



Pré-rentrée 2021-2022

UE4
Introduction aux Biostatistiques



Tuteurs : Anna, Aubert, Cyprien, Justine

Présentation de l'UE4

- ❖ “Matière à calcul” : **cours magistraux** + enseignements dirigés (**ED**) en groupe
- ❖ Epreuve d'UE4 :
 - Coef **4** / Durée = **1h30** / L'année dernière : **QROC** (≈10) + **QCM** (≈ 20)
 - Sujet : questions de cours et résolutions d'exercices
- ❖ UE4 au tutorat :
 - QCMs en ligne, conférences communes (septembre)
 - MCCB
 - 2 conférences ± 1 conférence bonus (révisions)
 - Concours blanc
 - Fiches récap

Programme de l'UE4

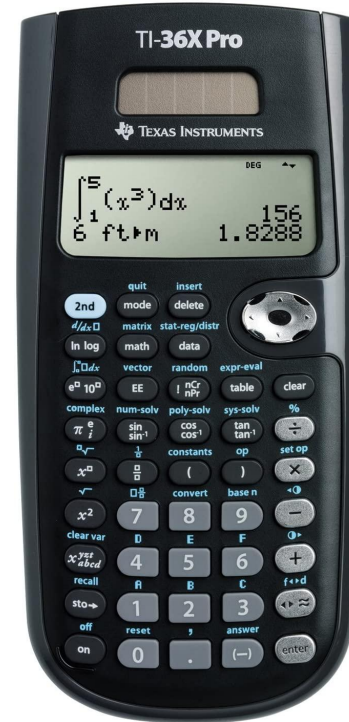
- ❖ **Rappels et probabilité** *pré-rentrée*
 - dérivées, intégrales, analyses combinatoires
- ❖ **Variables aléatoire, lois**
 - Variables discrètes et continues, lois uniformes / binomiale / de Bernoulli / normale ...
- ❖ **Statistiques descriptives**
 - paramètres de position / dispersion, représentation graphique
- ❖ **Estimations**
 - sondages, intervalles de confiance
- ❖ **Tests statistiques**
 - principes des tests, comparaison de moyennes, comparaison de fréquences
- ❖ **Corrélation et régression linéaire**
- ❖ **Epidémiologie**
 - indicateurs, enquêtes épidémiologiques

Conseils pour travailler l'UE4

- ❖ Apprendre par coeur : **formules par <3** , définitions importantes
 - **Ne pas apprendre les démonstrations !**
- ❖ **Entraînement** aux QCMs :
 - Conférences du tutorat et conférences classées
 - **Télécharger les versions de cette année !!!**
 - Annales MCCB, CCB et CC
- ❖ Supports à utiliser :
 - Calculatrice
 - Fiches bonus du tutorat
 - Forum
 - Titre de votre publication doit comprendre : type de sujet + année + n° QCM + photo du sujet + photo de la correction
- ❖ Prendre de **l'avance** avec tous les **outils du tutorat** car les derniers chapitres nécessitent de l'entraînement.

Calculatrice

On vous recommande vivement la **TI - 36 X Pro** !



Chapitre 1 : rappels et probabilités

- ❖ Dérivées et intégrales
- ❖ Analyse combinatoire
- ❖ Applications des probabilités : diagnostique thérapeutique



Première partie : Dérivées et intégrales

UE4
Introduction aux Biostatistiques



Tuteurs : Anna, Aubert, Cyprien, Justine

1) Rappels du lycée

- ❖ Acquis nécessaires en PASS :
 - Dérivées et propriétés
 - Etude de fonction

- ❖ Retenir la **fiche “dérivées et primitives usuelles”** (dispo sur site du tutorat)

Dérivées des fonctions usuelles

Dérivées des fonctions usuelles

Dans chaque ligne, f' est la dérivée de la fonction f sur l'intervalle I .

$f(x)$	I	$f'(x)$
λ (constante)	\mathbb{R}	0
x	\mathbb{R}	1
x^n ($n \in \mathbb{N}^*$)	\mathbb{R}	nx^{n-1}
$\frac{1}{x}$	$] -\infty, 0[$ ou $] 0, +\infty[$	$-\frac{1}{x^2}$
$\frac{1}{x^n}$ où $n \in \mathbb{N}$, $n \geq 2$	$] -\infty, 0[$ ou $] 0, +\infty[$	$-\frac{n}{x^{n+1}}$
\sqrt{x}	$] 0, +\infty[$	$\frac{1}{2\sqrt{x}}$
$\ln x$	$] 0, +\infty[$	$\frac{1}{x}$
e^x	\mathbb{R}	e^x
$\sin x$	\mathbb{R}	$\cos x$
$\cos x$	\mathbb{R}	$-\sin x$
$\tan x$	$] -\frac{\pi}{2} + k\pi, \frac{\pi}{2} + k\pi [$, $k \in \mathbb{Z}$	$1 + \tan^2 x = \frac{1}{\cos^2 x}$

Primitives des fonctions usuelles

Primitives des fonctions usuelles

Dans chaque ligne, F est une primitive de f sur l'intervalle I . Ces primitives sont uniques à une constante près notée C .

$f(x)$	I	$F(x)$
λ (constante)	\mathbb{R}	$\lambda x + C$
x	\mathbb{R}	$\frac{x^2}{2} + C$
x^n ($n \in \mathbb{N}^*$)	\mathbb{R}	$\frac{x^{n+1}}{n+1} + C$
$\frac{1}{x}$	$] -\infty, 0[$ ou $] 0, +\infty[$	$\ln x + C$
$\frac{1}{x^n}$ où $n \in \mathbb{N}$, $n \geq 2$	$] -\infty, 0[$ ou $] 0, +\infty[$	$-\frac{1}{(n-1)x^{n-1}} + C$
$\frac{1}{\sqrt{x}}$	$] 0, +\infty[$	$2\sqrt{x} + C$
$\ln x$	\mathbb{R}_+^*	$x \ln x - x + C$
e^x	\mathbb{R}	$e^x + C$
$\sin x$	\mathbb{R}	$-\cos x + C$
$\cos x$	\mathbb{R}	$\sin x + C$
$1 + \tan^2 x = \frac{1}{\cos^2 x}$	$] -\frac{\pi}{2} + k\pi, \frac{\pi}{2} + k\pi[$, $k \in \mathbb{Z}$	$\tan x + C$

Opérations et dérivées

Opérations et dérivées

$$(Cf)' = Cf'$$

$$(f/g)' = (f'g - fg') / (g^2)$$

$$\ln(f(x))' = f'(x)/f(x)$$

$$(f+g)' = f' + g'$$

$$(f \cdot g)'(x) = g'(x) \cdot f'(g(x))$$

$$\tan(f(x))' = f'(x) / (1 + \tan^2(f(x)))$$

$$(f \cdot g)' = f'g + fg'$$

$$(\sin f(x))' = f'(x) \cdot \cos(f(x))$$

$$= f'(x) / \cos^2(f(x))$$

$$(1/g)' = - (g'/g^2)$$

$$(\cos f(x))' = -f'(x) \cdot \sin(f(x))$$

$$\arccos(f(x))' = -f'(x) / \sqrt{1 - f(x)^2}$$

$$(f^m)' = m \cdot f' \cdot f^{m-1}$$

$$\exp(f(x))' = f'(x) \cdot \exp(f(x))$$

$$\arcsin(f(x))' = f'(x) / \sqrt{1 - f(x)^2}$$

$$\arctan(f(x))' = f'(x) / (1 + f(x)^2)$$

2) Dérivées partielles

Utilisée dans le cas d'une fonction avec **plusieurs variables**.

Ex : $f(x,y) = 2x + y^2$

Dans ce cas, on peut **dériver f selon x** (notation : df/dx) ou selon y (df/dy).
L'autre variable est alors considérée comme étant une **constante**.

Ex : $df/dx = 2 + y^2$ (y = constante)

$df/dy = 2x + 2y$ (x = constante)

⇒ Utile après pour le **calcul d'erreur**

3) Calcul d'erreur

❖ Notation :

- X = valeur mesurée (estimation de x)
- x = valeur réelle (jamais connue)

❖ Erreur **absolue** = ΔX , exprimée dans **l'unité** de X

❖ Erreur **relative** = $\Delta X/X$, exprimée **sans unité**

❖ Calcul d'erreur : $X = f(A, B, C, \dots)$ et $\Delta A, \Delta B, \Delta C, \dots$ les erreurs absolues des mesures de A, B, C, \dots

$$\Delta X = \left| \frac{\partial f}{\partial A} \right| \Delta A + \left| \frac{\partial f}{\partial B} \right| \Delta B + \left| \frac{\partial f}{\partial C} \right| \Delta C + \dots$$



3) Calcul d'erreur

$$\Delta X = \left| \frac{\partial f}{\partial A} \right| \Delta A + \left| \frac{\partial f}{\partial B} \right| \Delta B + \left| \frac{\partial f}{\partial C} \right| \Delta C + \dots$$

Ex : mesure du volume d'un parallélépipède dont longueur = **6cm**, largeur = **4cm** et hauteur = **5cm**. Les erreurs absolues de chaque mesure sont $\Delta L = 0,5\text{cm}$ et $\Delta l = \Delta h = 0,2\text{cm}$.

$$V = L \text{ (longueur)} * l \text{ (largeur)} * h \text{ (hauteur)}$$

On cherche $df/dL \rightarrow$ La dérivée de f en fonction de $L \rightarrow$ **l et h** sont des **constantes** $\Rightarrow df/dL = l * h$.

En appliquant la formule on obtient l'**erreur absolue de V** :

$$\Delta V = l * h * \Delta L + L * h * \Delta l + L * l * \Delta h = 4 * 5 * 0,5 + 6 * 5 * 0,2 + 6 * 4 * 0,2 = 20,8 \text{ cm}^3$$

Donc l'**erreur relative** de la mesure du volume est :

$$\Delta V / V = 20,8 / (5 * 4 * 6) = 20,8 / 120 = 0,1733 \text{ soit environ } 17\%$$

4) Densité de probabilité

La fonction f est une densité de probabilité sur l'intervalle $I=[a;b]$ si et seulement s'il remplit ces deux conditions

- **f est continue sur I**
- **f est positive ou nulle sur I**
- $\int_a^b f(x)dx = 1$

Ex : soit $f(x)=6x/c$ (c =constante) sur l'intervalle $I=[1;2]$. f est nulle si $x<1$ ou $x>2$ et est une densité de probabilité. On souhaite déterminer c .


$$\int_1^2 f(x)dx = 1 = \int_1^2 6x/c \quad \text{soit} \quad \left[\frac{6}{c} * \frac{x^2}{2}\right]_1^2 = \frac{6 * 2^2}{2c} - \frac{6 * 1^2}{2c} = \frac{9}{c} = 1 \quad \text{donc } c=9.$$

4) Densité de probabilité

Ex : soit $f(x)=6x/c$ (c =constante) sur l'intervalle $I=[1;2]$. f est nulle si $x<1$ ou $x>2$ et est une densité de probabilité. On souhaite déterminer c .

Ou à la calculatrice : on peut tester les différentes propositions **seulement si c'est un QCM** donc il faut apprendre la méthode sans calculatrice.

- Faire $\int_0^1 \square dx$ puis essayer toutes les possibilités



Deuxième partie :
**Analyse
combinatoire**

UE4
Introduction aux Biostatistiques



Tuteurs : Anna, Aubert, Cyprien, Justine

1) Introduction à l'analyse combinatoire

Analyse combinatoire : étudie comment **compter** les objets

3 types :

- Arrangements
- Permutations
- Combinaisons

2) Arrangements


Soit un ensemble E de n objets. Toutes **suites ordonnées de p objets parmi les n objets** sont appelées **arrangements**.

Remarque : on a toujours $1 \leq p \leq n$.

2) Arrangements

2.a - Arrangements sans répétition (ou "sans remise")

Notation : A_n^p

Formule : $A_n^p = n(n-1)\dots(n-p+1)$ soit $A_n^p = \frac{n!}{(n-p)!}$ 


Ex : les arrangements possibles du **tiercé** dans l'ordre d'une course comptant **20 chevaux** sont obtenus en utilisant **p=3** et **n=20**. Un cheval ne peut pas être sur 2 places différentes donc **sans rep**

$$A_{20}^3 = 20 \times 19 \times 18 = \frac{20!}{(20-3)!} = 6840$$

2) Arrangements

2.b - Arrangements avec répétitions (ou "avec remise")

Notation : L_n^p

Formule : $L_n^p = n \times n \times n \times \dots \times n = n^p$ soit $L_n^p = n^p$ 

Ex : une séquence d'ADN est formée par un enchaînement de nucléotides (**A,C,G,T**). Pour former un **dinucléotide**, il y a **4 possibilités** pour le premier et **4** aussi pour le second. **Sans rep** car on peut les **réutiliser** !

⇒ **n = 4** et **p = 2**

Donc $L = 4 \times 4 = 4^2 = 16$

3) Permutations

Soit un ensemble E de n objets. Toutes **suites ordonnées de ces n objets** sont appelées permutations.

Remarque : Une **permutation de n objet** correspond à un **arrangement de n objets parmi n objets** (soit $p=n$).

3) Permutations

3.a - Permutations sans répétition

Notation : P_n

Formule :

$$P_n = n!$$



Cas des **permutations circulaires**

$$P_n = \frac{n!}{n} = (n-1)!$$




Ex : Julie souhaite accrocher **6 cadres** différents sur son mur. Le nombre de permutations possibles est :

n = 6 ; pas de rep ; $P(6) = 6*5*4*3*2*1 = 6! = 720$

3) Permutations

3.b - Permutations avec répétitions

Notation : P_n

Formule :
$$P_n = \frac{n!}{k_1! k_2! \dots k_r!}$$
 

Ex : on cherche le nombre de mot (avec ou sans signification) que l'on peut écrire avec les **lettres** BIOSTATISTIQUES.

⇒ **15 lettres** ⇒ $n = 15$; **3 lettres I** ; **3 lettres S** ; **3 lettres T**

$P(15) = 15! / (3! * 3! * 3! * *) = 6\ 054\ 048\ 000$ mots

4) Combinaisons


Soit un ensemble E de n objets. Tout **ensemble de p objets parmi n objets** est une combinaison.

On ne tient donc **pas compte de l'ordre** des objets !

4) Combinaisons

4.a - Combinaisons sans répétition

Notation : C_n^p

Formule :
$$C_n^p = \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!} = \binom{n}{p}$$
 


Ex : parmi **34** étudiants, **4** sont tirés au sort. On cherche combien de groupes différents on peut créer. **n = 34** et **p = 4**. **Sans rep** car on ne peut pas tirer 2 * le même étudiant

$$C_{34}^4 = 34! / (4! * 30!) = 46\ 376 \text{ possibilités}$$

4) Combinaisons

4.b - Combinaisons avec répétitions

Notation : C_{n+p-1}^p

Formule : $C_{n+p-1}^p = \frac{(n+p-1)!}{p!(n-1)!}$ 

Ex : on dispose de **12** comprimés, on souhaite savoir combien de traitements différents sont possibles sachant qu'un traitement est composé de **4** comprimés, **qu'ils soient différents ou non**. **n=12** et **p=4**

$C_{12+4-1}^4 = 15! / (4! * 11!) = 1\ 365$ possibilités de traitement

4) Combinaisons

4.c - Propriétés des combinaisons

$$C_n^0 = C_n^n = 1$$

$$C_n^1 = C_n^{n-1} = n \quad \text{si } n \geq 1$$

$$C_n^2 = C_n^{n-2} = \frac{n(n-1)}{2} \quad \text{si } n \geq 2$$

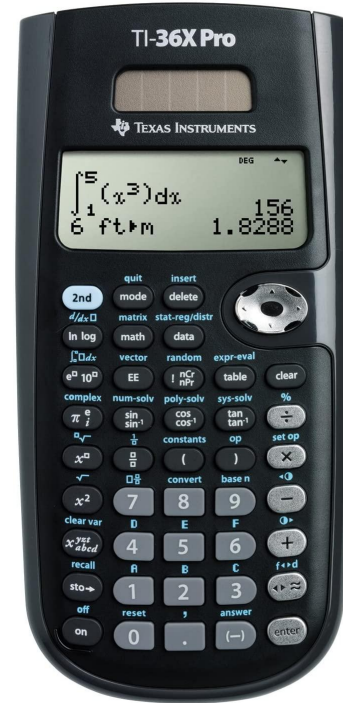
$$\boxed{C_n^p = C_n^{n-p}} \quad \text{si } 0 \leq p \leq n$$

$$\boxed{C_{n-1}^{p-1} + C_{n-1}^p = C_n^p} \quad \text{si } 0 \leq p \leq n-1$$

4) Combinaisons


4.d - Calculatrice

- Pour faire une **combinaison sans rep** :
taper « n » → nCr → « p » → enter.



Résumé des formules d'analyse combinatoire

		SANS REPETITION	AVEC REPETITIONS
ARRANGEMENT	Suite ordonnée et p objets parmi n objets	$A_n^p = \frac{n!}{(n-p)!}$	$L_n^p = n^p$
COMBINAISON	Ensemble de p objets parmi n objets	$C_n^p = \frac{n!}{p!(n-p)!} = \frac{A_n^p}{p!} = \binom{n}{p}$	$C_{n+p-1}^p = \frac{(n+p-1)!}{p!(n-1)!}$
PERMUTATION	Suite ordonnée de n objets (parmi n objets)	$P_n = n!$	$P_n = \frac{n!}{k_1!k_2!\dots k_r!}$



Troisième partie :
Probabilités
Diagnostic thérapeutique

UE4
Introduction aux Biostatistiques



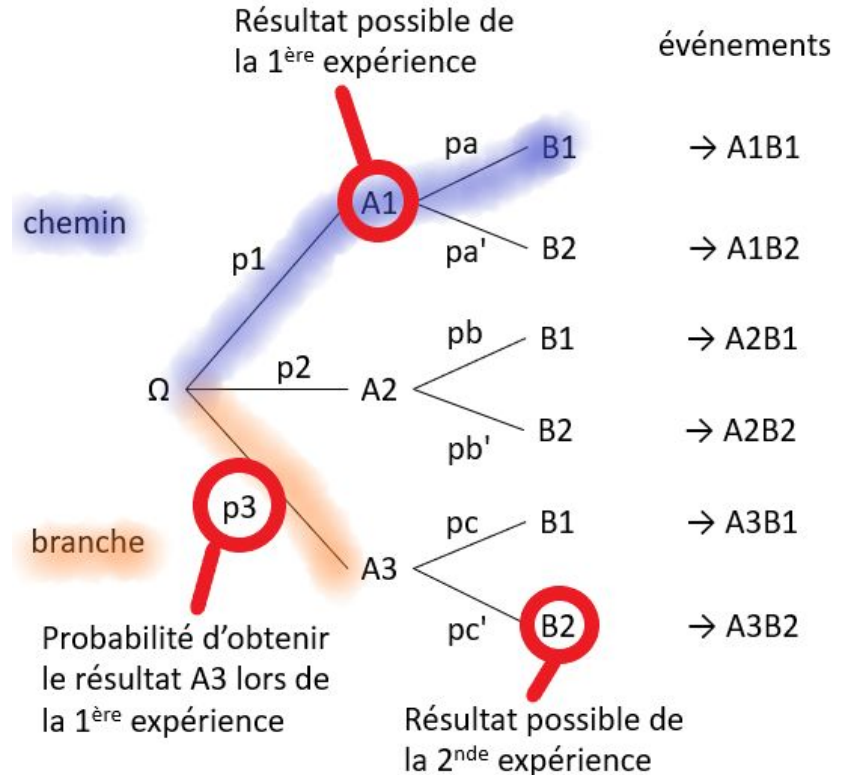
Tuteurs : Anna, Aubert, Cyprien, Justine

1) Calcul de probabilités : arbre de décision

Règles de calcul en utilisant l'arbre :

- la probabilité d'un événement est obtenue en **multipliant** les branches du chemin menant à celui-ci
- s'il existe plusieurs chemins menant à l'événement, on **additionne** les probabilités de chaque chemin

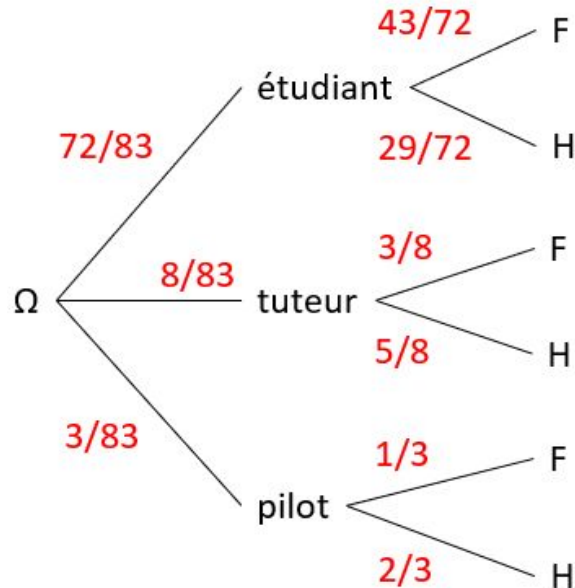
En cas de difficulté en proba, toujours revenir à l'arbre de décision !



1) Calcul de probabilités : arbre de décision

Ex : dans un amphi lors d'une conférence du tutorat, on trouve 72 étudiants de 1ère année dont 43 femmes, 8 tuteurs dont 3 femmes et 3 pilotes dont 1 femme.

Arbre de décision :



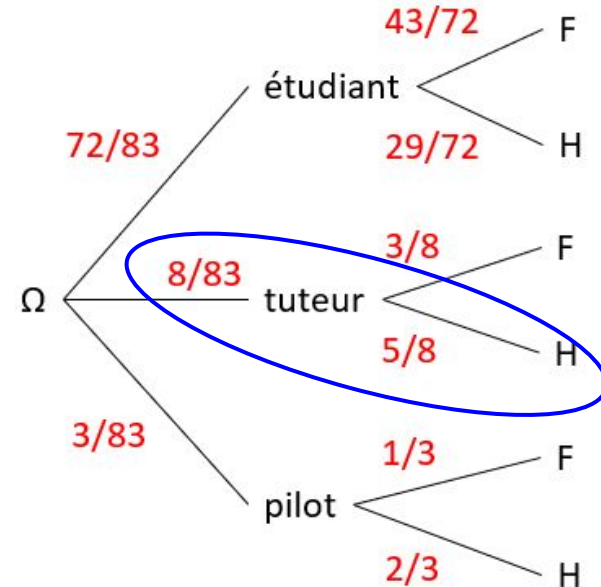
1) Calcul de probabilités : arbre de décision

Ex : en choisissant une personne au hasard, quelle est la probabilité d'obtenir un **tuteur (homme)** ?

$$P(\text{tuteur} \cap \text{homme}) = \mathbf{P(T) * P(H|T)}$$

Cette formule utilise la loi de Bayes qui sera vue dans le cours des probabilités, mais la question peut être résolue grâce à l'arbre.

$$P(T \cap H) = 8/83 * 5/8 = 0,06 = \mathbf{6\%}$$



1) Calcul de probabilités : arbre de décision

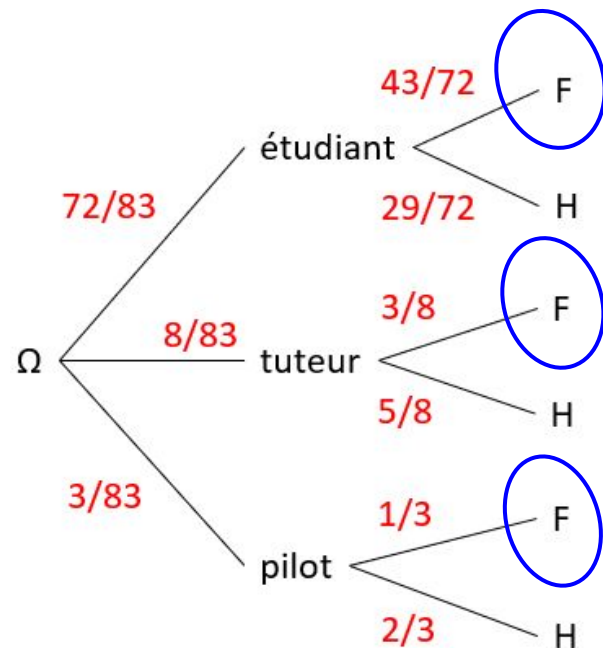
Ex : en choisissant une personne au hasard, quelle est la probabilité d'obtenir une **femme** ?

$$P(\text{femme}) = P(F \cap E) + P(F \cap T) + P(F \cap P)$$

Ici encore une formule existe et sera vue dans le cours des probabilités, mais la question peut être résolue grâce à l'arbre.

$$P(F) = 72/83 * 43/72 + 8/83 * 3/8 + 3/83 * 1/3$$

$$P(F) = 47/83 = 0,56$$

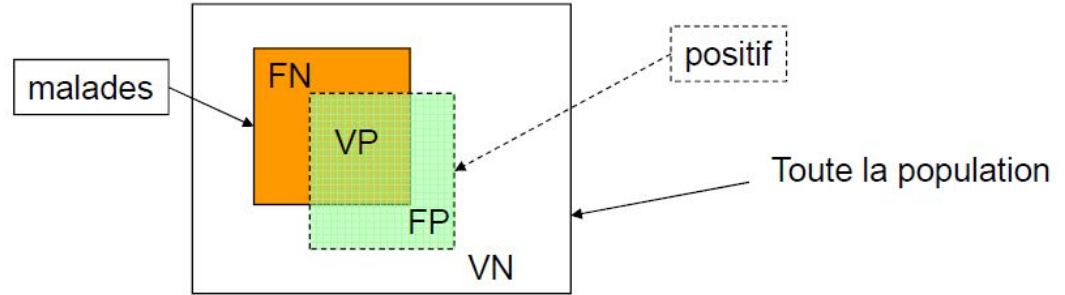


2) Diagnostic thérapeutique

2.a - Définitions

- ❖ Test *gold standard*
- ❖ Test alternatif

- ❖ Vrai/Faux positif
- ❖ Vrai/Faux négatif



		Résultat du gold standard		
		M	\bar{M}	
Résultat du test	T+	VP	FP	Total T+
	T-	FN	VN	Total T-
		Total M	Total \bar{M}	Total pop

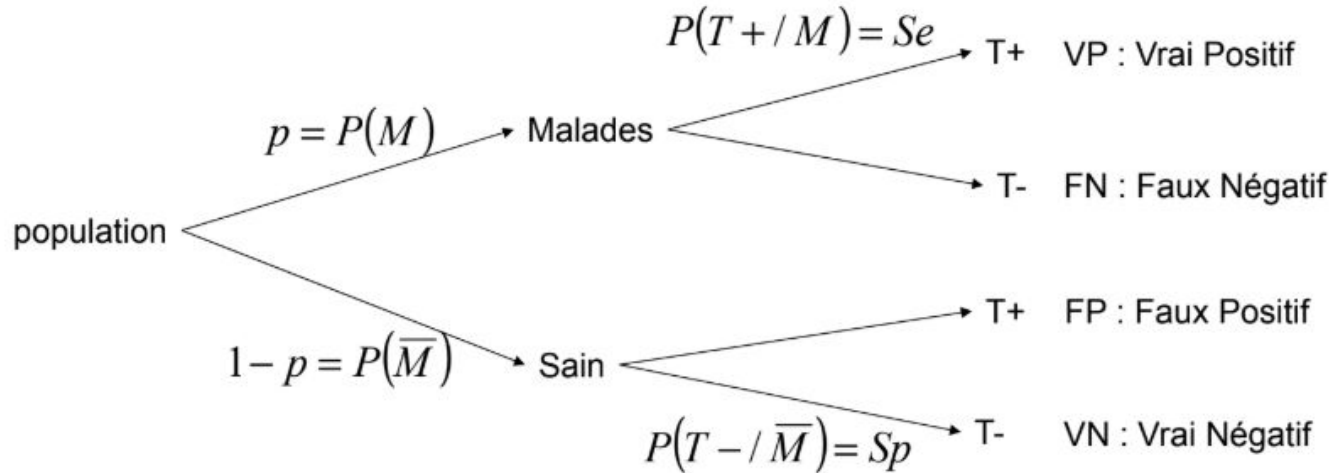
2) Diagnostic thérapeutique

2.a - Définitions

- ❖ **Prévalence p** (maladie) : proba pour un sujet d'être **malade**
- ❖ **Sensibilité Se** (test) : probabilité d'un sujet **malade d'être positif**
- ❖ **Spécificité Sp** (test) : probabilité d'un sujet **sain d'être négatif**
- ❖ **Valeur prédictive positive** (test) : probabilité d'un individu **ayant un test positif** d'être en effet **malade**
- ❖ **Valeur prédictive négative** (test) : probabilité d'un individu **ayant un test négatif** d'être en effet **sain**

2) Diagnostic thérapeutique

2.b - Résumé des définitions dans un arbre



2) Diagnostic thérapeutique

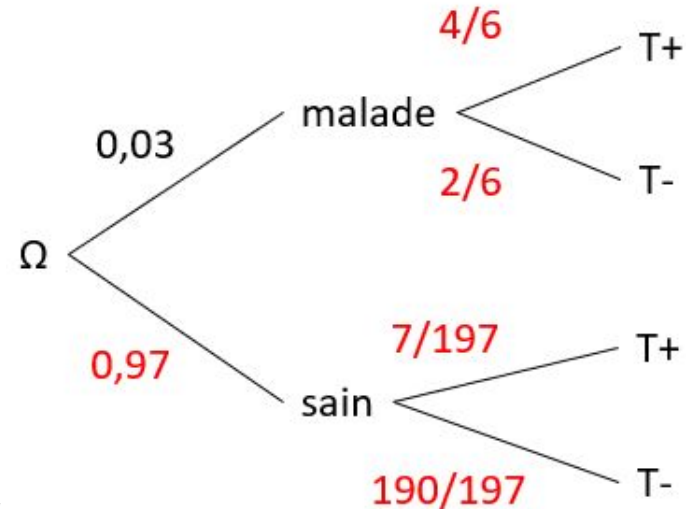
Ex : des chercheurs ont développé un test visant à diagnostiquer une infection virale par analyse salivaire. La **prévalence** de cette maladie dans la population est de **3%**.

On souhaite évaluer les paramètres de ce test, pour cela on dispose d'un test de référence (que l'on considère infaillible). Sur les **203** patients de l'étude, on obtient **11 tests positifs**, parmi lesquels seulement **4 sont des vrais positifs**. La prévalence de cette maladie dans l'échantillon est la même que dans la population.

2) Diagnostic thérapeutique

Ex : on commence par rédiger le tableau et/ou l'arbre.

		Test de référence		TOTAL
		M+	M-	
Test à évaluer	T+	4	7	11
	T-	2	190	192
TOTAL		6 = 0,03*203	197	203



Puis les questions principales sont de déterminer les valeurs des principaux paramètres (VN, VP, VPP...).

MERCI & BON COURAGE

Des **QCMs** portant sur les points de cours vus dans ce diapo seront en ligne ce **soir**, la **correction** aura lieu **demain**.

N'hésitez pas au début à faire les exercices en regardant les formules (c'est plus simple d'apprendre une fois qu'on a compris).