans les échantillons, réservée aux mesures quantitatives ET qualitatives

#### Exemples :

moléculaire/génomique/cellulaire/fonctionnel/populationnel

### Variable statistique : Qualitative ou Quantitative

#### QUALITATIVE

Nominale	Graphique à secteurs : Camembert Diagramme en secteurs (n<5) Diagramme en barres horizontales (n>5)
Binaire/Dichotomique	Diagramme en secteurs (n<5)
Ordinale	Diagramme en bâtons

#### QUANTITATIVE

Continue/Mesurable	Polygone des effectifs ou des fréquences Histogramme des fréquences cumulées Box plot ou Boite à moustache
Temporelle	Diagramme en secteurs (n<5)
Discontinue	Histogramme

**Définition de l'échantillonnage :** Méthode qui permet d'étudier la variable sur une partie jugée représentative. (Car la totalité est difficile et coûteuse)

### Echantillonnage probabiliste/non probabiliste :

Probabiliste	Non probabiliste
Aléatoire simple	Accidentel
Aléatoire systématique	Quotas
En grappe	Volontaire
Stratifié	Par choix raisonné
	Boule de neige

**\*Inférence statistique :** est la capacité de généraliser les résultats à partir d'un échantillon sur toute la population cible.

**\*Echantillon représentatif d'une population :** produit les caractéristiques de toute la population concernant la question de recherche.

### Facteurs/Critères intervenant dans le choix d'une méthode d'échantillonnage appropriée

- Disponibilité ou non des données
- Ressources financières et humaines
- La durée
- La nature de question de recherche

### Biais de sélection :

Erreur systématique faite lors de la sélection des sujets à étudier.

### Etapes de l'organisation des données :

- 1)- Tri des données
- 2)- Regroupement en classes
- 3)- Transformation de variable
- 4)- Effectifs et fréquences
- 5)- Distribution : Plus les distributions sont normales, plus les données se rapprochent du centre.

### Mesures de tendance et de dispersion :

Mesure de tendance centrale	Mesure de dispersion
Moyenne	Etendue
Moyenne pondérée	Intervalle interquartile et semi-interquartile
Médiane	Variance
Mode	Ecart-type
	Coefficient de variation

### Les Loïs de distribution :

Loi	Variable
Loi binomiale	Type binaire
Loi de poisson	Qualitative
Loi normale	Variable quantitative continue

### Propriétés de la loi normale :

- Les trois mesures de tendance centrale sont égales.
- L'aire entre les deux points d'inflexion de la courbe mesure la probabilité que les valeurs de X soient comprises :
  - entre  $-1\sigma$  et  $+1\sigma$  autour de la moyenne => Probabilité de 68%
  - entre  $-1,96\sigma$  et  $+1,96\sigma$  autour de la moyenne => Probabilité de 95%
  - à l'extérieur de l'intervalle de  $2\sigma$  autour de la moyenne => Probabilité 2,5% à gauche et de 2,5% à droite

### Principes des tests statistiques :

1. Etablir une hypothèse nulle
2. Proposer une hypothèse alternative
3. Calcul d'un test de comparaison
4. Résultats d'un test de comparaison
5. Choix du risque d'erreur
6. Interprétation finale du test

### Différents tests de liaison : (3)

Définition : C'est des tests qui vérifient si la liaison observée entre deux ou plusieurs variables d'un échantillon

Finalité : Vérifier si la relation observée entre les variables étudiées de l'échantillon se rapprochent suffisamment du modèle théorique

- Test de tendance
- Test de corrélation
- Test de régression

### Risque d'erreur $\alpha$

C'est le risque de rejeter l'hypothèse nulle ( $H_0$ ) alors que celle-ci est vraie. Ce risque est parfaitement connu : c'est la probabilité utilisée lors de la réalisation du test, c'est à dire la probabilité pour que la valeur de la variable aléatoire Z soit extérieure à l'intervalle  $[z_1, z_2]$

### Différents tests de comparaison + Condition d'application

Variable quantitative
Test de Student ( $n < 30$ + distributions normales + variances égales)
Test de Z ( $n \geq 30$ = comparaison de 2 moyennes)

Variable qualitative
Ecart réduit (concerne % + $np \geq 5$ + $nq \geq 5$ avec $q = 1 - p$ )
Test de Z ( $n \geq 30$ = comparaison de 2 moyennes)

### Cas de petit échantillons (<5)

- Corrections de Yates
- Test exact de Fisher