



## I-LA METHODE STATISTIQUE EN MEDECINE

\*La **statistique** est l'art de collecter, d'analyser et d'interpréter des données. Elle est le champ d'extension de l'étude des probabilités.

NE PAS CONFONDRE :

- LA statistique qui est une technique
- LES statistiques qui sont des collections de données

\*La **biostatistique** est l'application de la statistique au domaine biologique.

\*Une **donnée** est le résultat de l'observation d'un individu par un instrument de mesure (taille, poids...) OU par les sens de l'observateur (couleur de peau...)

\*Une **variable** est :

- une donnée qui n'est pas strictement équivalente d'un individu à l'autre  
=> **variabilité inter-individuelle** (*couleur des yeux...*)
- une donnée qui n'est pas strictement équivalente d'un instant à l'autre  
=> **variabilité intra-individuelle** (*la taille d'un enfant...*)

### Petit aparté sur les différents types de variables !!

Quantitatives (mesure de quantités)		Qualitatives (mesure de qualités)	
<b>Discrète</b>	<b>Continue</b>	<b>Nominale</b> (sans ordre)	<b>Ordinale</b> (avec ordre)
• <i>nombre d'accouchements d'une femme</i>	• <i>poids, taille...</i>	• <i>couleur des yeux</i> <b>Binaire</b> (sans ordre) • <i>sexe</i>	• <i>degré de satisfaction</i>

NB : une variable qualitative BINAIRE est le plus souvent un cas particulier de variable qualitative NOMINALE !!

NB : certaines variables qualitatives peuvent être numériques !

Par exemple : Nombre de cigarettes fumées par jour :

- moins de 5
  - entre 5 et 10
  - plus de 10
- => **La réponse attendue est sous forme de variable qualitative ordinale**

\*Un **paramètre** est une grandeur apportant une information résumée sur la variable étudiée (moyenne, variance, médiane...).

=> L'étude de ces paramètres fait partie de la statistique descriptive.

\*L'idée centrale de la statistique est que la **variabilité a deux types d'origines possibles** :

- le **hasard** = la **variabilité intrinsèque**
- une autre origine

=> La statistique déductive (= explicative = inductive) permet d'identifier laquelle des ces deux origines entre en jeu pour une situation donnée.

=> La Statistique Descriptive est l'étude d'une situation à l'aide de paramètres.

=> La Statistique Déductive / Explicative / Inductive consiste à définir si la variabilité d'une observation est due au hasard ou à une autre explication.

Ainsi, deux grands principes régissent la méthode statistique :

**Toute observation est soumise à une variabilité intrinsèque.**

**En effet, pour une observation répétée plusieurs de fois, le résultat est presque toujours variable.**

=> *Exemple : deux traitements anti cancéreux donnent à 5 ans une survie de 42 % pour l'un et de 48 % pour l'autre. Cette différence est-elle une conséquence du hasard ou d'une efficacité plus grande pour l'un des deux traitements ? Nous verrons plus tard que des **TESTS STATISTIQUES** (appartenant au domaine de la statistique déductive) nous permettent, à un risque d'erreur près, de répondre à cette question.*

**L'observation d'une différence ne permet pas en soi d'en préciser la cause.**

**Constater une différence statistiquement significative ne donne pas la clé de son interprétation.**

=> *Attention ! Bien que les tests évoqués ci-dessus nous permettent de dire si il existe ou non une variabilité due à autre chose que le hasard, ils ne nous renseignent pas sur l'origine de cette « autre chose » ! Imaginons que soit avérée une meilleure survie dans l'un des deux groupes ; la preuve n'est pas forcément établie que ce résultat soit la conséquence d'un traitement meilleur que l'autre. Par exemple, l'observance du traitement peut ne pas avoir été bien suivie dans l'un des deux groupes, sans que la molécule en elle-même soit pour autant moins efficace. Un biais peut également s'être présenté lors de la constitution des groupes. D'où la nécessité d'un regard critique !*

Toutes ces notions seront approfondies dans la troisième partie du cours....

\*Une **série statistique** est une collection d'objets de même nature avec des caractéristiques différentes d'un objet à l'autre. Deux séries statistiques primordiales :

- **Population** : série de TOUS les individus étudiés, sur lesquels on veut appliquer (= inférer) des décisions (*la population française ; l'ensemble des étudiants de France en PAES ; les habitants de Nice ; les femmes habitants Nice...*). La valeur de l'effectif en est souvent inconnu.

- **Echantillon** : ensemble d'effectif limité et connu extrait de la population (*200 personnes tirés au sort dans la population française...*).

=> Au cours d'une enquête de Santé Publique (par exemple), les études sont menées sur un **échantillon représentatif** de la population dont il est issu.

La **SEULE** technique de sélection d'un échantillon **représentatif** d'une population est le **TIRAGE AU SORT (= RANDOMISATION) !!!!!**

Si l'échantillon est bien représentatif (donc non biaisé) et que l'étude a été correctement menée, on « pari » sur l'extension ou l'extrapolation des résultats à la population entière. Cela sera détaillé dans la deuxième partie du cours.

=> On dit que la **population** est **inconnue** et que l'**échantillon** est **connu**.

