



Université Internationale  
de Casablanca


LAUREATE INTERNATIONAL UNIVERSITIES

# *STATISTIQUE DESCRIPTIVE ET PROBABILITÉ*

Filière : Nutrition et Diététique

Semestre 1

*MME IMANE GHANNAME, MS, PHD*



Un homme souffrant d'une douleur dans la région cardiaque se présente aux urgences. Quelle est la probabilité qu'il ait une sténose coronarienne ?

Sachant que :

- Le patient doit faire une épreuve d'effort avant la coronarographie
- $P(\text{EE}+) = 0,7$
- $P(\text{Coro-}/\text{EE}+) = 0,15$
- $P(\text{Coro+}/\text{EE-}) = 0,2$



Imaginons un ensemble important de données.



Exemple: la taille des 200 étudiants d'un auditoire de première candidature.

170 162 184 155 180 167 170 166 16 164 172 167  
180 154 180 151 153 164 177 149 173 160 152 173  
158 143 169 170 154 168 153 165 160 162 162 168  
175 169 166 177 175 148 167 166 159 178 156 172  
143 171 181 170 174 153 173 165 156 163 173  
162 160 162 165 146 177 165 163 176 172 178 135  
151 154 145 170 164 163 165 178 171 166 157 156  
167 157 154 164 166 184 167 164 167 153 170 162  
192 154 166 170 170 158 167 154 169 162 169 162  
158 151 179 159 171 165 165 166 180 180 172 165  
155 151 158 164 184 170 154 162 166 150 169 173  
155 173 149 174 168 162 172 158 183 175 176 165  
147 168 168 171 148 166 171 165 176 145 155 176  
163 176 167 171 169 171 169 172 171 178 155 164  
176 155 173 158 149 176 146 151 166 163 163 147  
161 149 155 146 155 177 168 166 168 163 152 169  
170 159 163 186 162 148 173 180 150

En regardant ces données, que peut-on dire ?

A priori pas grand chose. Il n'y a pas d'information qui ressort immédiatement de ce tableau

# INTRODUCTION

- Disponibilité des tableurs et logiciels: Calculs rebutants → Risque d'erreurs méthodologique & pertinence non-garantie 
- Les variables étudiées dans les sciences de la vie ne sont pas gouvernées par le hasard, elles sont déterminées par la notion de variabilité biologique (loi de la génétique, contraintes de l'environnement...)
- Sciences du vivant: la statistique est fondée sur une approximation qui assimile la distribution des valeurs observées à des modèles théoriques →  oublier cette approximation et vous persister à l'utilisation de ces modèles sans vérifier leur pertinence → obtenir des résultats justes d'un points de vue mathématique mais totalement faux dans leur interprétation statistique.



La statistique a défini des conditions d'utilisation des modèles théoriques: Doivent être **impérativement** respectées

# INTRODUCTION

## Place de la statistique

### Etape 1: Conceptualisation

- Problématique, Objectifs, **Hypothèse**...

### Etape 2: Planification

- **Horizon temporel, taille d'échantillon**...

### Etape 3: Acquisition des données

### Etape 4: Analyse des données

- **Description, distribution, comparaison, lecture critique**...

### Etape 5: Interprétation statistique

- **Conclusion statistique, interprétation clinique**...

### Etape 6: Communication des résultats

# INTRODUCTION

## Intérêt de la statistique

- La statistique désigne l'ensemble des méthodes mathématiques relative à la collecte, à la présentation, à l'analyse et à l'utilisation des données numériques. Ces opérations permettent de tirer des conclusions et de prendre des décisions dans les situations d'incertitudes qu'on rencontre dans les domaines scientifiques, économiques, sciences sociales ou des affaires...
  
- La statistique permet de :
  1. Organiser les données disparates prévenant des observations individuelles;
  2. Décrire clairement les phénomènes par des paramètres résumant ces observation;
  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 d'événements.

# INTRODUCTION

## Objectifs

1. Maîtriser les notions de base en statistique
  2. Collecter des données
  3. Construire des graphiques
  4. Déterminer des caractéristiques centrale
  5. Calculer des caractéristiques de dispersion
  6. Maîtriser l'organisation, la description et la présentation des données sous forme de tableaux ou de graphiques
  7. Analyser, interpréter, comparer et conclure
- } **Statistique descriptive**
- } **Statistique Inférentielle**

# ORGANISATION DU COURS

## Planning du cours

- Première partie: Définition et généralités
  - Echantillonnage
  - Biais
- Deuxième partie: Calculs des probabilités
  - Notions de base
  - Variables aléatoires
  - Lois classiques
  - Convergence de v.a
- Troisième partie: Statistique descriptive
  - Variables
  - Organisation des données
  - Description des données
  - Mesure en statistique
  - Représentation d'une distribution
  - Lois de distribution



# ORGANISATION DU COURS

## Evaluation

Vous serez sanctionnés par un simple « acquis » ou un « non acquis » une fois avancé dans le cours.

Pour obtenir un « acquis », vous devez participer activement aux séances et aux travaux intermédiaires.

# DÉFINITION & GÉNÉRALITÉ

## La statistique et les statistiques

On distingue :

**Les statistiques** : sont des dénombrements fournissant des renseignements d'ordre administratifs.

- Nbre de naissances
- Nbre d'accidents de roulage
- Le PIB

**La statistique** : méthode de raisonnement permettant d'interpréter des données numériques très particulières qu'on rencontre dans les sciences en générale et dont le caractère essentiel est la variabilité.

# DÉFINITION & GÉNÉRALITÉ

## La variabilité

Dans le domaine du vivant (biologie, santé publique, médecine, odontologie, psychologie, zoologie, botanique notamment) la caractéristique principale des phénomènes est la variabilité

La variabilité observée a deux type de sources:

- Individualité biologique : produite par l'interaction de très nbx facteurs groupés dans l'héritage génétique et l'influence du milieu : **V. biologique = {V. Inter-individuelle, V. Intra-individuelle}**
  - **La variabilité biologique ne peut pas être supprimée**
- Les erreurs que le chercheur peut commettre: **V. métrologique = {erreur de mesure, erreur de recueil de données, erreur de sélection, erreur instrumentale...}**
  - **La variabilité métrologique peut être supprimée, diminuée ou corrigée**

# DÉFINITION & GÉNÉRALITÉ

## Variables ou descripteur

Ce sont les observateurs sous forme de caractères, attributs, au moyen desquels on décrit et compare les éléments ou objets de l'étude. Ces observations sont sujettes à des variations quantitatives ou qualitatives.

Il existe plusieurs types de variables:

- **Aléatoire:** Susceptibles de prendre diverses valeurs non prévisibles
- **Simple:** Elle prend une seule valeur
- **Complexe:** Dérivée ou synthétique: variable calculée à partir de plusieurs mesures
- **Contrôlée:** Elle est contrôlée par des manipulations
- **Propre:** représente un attribut de l'élément étudié (taille, poids)
- **Associée:** une variable qui n'est pas mesurée sur l'éléments mais sur son environnement (tp°, HR)

# DÉFINITION & GÉNÉRALITÉ

## Variables ou descripteur

- **Qualitative ou nominale:** variable non ordonnée et ne peut pas être mesurée mais plutôt observée (couleur, espèce, présence/absence...)
- **Semi-qualitative:** variable ordonnée comme le rang
- **Quantitative discrète:** variable ordonnée qui peut prendre que les valeurs entières (nb d'enfants...)
- **Quantitative continue:** Variable ordonnée qui peut prendre toutes les valeurs dans un intervalle et qui peut être mesurée (hauteur, poids...)
- **Dépendante:** Variable dont on cherche à comprendre les variations: c'est la variable à expliquer
- **Indépendante:** Variable qui explique une partie de la variation du phénomène étudié: c'est la variable explicative
- **Subjective:** Variable non mesurable (Qualité de vie...)

# DÉFINITION & GÉNÉRALITÉ

## Population

- L'ensemble des individus sur lesquels on fait des statistiques. Le terme individu peut désigner un objet, un phénomène...
- Le nombre d'individu qui compose une population est appelé l'effectif de la population.
- Souvent de grande taille (voire infinie) : impossible de tester tous les individus d'une population.

# DÉFINITION & GÉNÉRALITÉ

**Population cible**

- L'ensemble des objets et des éléments visés par un échantillonnage. Souvent cette population est trop vaste ou trop changeante pour qu'elle puisse être considérée comme une population statistique

# DÉFINITION & GÉNÉRALITÉ

## Population statistique

- Collection d'éléments possédant au moins une caractéristique commune et exclusive, permettant de l'identifier et de la distinguer sans ambiguïté de toute autre.
- C'est dans cet ensemble que sera pris l'échantillon et c'est à cet ensemble que s'appliqueront les conclusions de l'inférence statistique.



# DÉFINITION & GÉNÉRALITÉ

## Echantillon

- C'est la partie de la population statistique qui sera récoltée et étudiée pour connaître les propriétés de cette population.

Le processus d'échantillonnage doit être aléatoire et représentatif de la population statistique, pour cela:

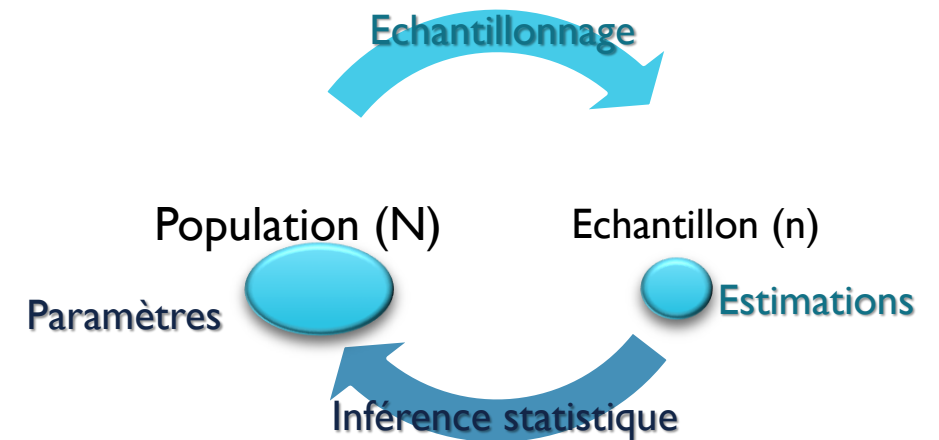
- L'échantillon ne doit pas avoir de biais systématique
- Chaque élément doit avoir une probabilité connue et non nulle de faire partie de l'échantillon
- La taille et l'effectif de l'échantillon ne doivent pas affecter sa représentativité

# DÉFINITION & GÉNÉRALITÉ

## Echantillon

- L'échantillon sera dit représentatif si on peut correctement estimer les paramètres d'intérêt de la population à partir de l'échantillon
- il faut que tout les profils de la population importants pour l'objectifs de l'enquête soient représentés dans l'échantillon

**Note:** Dans le cas contraire on parlera de biais d'échantillonnage ou biais de sélection.



# DÉFINITION & GÉNÉRALITÉ

## Inférence statistique

- La démarche inductive qui consiste à caractériser un ensemble appelé population statistique à partir d'un sous ensemble plus petit appelé échantillon.

# ECHANTILLONNAGE

La procédure d'échantillonnage doit permettre la constitution d'un sous- groupe recouvrant **les caractéristiques qui peuvent influencer la valeur** des paramètres que l'on veut estimer.

- Exemple:

Etude d'une maladie / échantillon par téléphone

→ Les hospitalisés: non inclus

# ECHANTILLONNAGE

## Type d'échantillonnage

### 1. Echantillonnage non-probabiliste

- *Les échantillons de convenance,*
- *Les Echantillons de volontaires*
- *Les Méthodes des quotas*

### 2. Echantillonnage probabiliste

*Echantillonnage à un degré*

- *Echantillonnage aléatoire simple*
- *Echantillonnage aléatoire stratifié*
- *Echantillonnage aléatoire en grappe*

*Echantillonnage à plusieurs degrés*

# ECHANTILLONNAGE

## Echantillonnage non-probabiliste

- Appelées aussi **méthodes empiriques** ou **par choix raisonné**.
- Sélection des individus qui n'obéit **pas au hasard**
- Définie selon des critères de **faisabilité**, de **ressemblance** à la population-cible et de **critères subjectifs** dépendant du choix des enquêteurs.
- Les méthodes d'échantillonnage non aléatoires les plus couramment utilisées sont :
  - *les échantillons de convenance*
  - *Les Echantillons de volontaires*
  - *Les Méthodes des quotas*

# ECHANTILLONNAGE

## Echantillonnage non-probabiliste

Echantillons de convenance

- Échantillon **d'individus facilement interrogeables**
- Généralement: des individus qui se présentent à un endroit donné à un moment donné

Par exemple: les études réalisées dans la rue

- **Les avantages**
  - la facilité d'application
  - l'absence d'influence de l'investigateur.
- **Les inconvénients:**
  - La non représentativité
  - L'impossibilité d'évaluer le biais associé.

# ECHANTILLONNAGE

## Echantillonnage non-probabiliste

Echantillons de volontaires

- **Les individus se sélectionnent eux-mêmes**
- Appel à participation par: annonces dans les journaux locaux ou nationaux et/ou d'un dépliant distribué
- **Les avantages**
  - Attractif du point de vue de l'éthique
  - Utile pour les phases exploratoires.
- **Les inconvénients:**
  - La non représentativité
  - L'impossibilité d'évaluer le biais associé.



# ECHANTILLONNAGE

## Echantillonnage non-probabiliste

Méthodes de Quotas

**Hypothèse:** si un échantillon est représentatif sur quelques grandes variables sociodémographiques, il sera, alors, représentatif sur les variables que l'on veut étudier.

### Démarche:

- Construire un échantillon qui soit un modèle réduit de la population étudiée
- Les mêmes proportions en ce qui concerne des caractéristiques (**les variables de contrôle**): les variables sociodémographiques
- **Les avantages**
  - Ne nécessite pas de base de sondage mais uniquement la connaissance de la répartition dans la population selon certaines caractéristiques
  - Le coût est plus faible: personne précise non obligatoire.
- **Les inconvénients:**
  - La non représentativité
  - L'impossibilité d'évaluer le biais associé.
  - Non connaissance des taux de réponse

# ECHANTILLONNAGE

## Echantillonnage probabiliste

- **Échantillonnage aléatoire:** tous les individus de la population-source ont une **probabilité** connue et non nulle d'être sélectionnés pour faire partie de l'échantillon
- Pas d'intervention du chercheur: seul le hasard régit l'inclusion ou non d'un individu dans l'échantillon.
- Les informations recueillies sur l'échantillon peuvent être inférées pour la population source

# ECHANTILLONNAGE

## Echantillonnage probabiliste

Echantillonnage à un degré

- Un seul tirage au sort (table de nombre au hasard, Excel...)
- Selon le type de base de sondage disponible, il peut être:
  - Simple
  - Stratifié
  - En grappe

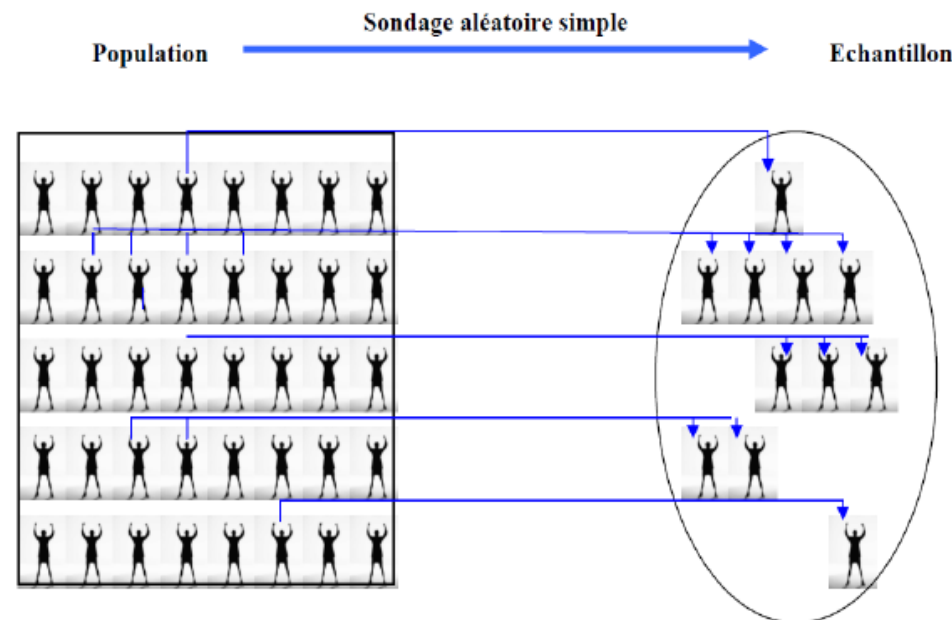
# ECHANTILLONNAGE

## Echantillonnage probabiliste

Echantillonnage à un degré

### Echantillonnage aléatoire simple

- Il s'agit de sélectionner à partir d'une liste ou base de sondage un échantillon de  $n$  individu par tirage au sort et ceci en une seule étape



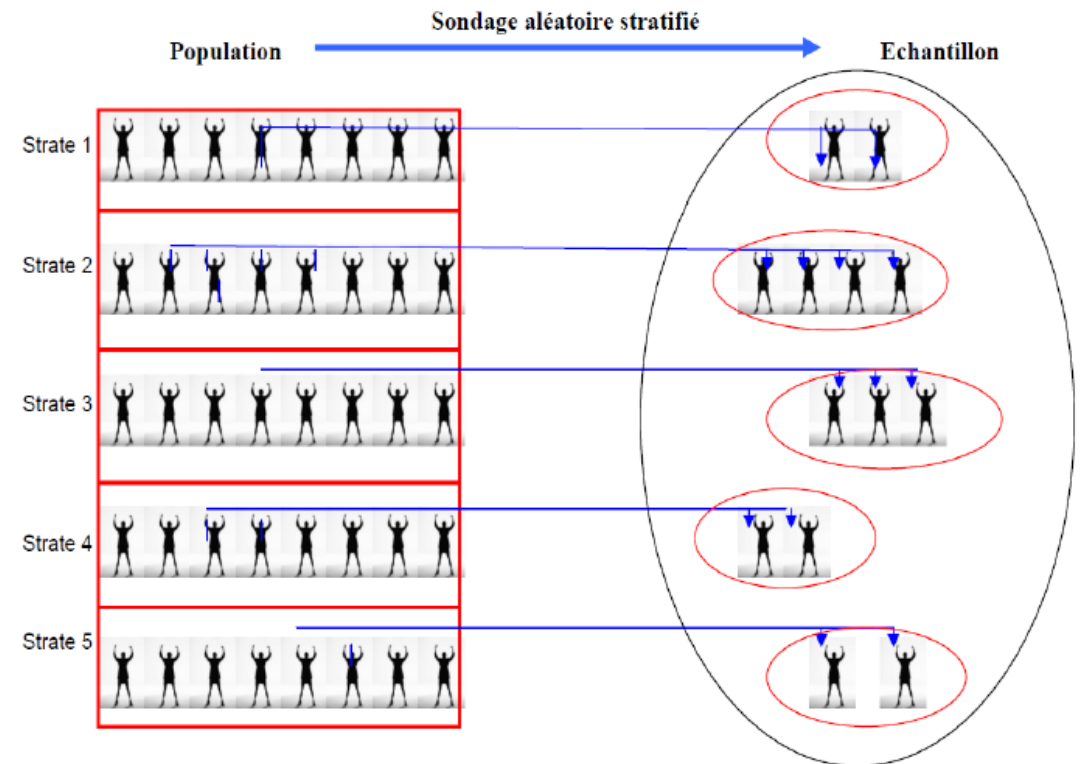
# ECHANTILLONNAGE

## Echantillonnage probabiliste

Echantillonnage à un degré

### Echantillonnage aléatoire stratifié

- La base de sondage peut être **divisé en strates**
- Ces strates sont plus **homogène** que la population globale
- Estimations plus précises
- Sélectionner des individus dans chaque strates est mieux que d'augmenter la taille d'échantillon
- Utile lorsqu'on souhaite sur-représenter dans l'échantillon des catégories de population qui doivent faire l'objet d'études particulières.



# ECHANTILLONNAGE

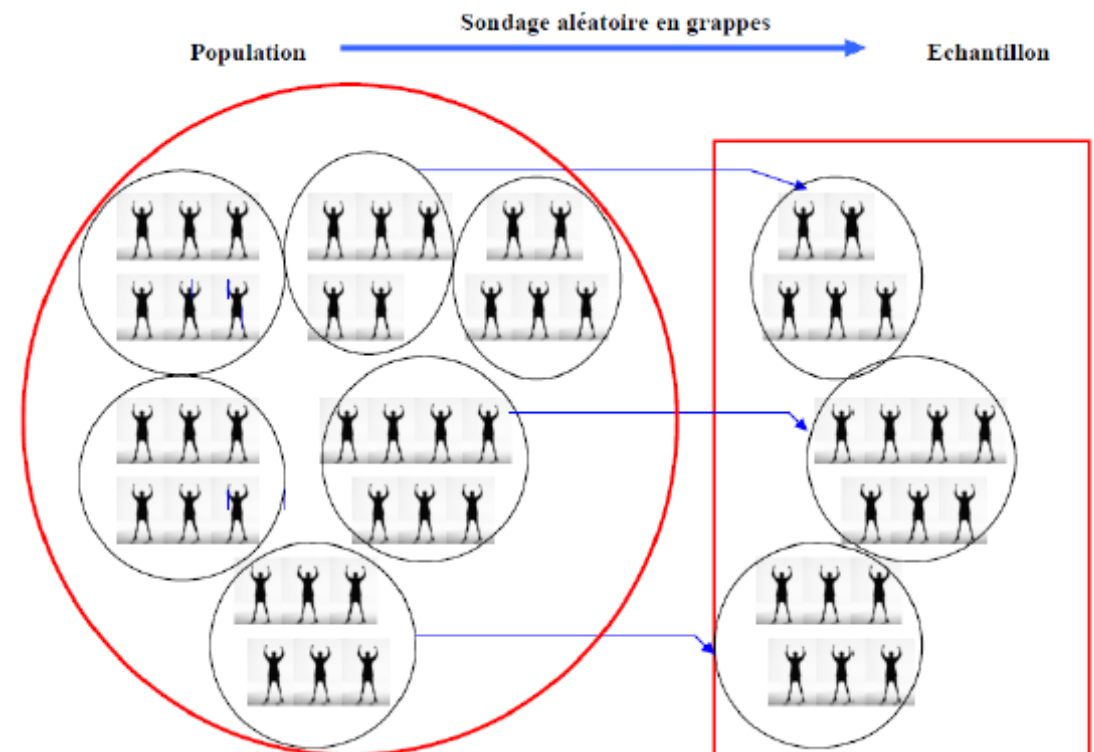
## Echantillonnage probabiliste

Echantillonnage à un degré

### Echantillonnage aléatoire en grappe

- Des sondages réalisés sur des groupes complets d'individus
- La population source est subdivisée naturellement en groupes (La composition des grappes est antérieure au plan de sondage)
- Un certain nombre de ces groupes va être sélectionné aléatoirement pour composer l'échantillon.
- **Intérêt:** lorsque l'accès à la liste des individus composant la population d'étude n'est pas possible

**Exemple:** Des listes de classes obtenues à partir d'établissements scolaires

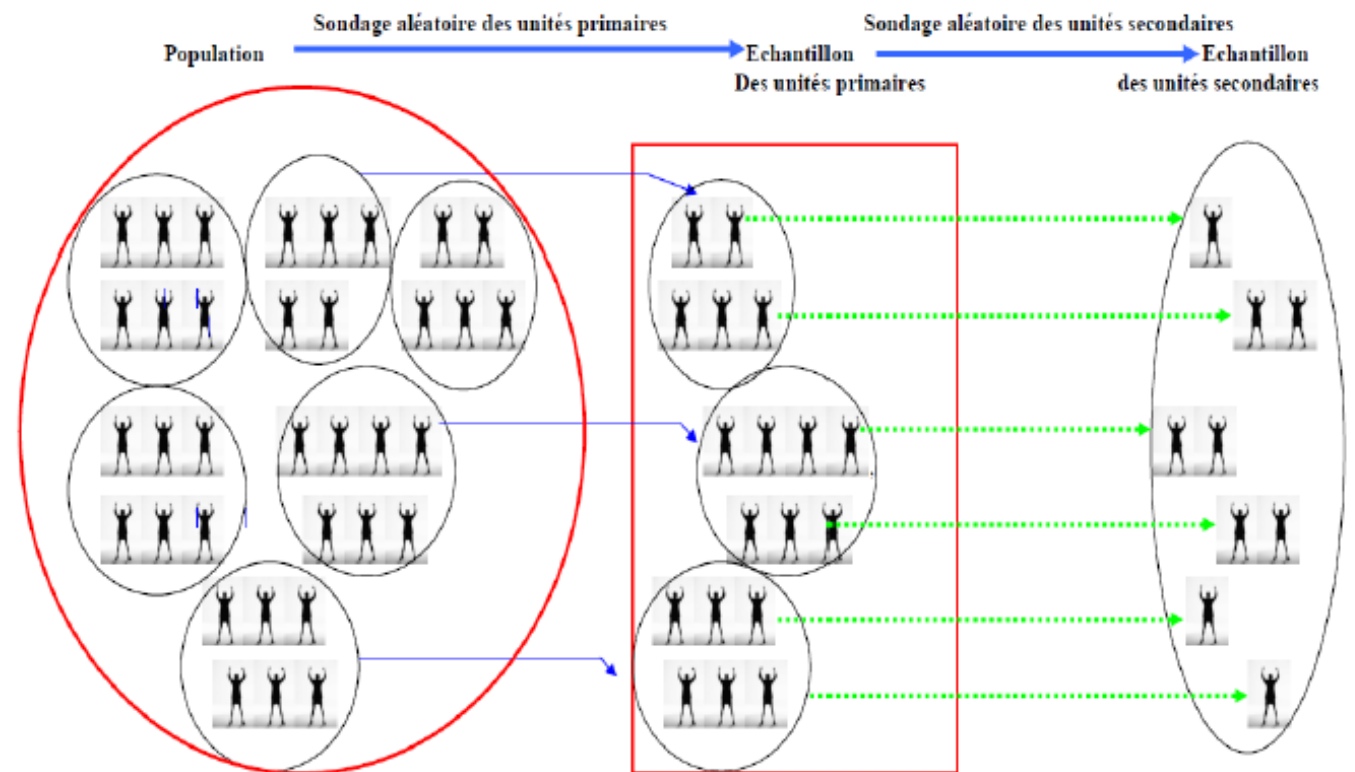


# ECHANTILLONNAGE

## Echantillonnage probabiliste Echantillonnage à plusieurs degrés

- Un sondage pour lequel plusieurs sélections aléatoires emboîtées sont réalisées.

**Intérêt:** lorsque l'accès à la liste des individus composant la population d'étude n'est pas possible



# ECHANTILLONNAGE

## Echantillonnage probabiliste

Echantillonnage à un degré

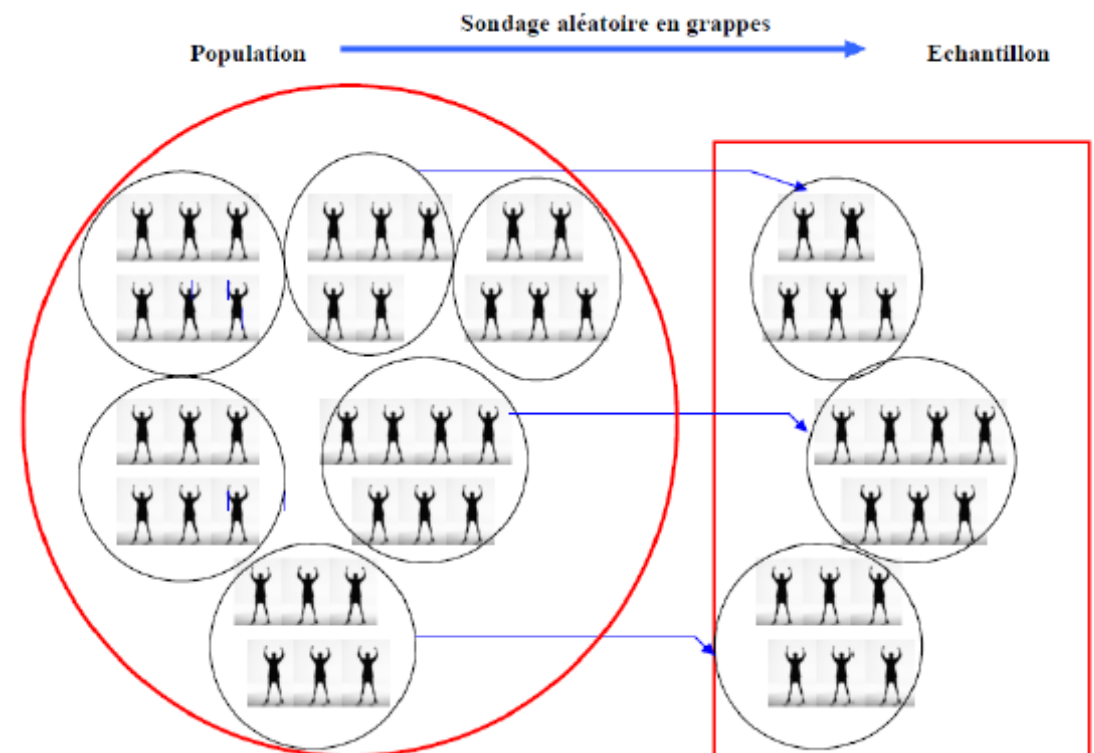
### Echantillonnage aléatoire en grappe

#### Avantages:

- Echantillonnage aléatoire malgré l'absence de liste exhaustive
- Réduction des coûts / temps.

#### Limites:

- Les grappes: risque de ne pas représenter correctement la variabilité
- Les grappes utilisées doivent être de tailles à peu près équivalentes





Deux types d'erreurs en épidémiologie:

- Erreurs dues au **hasard** (fluctuation d'échantillonnage)
- **Erreur systématique** ayant pour conséquence une estimation incorrecte d'un paramètre ou une distorsion d'une mesure de l'association

- **BIAIS**
  - De sélection
  - De mémoire
  - De confusion

- La population dont est extrait l'échantillon d'étude (population source) est différente de la population à laquelle on souhaite généraliser les résultats (population cible) ou les groupes de comparaison (exposés/non-exposés ou cas/témoins) ne sont pas comparables.

### Principales sources de biais de sélection

- La constitution d'un échantillon par un autre moyen que le **tirage au sort** (sujets volontaires);
- Les **non-réponses** à une enquête ;
- Le recrutement de témoins en **milieu de travail** qui ne sont pas représentatifs de la population générale («Healthy Worker Effect») ;
- Les sujets **perdus de vue** dans les enquêtes exposé / non-exposé.

- La population dont est extrait l'échantillon d'étude (population source) est différente de la population à laquelle on souhaite généraliser les résultats (population cible) ou les groupes de comparaison (exposés/non-exposés ou cas/témoins) ne sont pas comparables.

### Prévention des biais de sélection

- Le tirage au sort et la randomisation
- Lors du recueil des données pour limiter le nombre de non répondants
- Au cours du suivi lors des études de cohortes pour limiter le nombre de perdus de vue
- On choisissant les deux groupes du même «univers»

***Une erreur systématique de mesure de l'exposition ou de la maladie.***

- Biais d'enquêteur ou de subjectivité
- Biais de mémorisation
- Qualité médiocre des données disponibles surtout dans les enquêtes rétrospectives

### Prévention des biais d'information

- Choix de la méthode de mesure de l'exposition et de la maladie (questionnaire affinée, formation des enquêteurs, procédure d'entrevue et de surveillance, double insu...)

# BIAIS

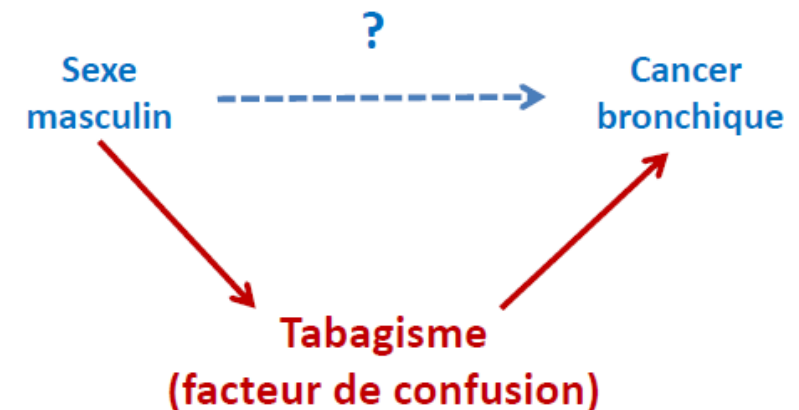
## Biais de confusion

Une erreur systématique dans l'estimation d'une **mesure d'association** (OR, RR) entre le facteur étudié et la maladie, du fait d'un défaut de prise en compte d'un **facteur de confusion**.

- Un facteur de confusion est un facteur lié à la fois à l'exposition et à la maladie étudiée.

### Prévention des biais de confusion

- Lors de la constitution des groupes de comparaison: **Randomisation**, **Appariement**
- Lors de l'analyse: **Ajustement** (si les données concernant le tiers facteur ont été recueillies au cours de l'étude)





# *PROBABILITÉS*

# DÉFINITIONS

## Expérience aléatoire, événement aléatoire

- Une **expérience** est dite **aléatoire** (random experiment-random trial) lorsqu'on ne peut pas en prévoir exactement les résultats du fait que **tous les facteurs qui déterminent ce résultat ne sont pas maîtrisés** ou contrôlés.
- Un **événement aléatoire** est un événement qui **peut ou ne pas se réaliser** au cours d'une expérience aléatoire.
- **Exemple :**
  - Expérience aléatoire "traverser la route"
  - Événement aléatoire "se faire écraser"

# DÉFINITIONS

**Expérience aléatoire, événement aléatoire**

N°	Expérience	Ensemble de résultats possibles
1	Jeter un dé et relever le nombre qui est sur la face supérieure	
2	Jeter une pièce de monnaie	
3	Compter le nombre de personnes entrant dans un magasin entre 8h30 et 22h	
4	Jeter un dé deux fois de suite	
5	Jeter une pièce de monnaie trois fois de suite	
6	Observer la durée de vie d'une ampoule électrique	



# DÉFINITIONS

## Événement

- **Événement** : toute proposition logique associée aux résultats de l'expérience.
- On distingue les événements simples ou événements élémentaires qui sont constitués d'un seul élément ou des éléments composés:
- **Exemple** :  
Dans l'expérience N°4:
  - $A = \{(1; 2)\}$  est un événement simple.
  - $B = \{(1; 2); (1; 4); (5; 3)\}$  est un événement composé.
  - $C = \{\text{la somme des points obtenus est égale à } 4\}$  : Il est clair que  $C = \{(1; 3); (2; 2); (3; 1)\}$  est un événement composé.

# DÉFINITIONS

## Opération sur les événements (Opérateurs logiques)

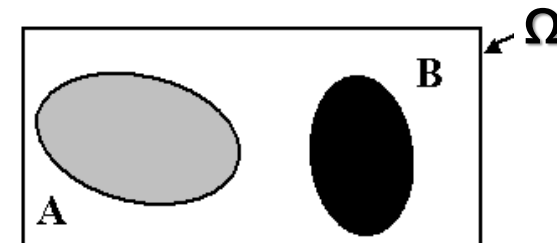
- Événement impossible :  $A = \emptyset$ ;
- Événement certain :  $A = \Omega$ ;
- Événement contraire :  $\bar{A} = \Omega \setminus A$ ;
- Événements incompatibles :  $A \cap B = \emptyset$ ;
- Réalisation simultanée de deux événements :  $A \cap B$ ;
- Réalisation d'un événement au moins :  $A \cup B$ ;
- Événement  $A \setminus B = A \cap \bar{B}$ . Cet événement est caractérisé par la réalisation de A et la non réalisation de B.

# DÉFINITIONS

Événement

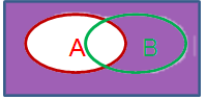
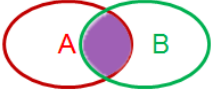
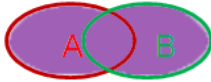
- **Représentation ensembliste** : Diagramme de Venn

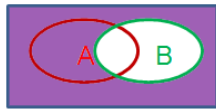
- $\Omega$  ensemble des événements possibles  
  A sous-ensemble de  $\Omega$   
  B sous-ensemble de  $\Omega$



# DÉFINITIONS

## Opération sur les événements (Opérateurs logiques)

- Événement impossible :  $A = \emptyset$ ;
- Événement certain :  $A = \Omega$ ;
- Événement contraire :  $\bar{A} = \Omega \setminus A$ ; 
- Événements incompatibles :  $A \cap B = \emptyset$ ;
- Réalisation simultanée de deux événements :  $A \cap B$ ; 
- Réalisation d'un événement au moins :  $A \cup B$ ; 
- Événement  $A \setminus B = A \cap \bar{B}$ . Cet événement est caractérisé par la réalisation de A et la non réalisation de B.

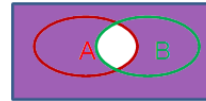


# DÉFINITIONS

## Opération sur les événements (Opérateurs logiques)

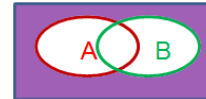
- Théorèmes de De Morgan:

- $\text{Non}(A \text{ et } B) = \text{Non}(A) \text{ ou } \text{Non}(B)$



$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

- $\text{Non}(A \text{ ou } B) = \text{Non}(A) \text{ et } \text{Non}(B)$



$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

**Attention:** La plupart des problèmes de probabilités n'ont comme difficulté que l'interprétation logique de l'énoncé.

# DÉFINITIONS

## ☐ Événements exclusifs

- ☐ Les événements A et B ne peuvent se produire simultanément
- ☐ Pour tous couples (A,B):  $A \cap B = \emptyset$



**La non-réalisation de l'un implique  
la réalisation de l'autre  
(Mort-Vivant)**

## ☐ Événements non exclusifs

- ☐ Les événements peuvent se produire simultanément .
- ☐ Pour tous couples (A,B):  $A \cap B \neq \emptyset$



### **Exemple:**

- Extraire une dame et un carreau
- Avoir un diabète et rouler avec des pneus lisses.
- Avoir un diabète et une angine.



**Ne pas confondre événements exclusifs ( $P(A \cap B) = 0$ ) et événements indépendants ( $P(A/B) = P(A)$  &  $P(B/A) = P(B)$ ).**

# PRINCIPES ÉLÉMENTAIRES

- $0 \leq P(A) \leq 1$  : Une probabilité est toujours comprise entre 0 et 1.
- $P(A) = 1$  : L'événement est toujours réalisé.
- $P(A) = 0$  : L'événement est impossible.

# PRINCIPES ÉLÉMENTAIRES

## Événement exclusifs

- Si 2 événements sont **exclusifs** :  $P(A \text{ ou } B) = P(A + B) = P(A \cup B) = P(A) + P(B)$

Exemple : Probabilité d'extraire un cœur ou un carreau =  $P(\text{Cœur ou Carreau}) = 0,25 + 0,25 = 0,5$ .

- Généralisation  $P(A+B+C) = P(A)+P(B)+P(C)$ .

- Si 2 événements sont **mutuellement exclusifs** (mort-vivant) et constituent l'ensemble des possibles :

$$P(A)+P(B) = 1 \Rightarrow P(A) = 1-P(B).$$

Exemple : Probabilité de survie à un moment donné est égale à 1 moins la probabilité de décéder à ce moment.



# PRINCIPES ÉLÉMENTAIRES

## Événement non exclusifs

- Les événements peuvent se produire simultanément.

Exemple : Probabilité d'avoir un infarctus du myocarde ou d'être diabétique

- $P(A \text{ ou } B) = P(B \text{ ou } A)$   
 $= P(A) + P(B) - P(A \text{ et } B)$
- Ceci se déduit des relations :
  - $P(A \text{ ou } B) = P(\text{A sans B}) + P(\text{B sans A}) + P(A \text{ et } B)$   
 $\quad \quad \quad \downarrow \quad \quad \quad \downarrow$   
 $\quad \quad \quad P(A) - P(A \text{ et } B) \quad \quad P(B) - P(A \text{ et } B)$

- En conclusion :
  - $P(A \text{ ou } B) \leq P(A) + P(B)$
  - $P(A \text{ ou } B \text{ ou } C) = P(A) + P(B) + P(C) - P(A \text{ et } B) - P(B \text{ et } C) - P(A \text{ et } C) - P(A \text{ et } B \text{ et } C)$



# PROBABILITÉS CONDITIONNELLES & INDÉPENDANCE

- En médecine, l'utilisation des probabilités conditionnelles est fréquente et apparaît naturelle.

## Exemple:

On dira que « un individu a 5 fois plus de chances de développer une maladie coronarienne s'il fume un paquet de tabac par jour que si il ne fume pas »...

# PROBABILITÉS CONDITIONNELLES & INDÉPENDANCE

## Probabilité conditionnelle

- Soit deux événements non exclusifs A et B :
  - On regarde la probabilité que l'un se réalise alors que l'autre est déjà réalisé.
- On note  $P(A/B)$  la probabilité de A si B est réalisé, l'inversement du conditionnement  $P(B/A)$  est la probabilité de B si A est réalisé:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B \cap A) = P(A/B) * P(B) = P(B/A) * P(A)$$

### Exemple:

- Quelle est la probabilité d'avoir une douleur de la fosse iliaque droite alors que l'on a une appendicite ?
- Quelle est la probabilité d'avoir une appendicite alors que je localise une douleur dans la fosse iliaque droite ?

# PROBABILITÉS CONDITIONNELLES & INDÉPENDANCE

- Deux événements sont indépendants si la réalisation de l'un n'influence pas la réalisation de l'autre. **Indépendance**
- $P(A/B) = P(A \cap B)/P(B)$
- $P(A/B) = P(A)$   
→  $P(A \cap B) = P(A) * P(B)$

## Exemple:

- Événements indépendants : Pluie, rouler avec des pneus lisses
- Événements non indépendants : Pluie, avoir un accident
- La connaissance n'est pas figée :
  - Avant la réalisation d'un test, la probabilité d'une maladie est p.
  - Que devient-elle si on sait que le test est positif ?

# INVERSION DU CONDITIONNEMENT

Indépendance

- Intérêt : évaluation des examens complémentaires.

## Théorème de Bayes :

$$P(A / B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A / B) * P(B)$$

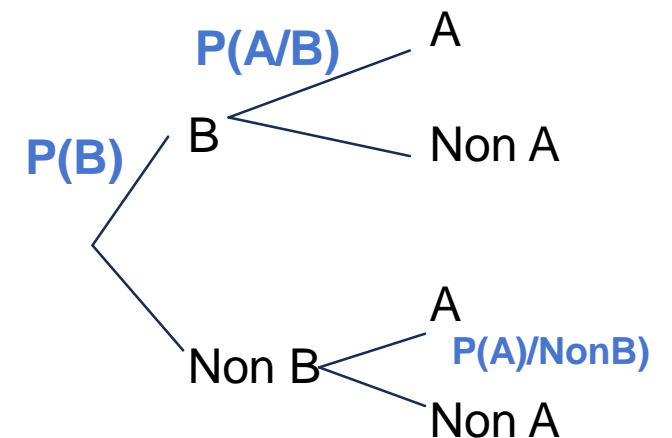
$$P(B / A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A / B) * P(B)}{P(A)}$$

$$P(\text{Non } B) = 1 - P(B)$$

$$P(A / \text{Non } B) = 1 - P(\text{Non } A / \text{Non } B)$$

$$P(A) = P(A / B) * P(B) + P(A / \text{Non } B) * P(\text{Non } B)$$

$$P(A) = P(A / B) * P(B) + [1 - P(\text{Non } A / \text{Non } B)] * [1 - P(B)]$$



# ARBRES DE DÉCISION

Un homme souffrant d'une douleur dans la région cardiaque se présente aux urgences. Quelle est la probabilité qu'il ait une sténose coronarienne ?

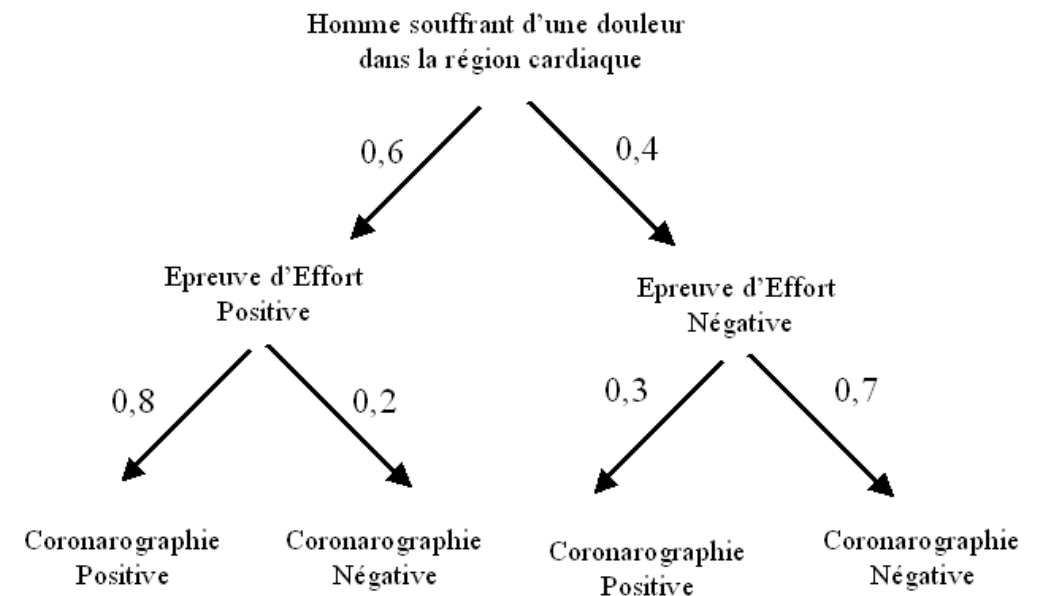
Sachant que :

- Le patient doit faire une épreuve d'effort avant la coronarographie
- $P(\text{EE}+) = 0,7$
- $P(\text{Coro-}/\text{EE}+) = 0,15$
- $P(\text{Coro+}/\text{EE-}) = 0,2$

# ARBRES DE DÉCISION

- Un homme se présente aux urgences. Quelle est la probabilité qu'il ait une sténose coronarienne ?

$$\begin{aligned} P(\text{Coro+}) &= P(\text{EE+} \cap \text{Coro+}) + P(\text{EE-} \cap \text{Coro+}) \\ &= 0,8 \times 0,6 + 0,4 \times 0,3 \\ &= 0,48 + 0,12 \\ &= \mathbf{0,60} \end{aligned}$$



# TABLEAU DE CONTINGENCE

- En médecine, 2 tableaux à 4 cases sont très utilisés et renvoient au conditionnement  
→ **Tableau croisé ou tableau de contingence**

**Objectif:** -Evaluation des signes et examens complémentaires.  
-Recherche de facteurs de risque.

	Maladie +	Maladie -	Total
Test +	A (VP)	B (FP)	$T+ = A+B = VP+FP$
Test -	C (FN)	D (VN)	$T- = C+D = FN+VN$
Total	$M+ = A+C = VP+FN$	$M- = B+D = FP+VN$	n

$P(T+/M+)$ ;  $P(T-/M-)$ ;  $P(M+/T+)$ ;  $P(M-/T-)$

	Maladie +	Maladie -	Total
Exposés	A	B	$E+ = A+B$
Non exposés	C	D	$E- = C+D$
Total	$M+ = A+C$	$M- = B+D$	$A+B+C+D$

$P(M+/E+)$ ;  $P(M+/E-)$ ;  $P(E+/M+)$ ;  $P(E-/M-)$



# EVALUATION DES EXAMENS COMPLÉMENTAIRES

- Il n'existe pas de signe ou d'examen parfait qui serait toujours présent en cas de présence de la maladie et absent en cas d'absence de la maladie.

	Maladie +	Maladie -	Total
Test +	A (VP)	B (FP)	$T+ = A+B = VP+FP$
Test -	C (FN)	D (VN)	$T- = C+D = FN+VN$
Total	$M+ = A+C = VP+FN$	$M- = B+D = FP+VN$	$N = A+B+C+D$

→ La **prévalence** est la probabilité d'avoir la maladie avant d'avoir fait le test : **Probabilité pré-test** (Fréquence de la maladie).

$$\text{Prévalence} = P(M+) = (VP+FN)/N$$

# EVALUATION DES EXAMENS COMPLÉMENTAIRES

	Maladie +	Maladie -	Total
Test +	A (VP)	B (FP)	$T+ = A+B = VP+FP$
Test -	C (FN)	D (VN)	$T- = C+D = FN+VN$
Total	$M+ = A+C = VP+FN$	$M- = B+D = FP+VN$	$N = A+B+C+D$

## ■ Caractéristiques intrinsèques du test:

- Elles supposent le problème résolu puisqu' un test de référence (gold standard) a permis de déterminer si la personne était malade ou non.
- Elles sont **influencées** notamment par le **stade évolutif de la maladie**.

→ La **Sensibilité** : probabilités des tests positifs chez les malades

$$\text{Sensibilité} = P(T+/M+) = VP/(VP+FN)$$

→ La **Spécificité** : probabilités des tests négatifs chez les non malades

$$\text{Spécificité} = P(T-/M-) = VN/(VN+FP)$$

# EVALUATION DES EXAMENS COMPLÉMENTAIRES

	Maladie +	Maladie -	Total
Test +	A (VP)	B (FP)	T+=A+B=VP+FP
Test -	C (FN)	D (VN)	T-=C+D=FN+VN
Total	M+ = A+C = VP+FN	M-=B+D=FP+VN	N=A+B+C+D

- La **valeur prédictive négative** :VPN = probabilité de ne pas avoir la maladie si le test est négatif.
- La **valeur prédictive positive** :VPP = probabilité d'avoir la maladie si le test est positif (probabilité post-test);

$$\text{VPN} = P(M-/T-) = VN/(VN+FN)$$

$$\text{VPP} = P(M+/T+) = VP/(VP+FP)$$

→ servent à la décision médicale.

- Si le généraliste utilise la positivité du test pour adresser les sujets au spécialiste, la fréquence de la maladie (prévalence) dans le groupe adressé au spécialiste sera la VPP.

# EVALUATION DES EXAMENS COMPLÉMENTAIRES

- Les valeurs prédictives dépendent de :
  - La sensibilité du test,
  - La spécificité du test,
  - La prévalence du test.

→ le même test (même sensibilité et spécificité) aura des VPP et VPN très différentes en fonction de la prévalence de la maladie.

# EVALUATION DES EXAMENS COMPLÉMENTAIRES

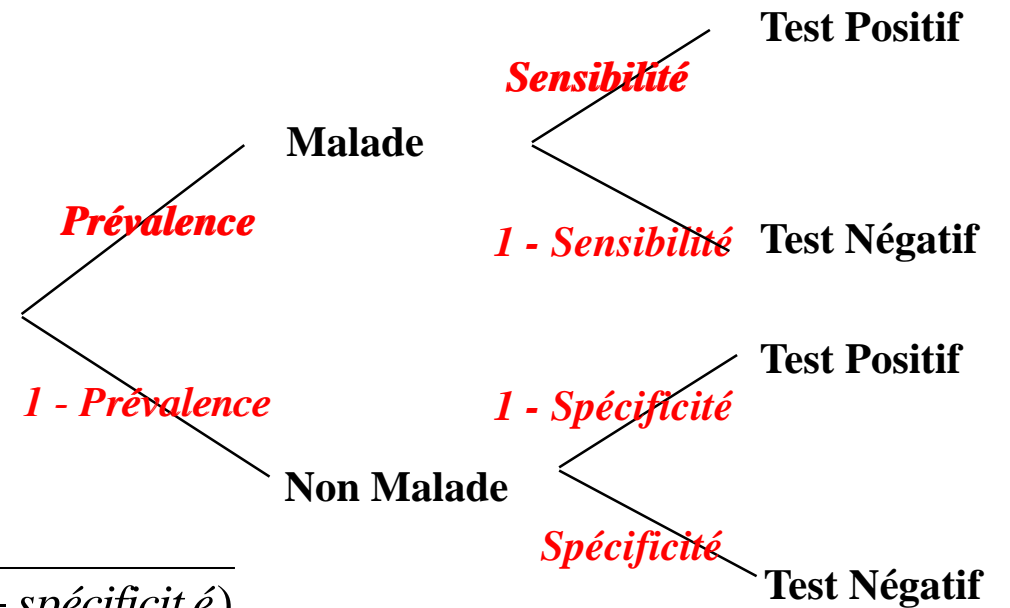
- VPP et VPN correspondent à l'inversion du conditionnement de la sensibilité et de la spécificité.
- L'arbre des probabilités permet facilement cette opération.

$$\text{VPP} = P(M+/T+) = \text{VP}/(\text{VP}+\text{FP})$$

$$\Rightarrow \text{VPP} = \frac{\text{prévalence} * \text{sensibilité}}{(\text{prévalence} * \text{sensibilité}) + (1 - \text{prévalence}) * (1 - \text{spécificité})}$$

$$\text{VPN} = P(M-/T-) = \text{VN}/(\text{VN}+\text{FN})$$

$$\Rightarrow \text{VPN} = \frac{(1 - \text{prévalence}) * (\text{spécificité})}{(1 - \text{prévalence}) * (\text{spécificité}) + (\text{prévalence}) * (1 - \text{sensibilité})}$$



# CORRECTION DE L'EXAMEN (SESSION 1)

## Exercice I

Dans le cadre de la mise au point d'un alcootest, on a observé les résultats suivants :

	Sujets ivres	Sujets sains
Alcootest positif	150	18
Alcootest négatif	10	322

Les sujets ayant été classés en sujets ivres et sains d'après leur alcoolémie (c'est-à-dire leur concentration sanguine en alcool).

1. Donner la fréquence des réponses « faux positifs » et des réponses « faux négatifs ».
2. Calculer la sensibilité et la spécificité de ce test.
3. Calculer les valeurs prédictives positive et négative du test lorsqu'il est appliqué :
  - A une population où la probabilité a priori de trouver un sujet ivre est  $P = 0,01$  ;
  - Un soir de réveillon, à une population de sujets dont la probabilité a priori d'être ivres est  $P = 0,20$ .
4. Que pouvez-vous conclure par rapport à la robustesse de l'Alcootest.