

Biostatistiques

I La méthode statistique en médecine :

Les biostatistiques (= statistiques appliquées au domaine de la santé) ont 3 objectifs :

1. description d'une population par rapport à une maladie
2. **Evaluation** des traitements, des techniques, des coûts
3. Mise en place des informations épidémiologiques et en tirer des conclusions

Les biostatistiques doivent être capables de décider si une observation peut être due au seul hasard ou si elle a une autre explication.

Quelques définitions de base :

Terme	Définition	Exemple
Statistique	art de <u>collecter, d'analyser, et d'interpréter</u> des données. Lorsqu'elle est appliquée au domaine de la biologie, on parle de <u>biostatistique</u> . Il en existe 2 types : <ul style="list-style-type: none"> - <u>descriptive</u> - <u>déductive</u> : une observation est-elle due au hasard ? Ou existe-t-il une explication ? 	- <u>Stat descriptive</u> : Collecte de 2 données sur la population française : taille et couleur des yeux - <u>Stat déductive</u> : On constate que les sujets ayant une taille > 1,70m ont tous les yeux bleus. Hasard ? Corrélation ?
Population	<u>Série exhaustive de tous les individus</u> étudiés sur lesquels on veut appliquer des décisions	Ensemble de la <u>population française</u>
Echantillon	<u>Ensemble fini et d'effectif limité</u> extrait de la population, le plus souvent <u>randomisé</u> créé par tirage au sort = TAS) donc <u>représentatif</u>	<u>10 personnes tirées au sort</u> dans la population française.
Variable quantitative	résultat de l'observation d'un individu par l'utilisation d'un <u>appareil de mesure, variable</u> d'un individu à l'autre.	<u>Les tailles des 10 individus</u> : 1,62m / 1,63m / 1,66m / 1,66m / 1,68m / 1,68m / 1,70m / 1,75m / 1,80m / 1,90m
Variable qualitative	résultat de l'observation d'un individu, par les <u>sens de l'observateur. Variable</u> d'un individu à l'autre	<u>La couleur des yeux</u> : bleus / verts / marrons / gris
Paramètre	grandeur apportant une <u>information résumée</u> sur la variable étudiée	<u>La moyenne</u> : $m = 1,708m$ <u>La médiane</u> : 1,68m
Série statistique	Collection d'objets de même nature, présentant des <u>caractéristiques différentes</u> d'un objet à l'autre	Les <u>hommes et les femmes</u> sont des objets de même nature mais avec des caractéristiques différentes (mince, gros, blond, brun ...)
Variabilité	Ensemble des <u>différences inter-individuelles et intra-individuelles</u> . Elles peuvent être : <ul style="list-style-type: none"> - dues au hasard - physiologiques 	- <u>inter-individuelles</u> : Comparaison de la taille des sujets entre eux . - <u>intra-individuelles</u> : évolution de la taille avec l'âge, comparaison du sujet à lui-même à diverses périodes

II Statistique descriptive :

A Notion de variabilité :

Toute donnée biologique possède une **variabilité**. Sa connaissance est indispensable pour pouvoir dire si la valeur d'une variable **quantitative** est normale ou pas.

- Une variabilité maîtrisée permet d'établir une estimation.
- Une variabilité non maîtrisée conduit à des biais

Exemple :

La **valeur moyenne de la glycémie (variable quantitative)** chez un sujet normal est de **1g/L +/- 0,25 g/L**. Ceci signifie qu'une glycémie appartenant à l'intervalle : **[0,75 g/L ; 1,25 g/L]** est normale.

Vos tuteurs de biostat préférés décident de se contrôler la glycémie un matin à jeun. Ils trouvent :

- Vincent : 1,2 g/L. Il a donc une glycémie **normale**
- Julia : 0,7 g/L. Elle a donc une glycémie **infra - normale**

B La représentation des données :

Il existe divers types de données / variables :

- **Les variables qualitatives :**
 - x binaires : homme / femme ; malade / non malade ...
 - x nominales : couleur des yeux, des cheveux ...
 - x ordinales : degré de satisfaction des étudiants vis à vis de la tut rentrée : peu satisfait / satisfait / très satisfait / il n'y a pas de mots tellement c'était bon !
- **Les variables quantitatives :**
 - x discrètes : âge des étudiants en PAES ...
 - x continues : poids, glycémie ...

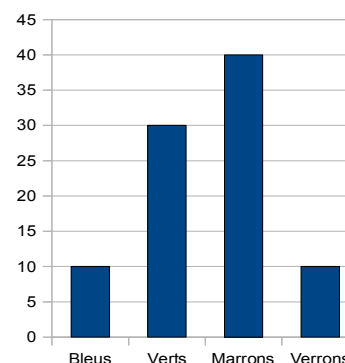
Les variables qualitatives peuvent être représentées de 2 manières :

- diagramme en bâton ou histogramme
- tableau

Exemple : On relève la couleur des yeux de 90 bébés à la sortie de la maternité. On constate que :

- 10 ont les yeux bleus
- 30 ont les yeux verts
- 40 ont les yeux marrons
- 10 ont les yeux verrons (un oeil vert et l'autre marron)

Couleur des yeux	Nombre de bébés	proportion
bleus	10	11,11%
verts	30	33,33%
marrons	40	44,44%
verrons	10	11,11%



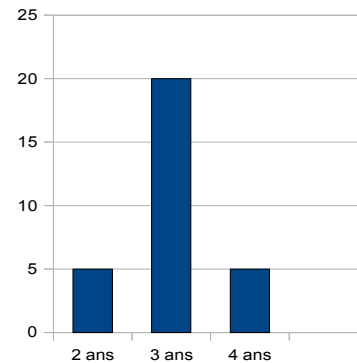
Les variables quantitatives peuvent être représentées/synthétisées de 2 manières :

- diagramme en bâton ou histogramme
- tableau
- synthétisées grâce à des paramètres

Exemple : On relève l'âge de 30 enfants d'une classe de 1ère année de maternelle :

- 5 ont 2 ans
- 20 ont 3 ans
- 5 ont 4 ans

Âge	Nombre d'enfants	proportion
2 ans	5	16,67%
3 ans	20	66,67%
4 ans	5	16,67%



Notion de paramètre : Un paramètre est une grandeur apportant une information résumée sur la variable étudiée. Il en existe plusieurs types :

- 1) La moyenne :
 - Pour une variable quantitative **discrète** : $m = \sum x_i/n$
 - Pour une variable quantitative **continue** : $m = \sum n_i x_i/n$
- 2) La variance : ou (écart type)². Elle indique la **dispersion** des données autour de la moyenne.
- 3) La médiane = observation centrale des valeurs qui sépare la série d'un effectif n en **2 sous-séries** de même effectif :
 - Si n est pair, la médiane est donnée par : $n/2$
 - Si n est impair, la médiane est donnée par $(n+1) / 2$
- 4) Les quartiles : Ce sont les valeurs **de la variable** qui partagent la série d'effectif n en **4 sous-séries** de même effectif.

Exemple : On relève le poids de 5 enfants d'une classe d'un service de pédiatrie. On trouve :

- 30 kg
- 31 kg
- 31 kg
- 32 kg
- 33 kg

- 1) La moyenne : Le poids est une variable quantitative continue.
 $m = [(30 \times 1) + (31 \times 2) + (32 \times 1) + (33 \times 1)] / 5 = 31,4 \text{ kg}$
- 2) La médiane : Soient les valeurs suivantes : 30 / 31 / **31** / 32 / 33.
 $(n+1) / 2 = 6/2 = 3$
 La médiane est donc la **3ème valeur des poids rangés par ordre croissant** : 31 kg.
- 3) Le 1er quartile : Soient les valeurs suivantes : 30 / 31 / 31 / 32 / 33.
 Le 1er quartile ou Q1 est la **valeur des poids qui délimite les premiers 25%** de la série.
 $Q1 = 0,25 \times 5 = 1,25 \rightarrow Q1$ se trouve entre la 1ère et la 2ème valeur des poids.
 Soit : $(30 + 31) / 2 = 30,5 \text{ kg}$.
- 4) Le 3ème quartile : Soient les valeurs suivantes : 30 / 31 / 31 / 32 / 33.
 Le 3ème quartile ou Q3 est la **valeur des poids qui délimite les premiers 75%** de la série.
 $Q3 = 0,75 \times 5 = 3,75 \rightarrow Q3$ se trouve entre la 3ème et la 4ème valeur des poids.
 Soit : $(31 + 32) / 2 = 31,5 \text{ kg}$.

	avantages	inconvénients
moyenne	<ul style="list-style-type: none">- Facile à calculer, à manipuler- significative si la répartition symétrique des données- dispersion faible	<ul style="list-style-type: none">- Sensibles aux valeurs anormales- sensible aux minimum- sensible aux maximum
médiane	<ul style="list-style-type: none">- Facile à calculer- peu sensible aux valeurs anormales- utilisable pour les valeurs ordinales	<ul style="list-style-type: none">- Moins adéquat pour les calculs statistiques