UE4
Biostatistiques
PASS - 2025/26

TUTORAT RENNAIS

Nous vous rappelons que pour chacune des UE PASS/L.AS, les évaluations portent uniquement sur les éléments produits par les enseignants lors de l'année universitaire en cours (diaporamas projetés pendant les cours, informations données oralement, supports déposés sur moodle, ...).

Ainsi, ces livrets, mêmes s'ils sont basés sur les cours des enseignants, sont réalisés par nos équipes de tuteurs et tutrices : les erreurs sont possibles. Ce livret élaboré par le tutorat n'est pas opposable en cas de litige.

SOMMAIRE

SOMMAIRE 3

INTRODUCTION AUX BIOSTATISTIQUES 6

METHODOLOGIE - BIOSTATISTIQUES 8

FICHE CALCULATRICE 9

NOTES PERSONNELLES 13

RAPPELS 14

FONCTIONS, DERIVEES ET PRIMITIVES 14

CALCUL D'ERREUR 16

ANALYSES COMBINATOIRES 18

NOTES PERSONNELLES 22

PROBABILITES 23

NOTION D'EVENEMENTS 23

CALCUL DES PROBABILITES 23

PROPRIETES DES PROBABILITES 24

PROBABILITES CONDITIONNELLES +++ ET THEOREME DE BAYES 24

DIAGNOSTIQUE THERAPEUTIQUE +++ 25

NOTES PERSONNELLES 26

BIOSTATISTIQUES DESCRIPTIVES 27

VARIABLES QUALITATIVES ET VARIABLES QUANTITATIVES : 27

MESURES EN STATISTIQUES 28

NOTES PERSONNELLES 30

VARIABLES ALEATOIRES ET LOIS 31

OPERATIONS SUR ESPERANCE & VARIANCE 31

VARIABLES ALEATOIRES ET LOIS 36

NOTES PERSONNELLES 37

LOIS CONTINUES 38

CARACTERISTIQUES DES VARIABLES ALEATOIRES CONTINUES : 38

STABILITE DE LA LOI NORMALE : 38

COMMENT UTILISER LE THEOREME CENTRALE LIMITE : 39

ESTIMATIONS 41

EXERCICE CORRIGE : 44

EXERCICE CORRIGE : 47

NOTES PERSONNELLES 48

TEST STATISTIQUE 49

QU'EST-CE QU'UN TEST STATISTIQUE 49

PRINCIPES DES TESTS STATISTIQUES 52

NOTES PERSONNELLES 54

TESTS DE COMPARAISON DE MOYENNES 55

NOTES PERSONNELLES 60

TEST DE COMPARAISON DE FREQUENCES 61

EXERCICE TYPE CORRIGE : COMPARAISON DE FREQUENCE 62

NOTES PERSONNELLES 66

CORRELATIONS ET REGRESSIONS 67

CORRELATION : EVALUER LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES 67

REGRESSION LINEAIRE : VOIR SI LA COURBE DES VARIATIONS DE Y EN FONCTION DE X EST UNE DROITE 68

NOTION DE COVARIANCE : 69

POINTS IMPORTANTS A MAITRISER : 70

EXERCICE TYPE CORRIGE : CORRELATIONS REGRESSIONS 71

METHODE DE RESOLUTION : 72

NOTES PERSONNELLES 76

EPIDEMIOLOGIE 77

MESURES EN EPIDEMIOLOGIE : 77

INDICATEURS EPIDEMIOLOGIQUES : 79

NOTION D'EPIDEMIOLOGIE : 80

RESUME 83

NOTES PERSONNELLES 84

TABLES DE LOIS 85

REMERCIEMENTS 90

INTRODUCTION AUX BIOSTATISTIQUES

Les biostatistiques sont une matière dont le bénéfice est souvent sous-estimé ! Il ne faut pas les négliger, le coefficient reste important (que ce soit en PASS ou en LAS). Et c'est ce qui peut vous permettre de faire la différence ! Même pour ceux qui n'aiment pas les calculs, pas de panique, cette matière reste accessible, les exercices au concours sont souvent similaires, et si vous les maîtrisez bien, vous serez récompensés de votre dur labeur et prendrez même du plaisir à faire vos calculs ;)

L'épreuve se divise en 2 parties :

Une partie **QCMs** (environ 25 QCMs).

Une partie **QROC** (Question à Réponse Ouverte Courte) (environ 7 à 8). Vous n'aurez pas de case d'item à cocher mais une case vide à remplir avec la donnée chiffrée. /!\ NE PAS RÉPONDRE PAR DES PHRASES + DONNER LA RÉPONSE SOUS LE FORMAT ET LA PRÉCISION ATTENDUE (Ex : en %, à 0,01 près…)

Répartition des questions à l'examen -> approximative

1 - Rappels et probabilités (Pr Hitti).....(~9 QCM, QROC+++)
2 - Biostatistiques descriptives (Pr Cuggia).....(2 QCM)
3 - Variables aléatoires et lois (Pr Hitti).....(5 QCM)
4 - Estimations (Pr Hitti).....(1 QCM)
5 - Principes des tests statistiques (Pr Cuggia).....(2 QCM)
6 - Tests de comparaison de moyennes (Pr Cuggia).....(7 QCM)
7 - Tests de comparaison de fréquences (Pr Cuggia).....(3 QCM)
8 - Corrélations régressions (Pr Cuggia).....(3 QCM)
9 - Epidémiologie (Pr Cuggia).....(1 QCM)

Deux types d'enseignements :

Des **CM** (**Cours Magistraux**) (environ 27h)

- Toutes les diapositives qui contiennent ce qu'il faut connaître seront disponibles sur Moodle sur le site de l'ENT.

Des **ED** (**Enseignements Dirigés**) (environ 8h)

- En amphi ou en petits groupes.

En biostatistiques, vous aurez 2 professeurs.

Pr. Cuggia (14h de CM) : Cet enseignant prend vraiment le temps d'expliquer les choses et donne beaucoup d'exemples.

Ses QCMs sont plus axés sur des questions de cours. Ne négligez pas les derniers chapitres qui sont parfois vus plus rapidement à la fin du semestre *(cf. Epidémiologie en décembre 2022).*

Pr. Hitti (13h de CM) : Ses cours sont majoritairement composés de démonstrations, ne vous inquiétez pas si vous ne comprenez pas tout, les diapositives évoquées en cours sont plus compliquées qu'à l'examen...

Néanmoins, il peut être intéressant d'y assister pour avoir les explications et poser ses questions. Il réalise également la correction d'exercices lors des CM ce qui peut être utile pour comprendre les étapes de résolution des exercices et certaines astuces. Sa partie au concours concerne surtout des exercices de calculs. Cependant, il est important de bien maîtriser les bases, en particulier les formules de dérivées et primitives. Pensez à les revoir à la fin du semestre ! *(Cf. primitives, décembre 2021 et 2022).*

La calculatrice : Calculatrice scientifique TEXAS INSTRUMENTS TI-36X Pro

Si vous ne vous l'avez pas, il y a une commande groupée organisée par le tutorat tous les ans. Cependant, en attendant de l'avoir, ou s'il vous arrive de l'oublier, pas d'excuses pour ne pas travailler les biostats ! Il existe une application assez fidèle à la calculette : **Calculatrice scientifique 36.**

Et le tutorat dans tout ça ?

Bien sûr, vous n'êtes pas seuls, si vous avez la moindre question sur une conférence ou autre, il y a le **forum**. Et, si vraiment vous voulez une explication plus détaillée, toute **l'équipe de tuteurs** est à votre disposition pendant les conférences !

Les ED :

Ils sont proposés **par des profs** pour les chapitres les plus calculatoires (lois, tests, corrélation - régression). Ils sont réalisés en **petits groupes**. Il s'agit d'une heure de correction d'exercices donnés au préalable, cela vous sera utile pour poser toutes vos questions et reprendre avec le prof la résolution étape par étape de manière détaillée. Ils permettent donc un soutien supplémentaire pour réussir l'examen de biostatistiques !

Cela vous permet également de poser des questions sur les exercices du poly ED que vous avez eu du mal à résoudre.

METHODOLOGIE - BIOSTATISTIQUES

Globalement similaire pour toutes les matières “à calculs” (chimie, biostatistiques et biophysique) :

- Voir les **cours**
- Voir les **compléments de cours** (notamment ce livret) s’il y a des morceaux de cours incompris
- Voir la résolution d’**exercices types**
- Vous **entraîner** avec régularité !
    - Sur les conférences classées, les conférences qui sortent chaque semaine et les exams blancs
    - Essayez si vous le pouvez de bloquer un créneau horaire dédié à la matière

Les biostatistiques sont, comme la chimie ou la physique au deuxième semestre, une **matière à calculs.** Cela veut donc dire que vous ne serez pas évalués sur votre connaissance du cours mais sur votre capacité à résoudre des problèmes. Dans les faits, cela signifie que vous devrez non seulement apprendre vos cours, mais également savoir les appliquer. Pour cela, rien de tels que les **qcms et annales**, ainsi que la sacro-sainte **régularité**.

LES EXERCICES

Le seul moyen de savoir si on a bien compris un cours de biostatistiques, c’est d’appliquer soit-même les formules. C’est pourquoi nous vous recommandons de faire beaucoup de qcms. De plus, les cours de biostatistiques ne changent presque pas d’une année sur l’autre, ce qui veut dire que le contenu des annales correspond toujours au programme. N’hésitez donc pas à **toutes** les faire avant le concours.

Pro tip : lors de vos sessions d’annales, notez votre note par chapitre, comme ça vous pourrez voir les chapitres qui vous font perdre le plus de points ! Pas besoin de réviser les lois si vous avez eu tous les points sur les questions associées à ce chapitre.

LA RÉGULARITÉ

S’il est vrai que les savoirs non utilisés disparaissent plus rapidement de notre mémoire, ça l’est particulièrement pour les matières à calculs. C’est pourquoi nous vous recommandons vivement de vous entraîner régulièrement. Vous pouvez par exemple décider de **bloquer un créneau** réservé aux biostatistiques, et vous dire que vous travaillerez cette matière tous les jours de 13h à 14h, le matin dans l’amphi avant les cours, ou le soir pour bien finir la journée ! D’autres préfèrent une grosse séance le WE, à chacun sa méthode ! Comme ça, pas besoin de réfléchir, les révisions de cette UE deviennent un automatisme, et vous progressez plus rapidement et avec une meilleure rétention des choses apprises.

PLANIFIER

On peut avoir le vertige au début du semestre en voyant la liste des chapitres… Pour vous rassurer, vous pouvez **planifier** au début du semestre **l’ensemble des chapitres**, selon si vous voulez terminer en avance ou pile à temps le programme ! Il faut cependant faire attention à ne pas perdre trop de temps sur les premiers chapitres pour en avoir suffisamment à consacrer aux chapitres sur les tests qui peuvent être très chronophages ! Par ailleurs, pensez à revenir sur les premiers chapitres tout au long du semestre, par exemple, consacrer une journée d’entraînement par semaine aux chapitres plus anciens.

FICHE CALCULATRICE

Lois

Pour ce cours il n'y en a que trois et elles varient très peu : **Poisson**, **Binomiale** et **Normale**.

Sur la calculatrice, on trouve deux dénominations : 2nd, data, DISTR :

- **Pdf** correspond aux probabilités simples,
  - Pdf correspond à **P(X=x)**
- **Cdf** aux probabilités cumulées,
  - Cdf correspond à **P(X≤x)**
  - /!\ C'est toujours inférieur ou égal la valeur trouver avec la caculatrice /!\.

On n'utilisera pas Normalpdf cette année.

Loi NORMALE	Loi BINOMIALE	Loi de POISSON
2nd Data DISTR NORMALCDF	**2nd Data DISTR**	**2nd Data DISTR**
$X \sim N\ (34;\ 1{,}2)$ On entre : espérance = 34 et écart type = 1,2. On cherche $p(X \leq 30)$: • Borne inf : -1^{99} • Borne sup : 30 On cherche $p(30 \leq X \leq 38)$: • Borne inf : 30 • Borne sup : 38 On cherche $p(X > 36)$: • Borne inf : 36 • Borne sup : 1^{99}	BINOMIALPDF • On entre SINGLE, puis les paramètres n et p. • On calcule $p(X = x)$ BINOMIALCDF • On entre SINGLE, puis les paramètres n et p. • On calcule $p(X \leq x)$. Pour calculer $p(X \geq x)$: • On fait : $1 - \ p(X < x)$ • On fait $1 - p(X < (x-1))$ ○ Ex : si on veut $P\ (X \geq$ à $4)$ On cherche $1 - p(X < 3)$ avec la calculatrice Pour calculer $p(x1 \leq X \leq x2)$ • $p(X \leq x2) - p(X \leq x1)$	POISSONPDF • On entre SINGLE, puis le paramètre λ. • On calcule $p(X = x)$ POISSONCDF • On entre SINGLE, puis le paramètre λ. • **On calcule** $p(X \leq x)$ *Idem que pour la loi binomiale.*

Corrélation et Régression :

Variables	Signification
n	Nombre de points de données x (1 Var) ou (x,y) (2 Var).
$\bar{x}$ ou $\bar{y}$	Moyenne de toutes les valeurs de x ou y.
Sx ou Sy	Écart type de l'échantillon x ou y.
σx ou σy	Écart type de la population x ou y.
Σx ou Σy	Somme de toutes les valeurs de x ou y.
Σx^2 ou Σy^2	Somme de toutes les valeurs de x^2 ou y^2.
Σxy	Somme de (x…y) pour toutes les valeurs xy.
a (2 Variables)	Coefficient directeur de la droite de régression de Y en X.
b (2 Variables)	Régression linéaire (ordonnée de l'intersection avec l'axe des y).
r (2 Variables)	Coefficient de corrélation linéaire.
x' (2 Variables)	Utilise a et b pour calculer une valeur anticipée de x quand vous saisissez une valeur de y.
y' (2 Variables)	Utilise a et b pour calculer une valeur anticipée de y quand vous saisissez une valeur de x.
MinX	Minimum des valeurs de x.
Q1	Médiane des points entre le minimum et la médiane, ou premier quartile.
Méd	Médiane de tous les points.
Q3	Médiane des points entre le maximum et la médiane, ou troisième quartile.
MaxX	Maximum des valeurs de x.

/!\ Ne pas confondre Sx et σx /!\.

Calcul de r avec PEARSON

Poids en kg	47,8	49,4	50	50,7	55,8	57,7	67,7	70	72,4	73,8	82,2	88,7
Taux de cholestérol	44	62	130	126	62	98	181	237	280	254	318	293

1. Entrer les valeurs dans la calculatrice :

DATA

- L1 pour entrer les valeurs de la première ligne
- L2 pour entrer les valeurs de la deuxième ligne

PS : Pour supprimer les valeurs : appuyer 2 fois sur DATA puis CLEAR L1

2. Pour accéder aux fonctionnalités de calculs : 2nd puis DATA puis STAT-REG puis 2-VAR Stats.

- Choisir L1 pour xDATA;
- L2 pour yDATA;
- ONE pour FRQ,
- Puis CALC.

Descendre et trouver r.

Calcul de r*s* avec SPEARMAN :

1er tableau

[**Créatinine**] *(en mg pour 100 mL)*	**TFG** *(mL/min)*
0.85	90
0.99	45
1.13	70
1.13	103
1.47	60
1.56	35
1.56	53
2.77	27
3.11	25
3.96	15
4.69	15
5.93	5
5.93	10
7.07	10
11.02	12

2ème tableau

Rangs 1	**Rangs 2**
1	14
2	10
3.5	13
3.5	15
5	12
6.5	9
6.5	11
8	8
9	7
10	5.5
11	5.5
12.5	1
12.5	2.5
14	2.5
15	4

A partir du premier tableau, établir les rangs des deux colonnes (deuxième tableau).

- *PS : si on trouve plusieurs valeurs identiques, faire la moyenne de leurs rangs.*

3. Entrer les valeurs dans la calculatrice :

DATA

- L1 pour entrer les valeurs de la première ligne
- L2 pour entrer les valeurs de la deuxième ligne

PS : Pour supprimer les valeurs : appuyer 2 fois sur DATA puis CLEAR L1

4. Pour accéder aux fonctionnalités de calculs :

2nd puis DATA puis STAT-REG puis 2-VAR Stats.

- Choisir L1 pour xDATA;
- L2 pour yDATA;
- ONE pour FRQ,
- Puis CALC.

Descendre et trouver r_s.

Notes personnelles

RAPPELS

FONCTIONS, DERIVEES ET PRIMITIVES

Opérations sur dérivées

$(C \times f)' = C \times f'$	$\left(\frac{f}{g}\right)' = \frac{(f' \times g - f \times g')}{g^2}$	$\ln(f(x))' = \frac{f'(x)}{f(x)}$
$(f + g)' = f' + g'$	$(f \times g)'(x) = g'(x) \times f'(g(x))$	$\tan(f(x))' = \frac{f'(x)}{1 + tan^2(f(x))} = \frac{f'(x)}{cos^2(f(x))}$
$(f \times g)' = f' \times g + f \times g'$	$(sin\, f(x))' = f'(x) \times cos\, f(x)$	$\arccos(f(x))' = \frac{-f'(x)}{\sqrt{1 - f(x)^2}}$
$(\frac{1}{g})' = -\frac{g'}{g^2}$	$(\cos f(x))' = -f'(x) \times \sin\ f(x)$	$\arcsin(f(x))' = \frac{f'(x)}{\sqrt{1 - f(x)^2}}$
$(f^m)' = m \times f' \times f^{(m-1)}$	$\exp(f(x))' = f'(x) \times \exp(f(x))$	$\arctan(f(x))' = \frac{f'(x)}{(1 + f(x)^2)}$

C : Constante

f, g : variables

Dérivées des fonctions usuelles

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

Primitives des fonctions usuelles

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

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

CALCUL D'ERREUR

Soit X la valeur mesurée et x la valeur réelle.

On a deux types d'erreurs :

L'erreur absolue ΔX

L'erreur relative $\frac{DX}{X}$

Attention : L'erreur relative n'a pas d'unité.

Pour calculer ΔX, l'erreur absolue :

Soit $X = f(A, B, C \ldots)$

ΔA, ΔB, ΔC... les **erreurs directes**. Ces dernières sont données dans l'énoncé de l'exercice.

Et $\frac{df}{dA}, \frac{df}{dB}, \frac{df}{dC}$ ... les **dérivées partielles** de f par rapport aux variables A, B ou C.

Pour réaliser la ***dérivée partielle de f par A***, on dérive f tel que A correspond à la variable (x), B et C sont considérés comme des constantes donc fixes.
Pour réaliser la ***dérivée partielle de f par B***, on dérive f tel que B correspond à la variable (x), A et C sont considérés comme des constantes, et ainsi de suite avec chaque variable de f.

$$\Delta X = \left|\frac{df}{dA}\right| * \Delta A + \left|\frac{df}{dB}\right| * \Delta B + \left|\frac{df}{dC}\right| * \Delta C \ldots$$

Pour calculer l'erreur relative :

On divise l'erreur absolue par la valeure X.

Exercice type : densité de probabilité

Déterminez la valeur de C pour que la fonction f[x] définie sur l'intervalle [1;2] soit une densité de probabilité sur cet intervalle :

$$f(x) = \frac{x^3}{4} + \frac{4}{x^2} + Cx$$

A. C = -31/24
B. C= 9
C. C= 1
D. C= -6/ 3
E. C= 1/ 2
F. Toutes les propositions précédentes sont inexactes.

Correction :

Pour qu'une fonction soit densité de probabilité sur l'intervalle donné, la valeur de l'aire sous la courbe doit **TOUJOURS** être égale à 1.

Donc l'intégrale de f(x) doit être égale à 1 :

$$\Leftrightarrow \int_1^2 f(x) = 1$$

$$\Leftrightarrow \left[\frac{x^3}{4} + \frac{4}{x^2} + Cx\right]_1^2 = 1$$

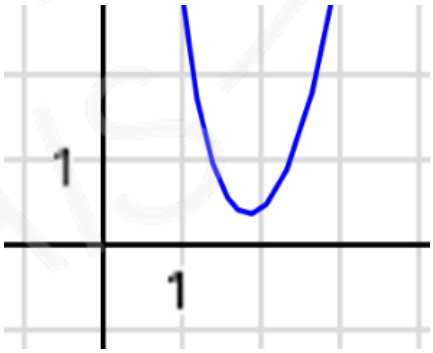
$$\Leftrightarrow \left[\frac{1}{16}x^4 - \frac{4}{x} + \frac{C}{2}x^2\right]_1^2 = 1$$

$$\Leftrightarrow \frac{2^4}{16} - 2 + \frac{4C}{2} - \left[\frac{1}{16} - 4 + \frac{C}{2}\right] = 1$$

$$\Leftrightarrow 1{,}5C = -\frac{31}{16} \Leftrightarrow C = -\frac{31}{24}$$

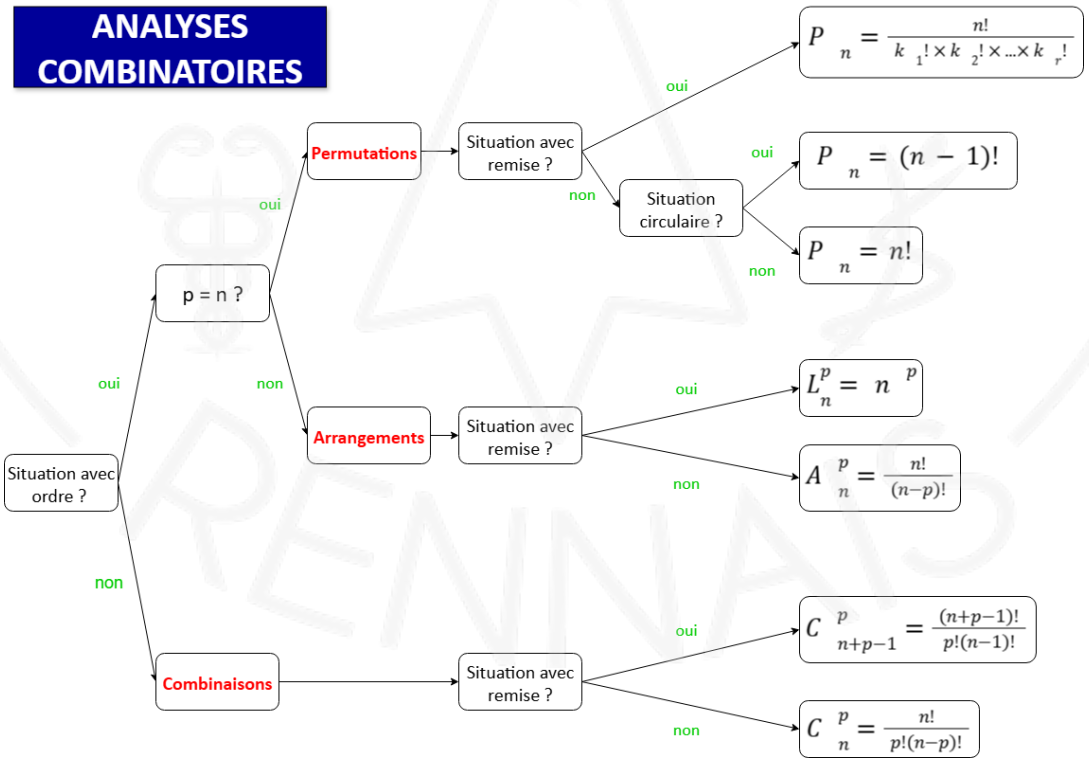
Remarque : On a ici utilisé les propriétés de l'intégrale pour résoudre cette densité de probabilité, celà implique de bien savoir manier les primitives.

En modélisant cette courbe avec C = -31/24, on obtient bien un résultat graphique approprié.



ANALYSES COMBINATOIRES

		SANS REMISE	AVEC REMISE
Avec ordre	**Arrangements** = nombre possible de suites ordonnées de **p objets parmi n** $1 \leq p \leq n$	$A_n^p = \frac{n!}{(n-p)!}$ *Taper "n" nPr "p" enter* *NB : pour nPr : retaper sur nCr*	$L_n^p = n^p$
	Permutations = nombre possible de suites ordonnées de n objets **(arrangement où p = n)**	$P_n = n!$ $NB : 0! = 1$	$P_n = \frac{n!}{k_1! * k_2! * \dots * k_r!}$
	/!\ **Permutations circulaires**	$P_n = \frac{n!}{n} = (n-1)!$	
Sans ordre	**Combinaisons** = nombre possible d'ensembles de **p objets parmi n** sans notion d'ordre	$C_n^p = \frac{n!}{p!(n-p)!}$ *taper "n" nCr "p" enter*	$C_{n+p-1}^p = \frac{(n+p-1)!}{p!(n-1)!}$



Symétrie :

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

Exercice type : Combinatoire

Méthode

1. Bien identifier la situation :

1. Sans **ordre** / avec **ordre**
2. Avec **tous les objets** / avec **une partie des objets**
3. Sans **remise** / avec **remise** (ou répétitions)

2. Choisir la bonne formule et l'appliquer

Exercice type Combinatoire : Annale 2018-2019

QCM 1 : Vous devez décorer une jardinière avec 7 fleurs que vous devez disposer en ligne. Vous avez 4 fleurs rouges et 3 fleurs blanches. Combien de jardinières différentes pouvez-vous proposer ?

A. 5040
B. 35
C. 70
D. 720
E. 15
F. Toutes les propositions précédentes sont inexactes.

QCM 2 : On vous impose que la première et la dernière fleur soient rouges. Combien de jardinières différentes avez-vous dans ces conditions ?

A. 2542
B. 35
C. 15
D. 180
E. 10
F. Toutes les propositions précédentes sont inexactes.

Correction type QCM 1 : B

1. Avec ordre / sans ordre

On remarque qu'il y a une notion d'ordre, en effet les jardinières seront différentes en fonction de l'ordre dans lequel les fleurs seront plantées.

2. Avec tous les objets / avec une partie des objets

Toutes les fleurs proposées sont à placer dans la jardinière, on prend donc tous les objets.

Il s'agit donc de chercher toutes les suites ordonnées de 7 fleurs qu'il est possible de faire avec 4 fleurs rouges et 3 fleurs blanches.

3. Sans remise / avec remise *(ou répétitions)*

On a 4 fleurs rouges et 3 fleurs blanches. On a donc des fleurs identiques qu'il faut prendre en compte puisque cela ne changera pas la suite. Il s'agit donc d'une situation avec des répétitions.

4. Se référer au tableau précédent.

Par lecture du tableau précédent, il s'agit donc d'une situation de permutation avec répétitions.

5. Calculer

On utilise donc la formule suivante : $P_n = \frac{n\,!}{k_1\,! * k_2\,! * \ldots * k_r\,!}$

- *Avec n = le nombre total d'objets = 7*
- *k1 = le nombre de répétition de l'objet fleur rouge = 4*
- *k2 = le nombre de répétition de l'objet fleur blanche = 3*

D'où : $P_7 = \frac{7\,!}{(4\,! * 3\,!)} = 35$

On peut proposer **35 jardinières différentes**.

Correction type QCM 2 : E

1. Situation

Il s'agit toujours de permutation avec répétition mais avec une contrainte supplémentaire. Avec deux fleurs à la place fixe, il y aura moins de possibilité.

2. Calculer

On utilise toujours la formule suivante : $P_n = \frac{n\,!}{k_1\,! * k_2\,! * \ldots * k_r\,!}$

Néanmoins, étant donné que l'on nous impose que la première et la dernière fleur soient rouges, cela nous enlève deux fleurs rouges pouvant permuter.

Nous avons donc maintenant :

- *n = le nombre total d'objets= 5*
- *k1 = le nombre de répétition de l'objet fleur rouge = 2 (4-2)*
- *k2 = le nombre de répétition de l'objet fleur blanche = 3*

D'où $P_5 = \frac{5\,!}{2\,! * 3\,!} = 10$

Dans ces conditions, nous avons **10 jardinières différentes**.

NOTES PERSONNELLES

PROBABILITES

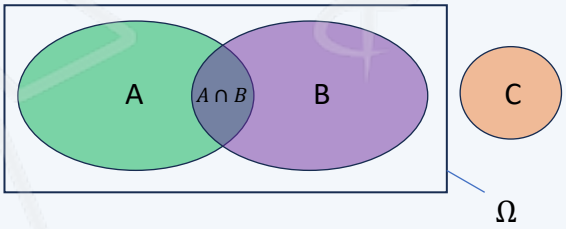
NOTION D'EVENEMENTS

Une probabilité correspond au « risque » ou « chance » de survenue d'un événement associé à une **fréquence** comprise **entre 0 et 1** : plus la fréquence s'approche de 1, plus le risque de survenue est élevé.

Les probabilités sont associées à des expériences aléatoires, dont on ne peut prédire le résultat à l'avance de façon certaine, dans lesquelles un événement est un sous-ensemble des résultats possibles de ces expériences.

Particularités :

Événements	Signes

• Événement **certain** : P = 1, noté Ω
  ○ *Ensemble des résultats possibles de l'expérience*
• Événement **impossible** : P = 0, noté ∅
  ○ *Ensemble vide, impossible à réaliser*
  ○ *Voir C sur le schéma A*
	• A ∩ B *("inter ")* : A ET B sont réalisés
	• A ∪ B *("union")* : A OU B sont réalisés
	• A ∩ C = ∅ : A et C sont incompatibles
	

CALCUL DES PROBABILITES

Probabilité d'un événement A : $$P(A) = \frac{n}{N} = \frac{\text{nb de cas favorables}}{\text{nb de cas possibles}}$$

Ex : dans une soirée poker, les tuteurs d'UE 4 utilisent un paquet de 52 cartes.

- *Quelle est la probabilité, en ne tirant qu'une seule carte au hasard, d'avoir un as ?*
- *n correspond au nombre de cas favorables, cad tous les AS du jeu donc 4 N correspond à toutes les issues possibles du tirage soit les 52 cartes du paquet*
- *P(As) = 4/52 = 1/13 = 0,0769*

PROPRIETES DES PROBABILITES

Si A et B incompatibles : **P(AUB) = P(A) + P(B)**

Si A et B quelconques : **P(AUB) = P(A) + P(B) - P(A∩B)**

$$0 \leq P \leq 1$$
$$P(\emptyset) = 0$$
$$P(\Omega) = 1$$
$$P(\overline{A}) = 1 - P(A)$$
$$si\ 1 \subset B\ alors\ P(A) \leq P(B)$$

PROBABILITES CONDITIONNELLES +++ ET THEOREME DE BAYES

Elles correspondent à la probabilité de survenue d'un événement sachant la survenue d'un autre événement.

$$\mathrm{P}\,(\mathrm{A}/\mathrm{B}) \;=\; \mathrm{P}_{\mathrm{B}}(\mathrm{A}) = \frac{\mathrm{P}\,(\mathrm{A} \cap \mathrm{B})}{\mathrm{P}\,(\mathrm{B})}\ avec\ P\,(B) \neq 0$$

Nb : Pour P (A/B), Il faut lire "probabilité de A sachant B"

/!\ **Incompatibles ≠ Indépendants**

Si A et B Incompatibles	Si A et B Indépendants
$\mathrm{P}\,(\mathrm{A} \cap \mathrm{B}) \;=\; 0 \;=\; \mathrm{P}_A(B)$	$\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}\,(\mathrm{A}) * P\,(B) \;\rightarrow \mathrm{P}_B(\mathrm{A}) = P\,(A)$

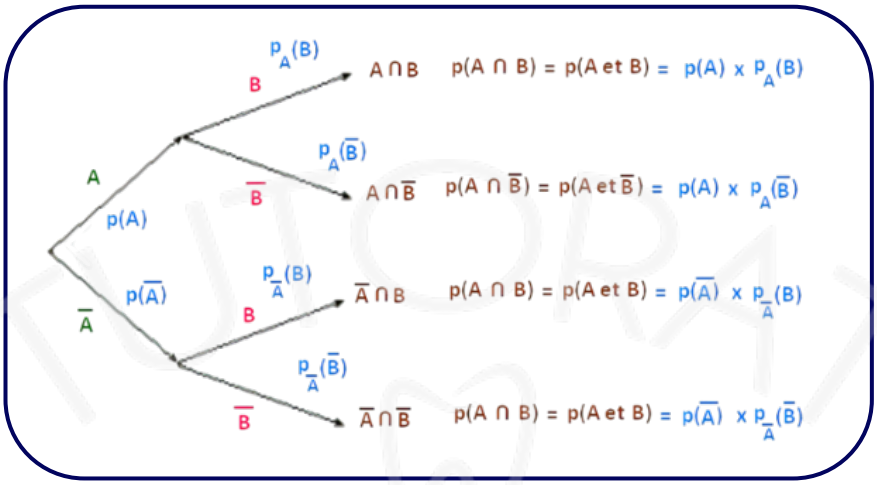
Des probabilités conditionnelles, on peut déduire : $P(A \cap B) = P_A(B) \times P(A) = P_B(A) \times P(B)$

D'où le Théorème de Bayes :

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

À ces probabilités composées sont associés des **arbres de décision**, base de ce chapitre. *Si vous avez des difficultés à résoudre un exercice de probas, revenez toujours à un arbre de décision !*

En voici un **exemple type** que vous pourrez appliquer à toutes les situations.

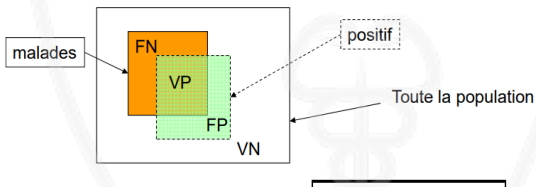


- P d'un événement = *multiplication des P des branches du chemin*
- Plusieurs chemins sont possibles pour 2 évènements = *somme des P de chaque chemin*

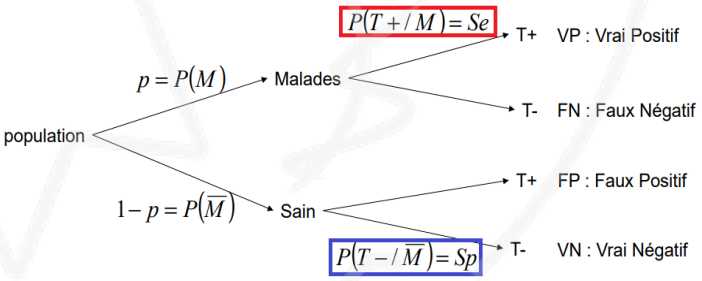
DIAGNOSTIQUE THERAPEUTIQUE +++

On dispose d'un **test gold standard** permettant d'identifier avec certitude une pathologie.

On élabore un **test alternatif** dont on doit tester l'efficacité. On prend une population test de sains et malades connus, et on remplit le tableau après avoir appliqué le test.



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



NB : pour trouver **VP**, **FN**, **FP** et **VN**, suivez le schéma de l'arbre de décision vu ci-dessus.

Dans une population constituée de sujets atteints d'une maladie M et de sujets sains, on a :

- p, la **prévalence** de la maladie : *la probabilité pour un sujet d'être malade.*
- Se, la **sensibilité** du test : *probabilité d'un sujet malade d'être positif* = $\frac{vrais\ positifs}{tous\ les\ malades}$
- Sp, la **spécificité** du test : *probabilité d'un sujet sain d'être négatif.* = $\frac{vrais\ négatifs}{tous\ les\ sains}$

/!\ Ne pas confondre la sensibilité et la spécificité avec VPP et VPN

- **VPP** : Valeur Prédictive Positive *: la probabilité pour un sujet positif d'être malade*
  - $P_{T+}(M) = VP / (VP + FP)$
- **VPN** : Valeur Prédictive Négative *: la probabilité pour un sujet négatif d'être sain*
  - $P_{T-}(S) = VN / (VN + FN)$

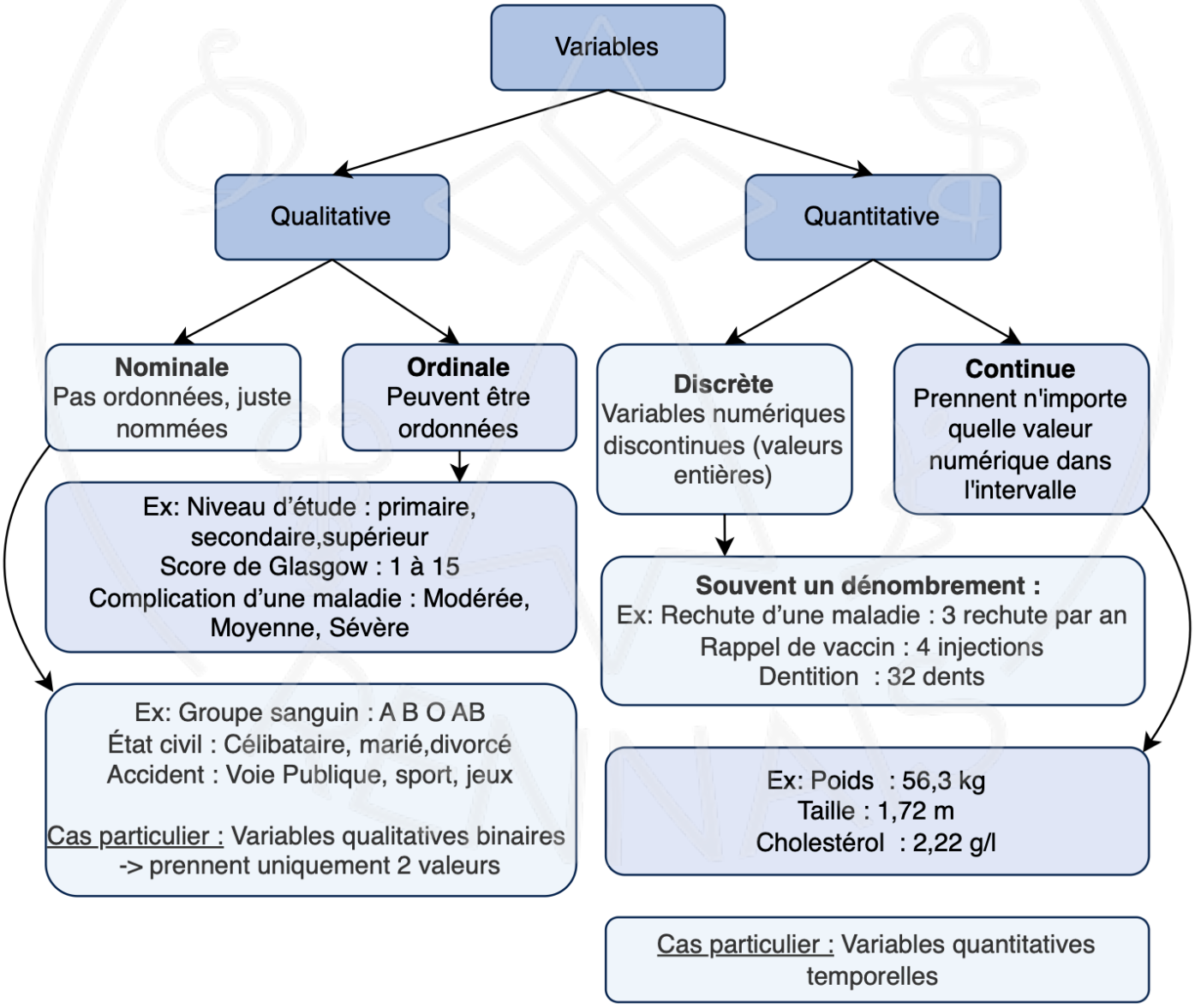


NOTES PERSONNELLES

BIOSTATISTIQUES DESCRIPTIVES

VARIABLES QUALITATIVES ET VARIABLES QUANTITATIVES :

Variables qualitatives	Variables quantitatives
Tableau de fréquence Diagramme en camembert Diagramme en bâton	Histogramme Polygone de fréquence Nuage de point Diagramme de Venn



MESURES EN STATISTIQUES

Les paramètres positions et dispersions

Paramètres de POSITION	Paramètres de DISPERSION
Moyenne Médiane Quartiles, déciles, percentiles Mode Fréquences relatives	Extrêmes (Minimum, Maximum) Etendue Intervalle interquartile Variance Ecart type Coefficient de variation

Paramètres	Formules
Moyenne	$m_x = \frac{1}{n}\sum_{1}^{n} x_i = \frac{\sum_{i=1}^{n} f_i X_i}{\sum_{i=1}^{n} f_i}$ $\qquad P = \frac{K}{N}$ Proportion pour variable qualitative binaire, de modalité K et de taille de la pop N.

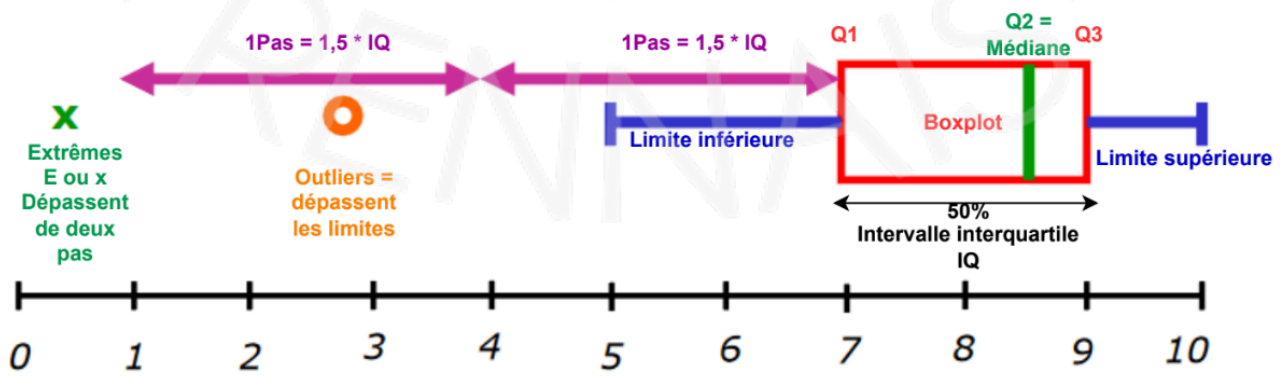
Médiane	• Si la série est **impaire**, la médiane = valeur observée chez le sujet médian
	  ○ *On ordonne dans l'ordre croissant les valeurs*
	• Si la série est **paire**, médiane = moyenne des valeurs qui séparent en 2 la série.
	• Avec un **tableau de fréquence :** Médiane = Borne inférieure de la classe médiane + longueur X
	  ○ *Voir comment trouver X dans le diapo du Pr.*
	La formule (nous paraissant plus simple pour calculer une médiane) ci-dessous n'est pas donné par le Pr. mais par vos tuteurs :
	Soit B = [Borne inférieure de l'intervalle comprenant les 50%]
	Soit I = [Intervalle précédent l'intervalle comprenant les 50%]
	Soit J = [Intervalle comprenant les 50%]
	• $Med = B + \frac{(0{,}5 - FR_{cumulée}[I]) \times Etendue[J]}{FR_{relative}[J]}$

Quartiles	3 valeurs qui partagent la distribution en 4 parties avec 25% chacunes • Q1 : sépare 25% des valeurs les plus faibles et 75% des valeurs les plus élevées • Q2 : Médiane, sépare 50 % des valeurs les plus faiblse de 50% des valeurs les plus élevées • Q3 : sépare 75% des valeurs les plus faibles et 25% des valeurs les plus élevées
Déciles	9 valeurs qui partagent la distribution en 10 groupes de tailles égales.

Pourcentiles	Sont les valeurs qui partagent la distribution en 100 groupes de tailles égales
Mode	Aucun, unique ou bimodale. Classe où l'effectif est le plus élevé.

Paramètres		**Formules**
Variance	***Population***	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$ $\sigma^2 = P(1-P)$ *Pour une variable qualitative binaire.*
	L'échantillon	$s^2 = \frac{\sum (x-m_x)^2}{n-1}$ $s^2 = p(1-p)$ *Pour une variable qualitative binaire.*
Ecart Type	***Population***	$\sigma = \sqrt{\sigma^2}$ $\sigma = \sqrt{P(1-P)}$ *Pour une variable qualitative binaire.*
	L'échantillon	$s = \sqrt{s^2}$ $s = \sqrt{p(1-p)}$ *Pour une variable qualitative binaire.*
Coefficient de variation (variable quantitative)		$CV = \frac{\sigma}{\mu}$
Etendue		Valeur Max - Valeur Min
Espace interquartile		Qi = Q3 – Q1 Contient 50% des valeurs de la série

Boite à moustache



NOTES PERSONNELLES

VARIABLES ALEATOIRES ET LOIS

OPERATIONS SUR ESPERANCE & VARIANCE

Pour des variables aléatoires discrètes :

Soit X et Y deux variables aléatoires (va) et a et b deux réels :

	Formules	Constante seule	Multiplication par une constante	Addition de deux va	Soustraction de deux va	Multiplication de deux va
Espérance = E(X)	E(X) $= \sum xiP(X = xi)$ $= \sum xipi$	E(*a*) = *a* *E(1) = 1* E(X + *a*) = E(X) + *a*	E(*a*X) = *a*E(X)	E(X+Y) = E(X)+E(Y)	E(X-Y) = E(X)-E(Y)	Si X et Y deux va indépendantes: E(X*Y) = E(X) * E(Y)
Variance = V(X)	V(X) $= E(X^2) - E(X)^2$	V(*a*) = 0 V(1) = 0 V(X+*b*) =V(X)	V(*a*X) = a^2V(x)	Si X et Y deux va indépendantes: V(X+Y) = V(X-Y) = V(X) + V(Y)	Si X et Y deux va indépendantes: V(X+Y) = V(X-Y) = V(X) + V(Y)	

Toutes les formules fonctionnent si les va sont dépendantes ou indépendantes. Exception faite pour celles dont il est précisé qu'elles ne sont utilisables qu'en cas d'indépendance entre les deux va, ainsi elles ne peuvent pas être utilisées si les va sont dépendantes.

La situation d'indépendance sera indiquée dans l'énoncé clairement si elle est présente.

Les exercices sur ces propriétés consistent souvent en la **détermination de l'espérance ou de la variance d'une nouvelle variable aléatoire créée à partir de deux anciennes** variables aléatoires, dont on connaît les paramètres d'espérance et de variance.

La résolue de ces exercices nécessite de la rigueur et une bonne connaissance des propriétés sur ces paramètres pour les appliquer un à un sans se tromper. Bon courage !

Exercices types : Opérations sur espérance et variance

Exercice type : Annale 2017-2018

QCM 9 : Soit deux variables aléatoires indépendantes X et Y respectivement de variance 4 et 8. On construit une nouvelle variable aléatoire Z avec Z = 3X - Y + 2. Quelle est la variance de la variable aléatoire Z ?

A. 20
B. 44
C. 24
D. 6
E. 26
F. Toutes les propositions précédentes sont inexactes.

Correction type QCM 9 : B

Bien comprendre la situation

On a deux variables aléatoires **indépendantes** X et Y (cf énoncé).

On a V(X) = 4 et V(Y) = 8.

On cherche la **variance** d'une **nouvelle variable aléatoire Z** définie comme Z = 3X - Y + 2.

1. Déterminer quelles propriétés seront utiles

Z = **3X-Y+2**

On nous demande ici de calculer la variance de Z. On se concentre donc sur les propriétés des variances.

On remarque dans la formule de la va Z, la présence d'une constante : **+2**. On aura donc besoin de la propriété pour la constante :

V(a) = 0 ou V(X+b) = V(X)

On remarque enfin une soustraction de deux va indépendantes (cf énoncé + formule de Z) : **X-Y**. On aura donc besoin de la propriété suivante :

Si X et Y deux va indépendantes : V(X+Y) = V(X-Y) = V(X) + V(Y)

On remarque la présence de la multiplication d'une va par une constante **3X**. On aura donc besoin de la propriété suivante :

$V(aX)=\alpha^2(X)$

2. Établir la formule (de la variance de Z ici)

Procéder étape par étape pour ne rien oublier selon les étapes suivantes.

Soit V(Z) = V(3X - Y +2)

Or **V(X+b) = V(X)**

Donc V(Z) = V(3X - Y)

Soit V(Z) = V(3X - Y)

Or **V(X-Y) = V(X) + V(Y)**

Donc V(Z) = V(3X) + V(Y)

Soit V(Z) = V(3X) + V(Y)

Or **V(aX)=α^2(X)**

Donc V(3X) = 32V(X) = 9V(X)

Donc **V(Z) = 9V(X) + V(Y)**

3. Calculer avec les valeurs le paramètre demandé

Ainsi V(Z) = 9V(X) + V(Y)

On a V(X) = 4 et V(Y) = 8.

Donc V(Z) = 94+8

Donc V(Z) = 44

Exercice type : Annale 2018-2019

QCM 11 : Soit deux variables aléatoires indépendantes X et Y respectivement de variance 10 et 5 et de moyenne 120 et 100. On construit une nouvelle variable aléatoire Z avec Z = 3X-4Y+50. Quelle est la moyenne de la variable aléatoire Z ?

A. 40
B. 10
C. 110
D. 405
E. 160
F. Toutes les propositions précédentes sont inexactes.

Correction détaillée QCM 11 : B

Bien comprendre la situation

On a deux variables aléatoires **indépendantes** X et Y (cf énoncé).

On a :

Variables aléatoires	X	Y
Variance	10	5
Moyenne	120	100

On cherche la moyenne d'une **nouvelle variable aléatoire Z** définie comme Z = 3X - 4Y + 50.

Rappel : Espérance = moyenne.

4. Déterminer quelles propriétés seront utiles

Z = **3X – 4Y + 50**

On nous demande ici de calculer la moyenne de Z. On se concentre donc sur les propriétés des moyennes.

On remarque dans la formule de la va Z, la présence d'une constante : **+50**. On aura donc besoin de la propriété pour la constante :

E(a) = a avec a est un réel.

On remarque enfin une soustraction de deux va : **X-Y**. On aura donc besoin de la propriété suivante :

E(X-Y) = E(X) + E(Y)

On remarque la présence de multiplications d'une va par une constante : **3X / 4Y**. On aura donc besoin de la propriété suivante :

E(*a*X) = *a*E(X)

5. Établir la formule (de la moyenne de Z ici)

Procéder étape par étape pour ne rien oublier selon les étapes suivantes.

Soit E(Z) = E(3X - 4Y +50)

Or **E(a) = a**

Donc E(Z) = E(3X - 4Y) + 50

Soit E(Z) = E(3X - 4Y) + 50

Or **E(X-Y) = E(X) + E(Y)**

Donc E(Z) = E(3X) - E(4Y) +50

Soit E(Z) = E(3X) - E(4Y) +50

Or **E(*a*X) = *a*E(X)**

Donc E(3X) = 3E(X)

Et E(4Y) = 4E(Y)

Donc **E(Z) = 3E(X) - 4E(Y) +50**

6. Calculer avec les valeurs le paramètre demandé

Ainsi E(Z) = 3E(X) - 4E(Y) +50

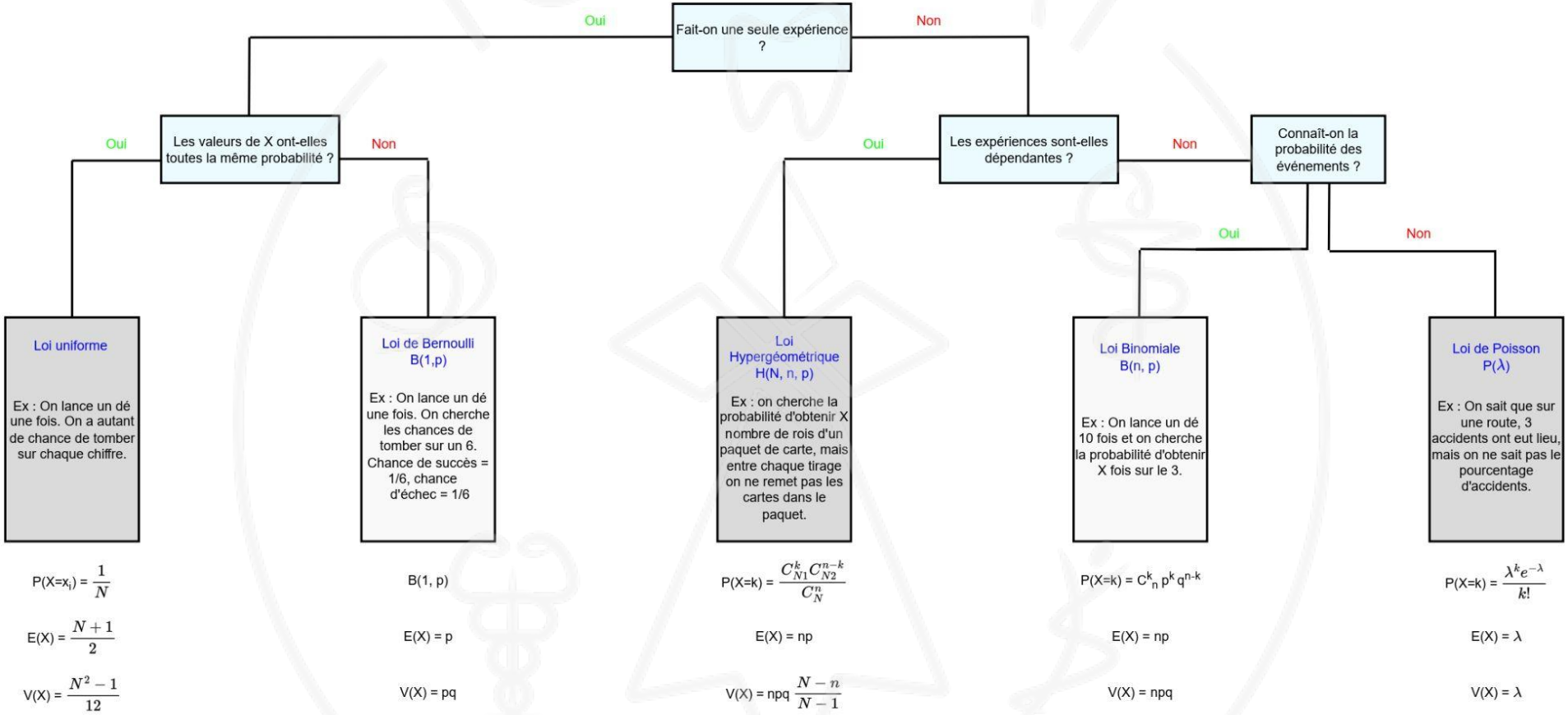
On a E(X) = 120 et E(Y) = 100.

Donc E(Z) = 3120 - 4100 +50

Donc E(Z) = 10

Réponse B

VARIABLES ALEATOIRES ET LOIS

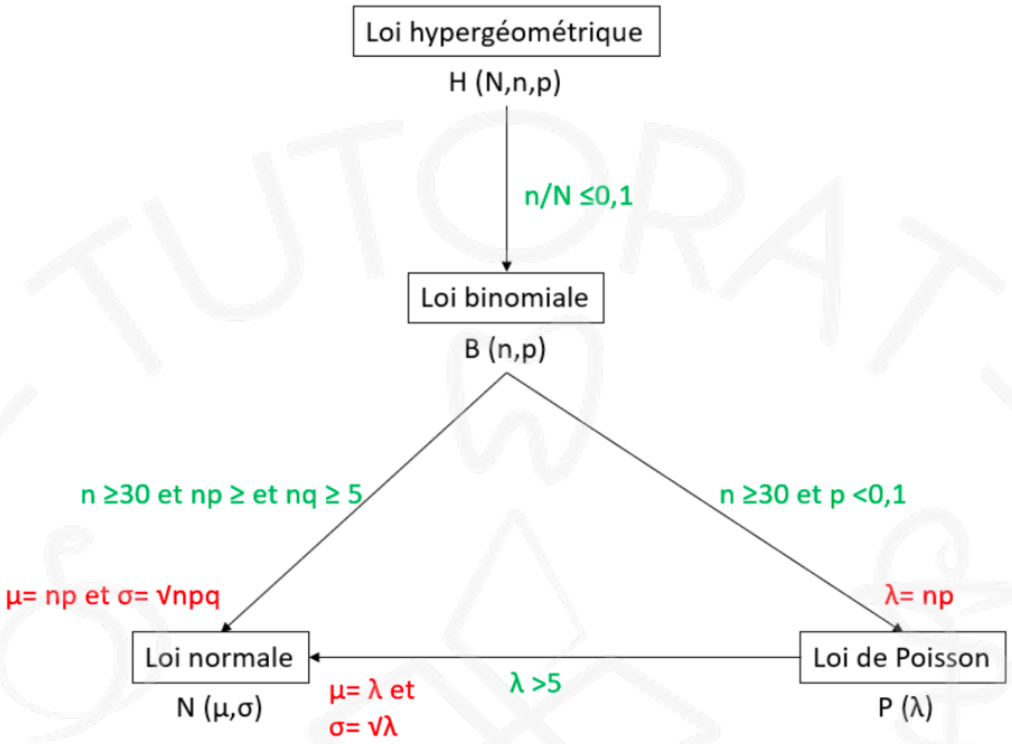


n: nombre de tentatives, taille de l'échantillon
k: nombre de succès
q = 1 - p
N: population
N1, N2: sous population

Exemple pour n, N, N1 et N2 : Soit 12 billes de couleurs (N), on en tire 10 (n), sachant qu'il y en a 3 bleues (N1) et 9 rouges (N2).

Attention : les lois binomiale et de Poisson sont stables.
Pour la loi Binomiale, $X_n + X_m = B(n+m, p)$
Pour la loi de Poisson, $X1 + X2 = P(\lambda 1 + \lambda 2)$

Approximations des lois



NOTES PERSONNELLES

LOIS CONTINUES

CARACTERISTIQUES DES VARIABLES ALEATOIRES CONTINUES :

Une **v.a. continue** est définie par trois caractéristiques :

- **Fonction de répartition** : $F(x_i) = P(X \leq x_i) = \int_a^x f(x)dx$
- **Espérance mathématique** : $E(X) = \int_a^b xf(x)dx$
- **Variance** : $V(X) = \int_a^b x^2 f(x)dx - E(X)^2$

	Loi uniforme	Loi normale	Théorème centrale limite
Densité de probabilité	$f(x) = \frac{1}{b-a}$ si $x \in [a;b]$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ *Pas à apprendre*	
Espérance	$E(X) = \frac{b+a}{2}$	$E(X) = \mu$	$E(X) = 0$
Variance	$V(X) = \frac{(b-a)^2}{12}$	$V(X) = \sigma^2$	$V(X) = 1$

STABILITE DE LA LOI NORMALE :

- ***X+Y*** suit la loi normale **N** $(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.
- **X-Y** suit la loi normale **N** $(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.
- **aX + b** suit la loi normale **N** $(a\mu + b, |a|\sigma)$.

COMMENT UTILISER LE THEOREME CENTRALE LIMITE :

Soit X une variable suivant la loi normale, Z suit une **loi normale centrée réduite N (0, 1)** avec $Z = \frac{X-\mu}{\sigma}$.

La fonction $\Pi(u)$ donne la probabilité d'avoir une valeur inférieure à u.

Prenons un exemple : la concentration sanguine en hémoglobine X, exprimée en g/100mL, est distribuée normalement avec une moyenne $\mu = 30$ et un écart-type $\sigma = 5$.

A. Quelle est la probabilité que X > 34 ?
B. Quelle est la probabilité que 22 < X < 35 ?
C. Quel est l'intervalle centré sur la moyenne contenant 80% de la population ?

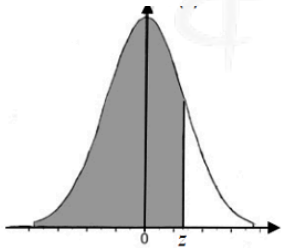
Quelle est la probabilité que X > 34 ?

Commencer par calculer Z : $Z = \frac{34-30}{5} = 0{,}8$

- Aller chercher dans la table de la fonction de répartition de la loi normale la valeur de Π.

Loi normale : fonction de répartition

Pour une valeur $z \geq 0$, la table ci-dessous renvoie la valeur de la fonction de répartition $F(z)$ d'une loi normale centrée réduite au point z.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177

- Pour $\boldsymbol{Z = 0.8}$, $\boldsymbol{\Pi = 0{,}7881}$. $\boldsymbol{P(X < 34) = 0{,}7881}$.
- A partir de là, on peut déduire $\boldsymbol{P(X > 34) = 1 - P(X < 34) = 1 - 0{,}7881 = 0{,}2119}$.
- La probabilité que $X > 34$ est de 21,19 %

Quelle est la probabilité que 22 < X < 35 ?

Commencer par calculer les Z :

$$Z_1 = \frac{35-30}{5} = 1 \text{ et } Z_2 = \frac{22-30}{5} = -1{,}6$$

Aller chercher dans la table de la fonction de répartition de la loi normale la valeur de $\boldsymbol{\Pi}$ *(voir ci-dessus).*

- $\boldsymbol{\Pi(1) = 0{,}8413}$ et $\boldsymbol{\Pi(-1{,}6)} = 1 - \Pi(1{,}6) = 1 - 0{,}9452 = 0{,}0548$.

Pour obtenir la probabilité de l'intervalle, on soustrait à la probabilité associée à la borne supérieure la probabilité associée à la borne inférieure, soit :

- $\boldsymbol{\Pi(22 < X < 35)} = 0{,}8413 - 0{,}0548 = 0{,}7865$

La probabilité que X soit compris entre 22 et 35 est de 78,65%.

Quel est l'intervalle centré sur la moyenne contenant 80% de la population ?

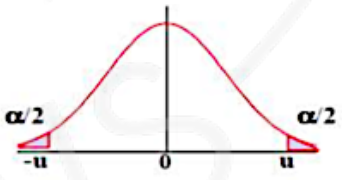
Pour cette question on doit appliquer le même raisonnement que pour les deux questions précédentes, mais à l'envers (plutôt que de partir de X pour trouver $\boldsymbol{\Pi}$, on part de $\boldsymbol{\Pi}$=0,80 pour retrouver X).

- On cherche dans la **table de l'écart réduit** (!) la valeur de Z pour laquelle α = 0,2 (les 20% non compris dans l'intervalle), soit **Z = 1,281552**.
- On a $\boldsymbol{Z_+}$ = 1,281552 et $\boldsymbol{Z_-} = -1{,}281552$.
- $\boldsymbol{Z_+} = 1{,}281552 = \frac{X_+ - \mu}{\sigma} = \frac{X_+ - 30}{5}$ soit $\boldsymbol{X_+} = Z_+ \times 5 + 30 = 36{,}41$.
- $\boldsymbol{X_-} = -1{,}281552 \times 5 + 30 = 23{,}60$.

80% des valeurs de X se trouvent dans l'intervalle [23,60 ; 36,41].

Table de l'écart réduit (loi Normale centrée réduite)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée u, c'est à dire la probabilité extérieure à l'intervalle (-u, +u).



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,575 829	2,326 348	2,170 090	2,053 749	1,959 964	1,880 794	1,811 911	1,750 686	1,695 398
0,1	1,644 854	1,598 193	1,554 774	1,514 102	1,475 791	1,439 531	1,405 072	1,372 204	1,340 755	1,310 579
0,2	1,281 552	1,253 565	1,226 528	1,200 359	1,174 987	1,150 349	1,126 391	1,103 063	1,080 319	1,058 122
0,3	1,036 433	1,015 222	0,994 458	0,974 114	0,954 165	0,934 589	0,915 365	0,896 473	0,877 896	0,859 617
0,4	0,841 621	0,823 894	0,806 421	0,789 192	0,772 193	0,755 415	0,738 847	0,722 479	0,706 303	0,690 309
0,5	0,674 490	0,658 838	0,643 345	0,628 006	0,612 813	0,597 760	0,582 842	0,568 051	0,553 385	0,538 836
0,6	0,524 401	0,510 073	0,495 850	0,481 727	0,467 699	0,453 762	0,439 913	0,426 148	0,412 463	0,398 855
0,7	0,385 320	0,371 856	0,358 459	0,345 126	0,331 853	0,318 639	0,305 481	0,292 375	0,279 319	0,266 311
0,8	0,253 347	0,240 426	0,227 545	0,214 702	0,201 893	0,189 118	0,176 374	0,163 658	0,150 969	0,138 304
0,9	0,125 661	0,113 039	0,100 434	0,087 845	0,075 270	0,062 707	0,050 154	0,037 608	0,025 069	0,012 533

α	0,002	0,001	0,0 001	0,00 001	0,000 001	0,0 000 001	0,00 000 001	0,000 000 001
u	3,090 23	3,290 53	3,890 59	4,417 17	4,891 64	5,326 72	5,730 73	6,109 41

ESTIMATIONS

Les **recensements** sont des études **exhaustives** d'un caractère donné dans une population *(ex : recensement de population, recensement à 18 ans, etc…)*.

Ils permettent une information extrêmement précise mais sont très coûteux, difficiles et demandent beaucoup de temps : de fait, ils sont très peu réalisés.

Les études sont donc généralement menées sur un **sous-ensemble ou échantillon** de la population et on obtient des **estimations** des caractères étudiés : plus l'échantillon est grand, plus l'estimation sera précise.

Population	Ensemble des individus ou unités statistiques
Echantillon	Sous-ensemble de cette population obtenu par des méthodes d'échantillonnage
Sondage	Méthode utilisée pour échantillonner

Mais pour que cette méthode soit viable, il faut que l'échantillon soit **représentatif** de la population : s'il ne l'est pas, il est dit **biaisé** *(c'est un **biais de sélection**, source d'erreurs potentielles dans les études).*

Il existe deux grands types de sondages : les sondages **raisonnés** et les sondages **aléatoires**.

Les différents types de sondage

Sondages raisonnés
critères fixés à priori

Méthode des **unités types**
division de la population en sous-ensembles, chacun d'entre eux représenté par une unité type considérée comme représentative

Ex : On désigne un champ d'une région comme représentant de tous les champs

Méthode des **quotas**
l'échantillon doit respecter des contraintes, des proportions
Ex : si on a 35% de cadres dans la population, on doit avoir 35% de cadres dans l'échantillon

Sondages aléatoires
Sélection des unités statistiques au hasard parmi une base de sondage

Sondage **élémentaire**
Tirage au sort de n individus parmi les N unités statistiques de la population
Taux de sondage : $\frac{n}{N}$

Sondage **stratifié**
la population est divisée en **q** sous-ensembles et l'échantillon est constitué de **q** sous-échantillons choisis au hasard, un dans chaque strate

Sondage **systématique**
Tirage au hasard d'un individu **tous les k individus**, k étant le pas du sondage
$k = \frac{N}{n}$
Ex; si on a une pop de 300 unités et qu'on veut un échantillon de 60 unité, on choisit un pas $k = \frac{300}{60} = 5$

Sondage **à différents degrés**
la population est divisée en unités primaires, elles-mêmes divisées en unités secondaires, etc... puis on réalise un tirage au hasard en cascade
Ex : on tire d'abord le département, puis le canton, puis la commune, et enfin le ménage

cas particulier : tirage **par grappes.**

Notation :
Dans une population : taille N/ moyenne μ / écart type σ / proportion p
Dans un échantillon : taille n / moyenne m / écart type s / proportion p_0

A partir de ces **sondages**, on peut créer des **estimateurs :**

	Population	**Echantillon**
Moyenne	$E(x) = \mu = \frac{\Sigma x_i}{N}$	$m = \frac{\Sigma x_i}{n}$
Variance	$Var(x) = \sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}$	$s^2 = \frac{1}{n-1}\Sigma(x_i - m)^2$ $s^2 = \frac{1}{n-1}(\Sigma {x_i}^2 - nm^2)$ Formule développée +++ : $s^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1}$
Écart type	$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$	$s = \sqrt{s^2}$
Proportion	$p = \frac{\Sigma X_i}{N}$	$p = \frac{\Sigma X_i}{n}$

EXERCICE CORRIGE :

Les périmètres crâniens xi mesurés sur deux groupes d'enfants ont donné les résultats suivants :

Groupe A : nA=87	xi=2673	xi^2=82457
Groupe B : nB=82	xi=2465	xi^2=74437

Les moyennes et variances des deux groupes A et B valent respectivement environ :

A. mA=31,08 et mB=30,43
B. mA=30,72 et mB=30,06
C. sA^2=3,85 et sB^2=4,16
D. sA^2=3,49 et sB^2=3,56
E. Toutes les propositions précédentes sont inexactes.

Réponse : BC

A. FAUX.
B. VRAI : Pour calculer la moyenne on utilise la formule m=xin. On a donc mA=26738730,72 et mB=24658230,06
C. VRAI : Pour calculer la variance, on utilise la formule s^2=xi^2-xi2nn-1. On a donc sA^2=82457-(2673)²87863,85 et sB^2=74437-(2465)²82814,16
D. FAUX.
E. FAUX.

On peut également calculer des **intervalles de confiance** ou **de pari :**

	Intervalle de confiance	**Intervalle de pari**
Définition	Le **paramètre inconnu** est celui de la **population**, on l'estime alors à partir de l'échantillon. Il a une probabilité (1-α) de se trouver dans cet intervalle. (On l'appelle aussi intervalle de confiance au risque α, en général 5%).	Le **paramètre inconnu** est celui de **l'échantillon** : on connaît ceux de la population et on élabore un intervalle qui pourrait contenir l'estimation, avec une certaine probabilité et un certain risque.
Grands échantillons $(n \geq 30)$ Pour la **moyenne**	$\left[m - z_{\alpha/2}\frac{s}{\sqrt{n}}, m + z_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ *La valeur $z_{\alpha/2}$ se trouve dans la table de la variable normale réduite.*	$\left[\mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ *La valeur $z_{\alpha/2}$ se trouve dans la table de la variable normale réduite.*
Grands échantillons $(n \geq 30)$ Pour la **proportion**	$\left[p_0 - z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}\right]$ *La valeur $z_{\alpha/2}$ se trouve dans la table de la variable normale réduite.*	$\left[p - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]$ *La valeur $z_{\alpha/2}$ se trouve dans la table de la variable normale réduite.*
Petits échantillons $(n \leq 30)$	$\left[m - t_{\alpha/2}\frac{s}{\sqrt{n}}, m + t_{\alpha/2}\frac{s}{\sqrt{n}}\right]$ *La valeur $t_{\alpha/2}$ est donnée par la **table de la variable t de Student** pour n-1 ddl ou degré de liberté.*	$\left[\mu - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \mu + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ *La valeur $t_{\alpha/2}$ est donnée par la **table de la variable t de Student** pour n-1 ddl ou degré de liberté.*

Tableau récap de fin du cours :

	Paramètre de la population	Estimateur	Variance de l'estimateur	Estimation *(à partir d'un échantillon de n individus)*	Intervalle de confiance *(sur le paramètre de la population)*	Intervalle de pari *(sur l'estimation à partir d'un échantillon)*
Variable quantitative	Moyenne μ	$\hat{\mu}$	$\frac{\sigma^2}{n}$	m	Grand échantillon $n \geq 30$	Grand échantillon $n \geq 30$
					Petit échantillon et variable distribuée selon une loi normale	
	Variance σ^2	$\hat{s}^2$		s^2		
	Écart-type σ	$\hat{s}$		s		
Variable qualitative à deux modalités	Proportion p	$\hat{p}$	$\frac{p(1-p)}{n}$	p_0	Grand échantillon $n \geq 30$ $np \geq 5$ $n(1-p) \geq 5$	Grand échantillon $n \geq 30$ $np \geq 5$ $n(1-p) \geq 5$

EXERCICE CORRIGE :

En France, la prévalence de l'aspartite chez le babouin est de 6 %. Quelle taille minimum doit faire un échantillon de babouins pour que, avec une probabilité de 95 %, la proportion de babouins aspartiques dans l'échantillon soit de 6 % ± 2% ? (à 2 unités près)

A. 462
B. 256
C. 156
D. 542
E. 57
F. Toutes les propositions précédentes sont inexactes.

On applique la formule $[p - z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}]$.

Ici, on cherche à construire un intervalle dans lequel 6% des babouins ont 95% de chance d'être situé dans l'intervalle [0.04 ; 0.08]. Il s'agit donc d'un intervalle de pari. On a donc un risque $\alpha = 5\%$ et on a $p = 0.06$ et on veut connaître n. On veut donc que :

$[p - z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}, p + z_{\alpha/2} * \sqrt{\frac{p(1-p)}{n}}]$ = **[0.04 ; 0.08]**

On peut essayer avec la calculatrice toutes les propositions ou on peut résoudre l'équation.
$z_{\alpha/2} = 1.96$ (*Cf. table loi normale centrée réduite*) donc on a :

- $0.06 - 1.96 * \sqrt{\frac{0.06*0.94}{n}} = 0.04 \Leftrightarrow (\frac{0.06-0.04}{1.96})^2 = \frac{0.06*0.94}{n} \Leftrightarrow n = \frac{0.06*0.94}{0.000104123} = 541.6656 \approx 542$

On peut aussi faire avec 0.08 et on constate que cela revient au même :

- $0.06 + 1.96 * \sqrt{\frac{0.06*0.94}{n}} = 0.08 \Leftrightarrow (\frac{0.08-0.06}{1.96})^2 = \frac{0.06*0.94}{n} \Leftrightarrow n = \frac{0.06*0.94}{0.000104123} = 541.6656 \approx 542$

Nb : On peut constater que $n \approx 542 \geq 30$ *et que* $np \approx 32 \geq 5$ *et* $n(1-p) \approx 22 \geq 5$ *donc il s'agit bien d'un grand échantillon dont on peut approximer la distribution comme suivant une loi normale. Si les conditions n'avaient pas été respectées, il aurait fallu faire avec la formule.*

$[\mu - t_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \mu + t_{\alpha/2} * \frac{\sigma}{\sqrt{n}}]$ et regarder la valeur $t_{\alpha/2}$ dans la table de Student pour n-1 ddl.

Réponse : D

NOTES PERSONNELLES

TEST STATISTIQUE

QU'EST-CE QU'UN TEST STATISTIQUE

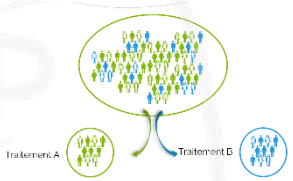
On s'intéresse à **la relation entre deux variables**, c'est-à-dire qu'on se demande s'il y a ou non une **association** entre les deux. Réaliser un test statistique permet de vérifier si deux variables sont statistiquement liées ou non : pour cela on calcule « la probabilité qu'elles soient associées en comparant un outil de mesure de cette variable (la moyenne par exemple) sous le risque qu'on a de se tromper : le risque alpha ».

On utilise soit des mesures de la **population** (recensement), soit des **échantillons** de la population pour faciliter le test et que l'on considère comme représentatif de la population. L'utilisation d'échantillons peut engendrer des biais de sélection, c'est-à-dire des erreurs dans le résultat dû à un échantillon non représentatif alors qu'on le pensait l'être.

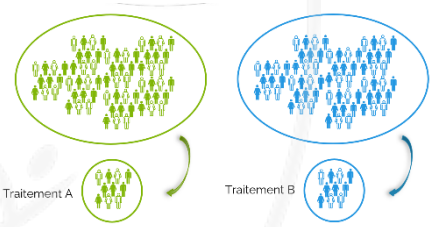
On peut comparer grâce au test un paramètre **théorique** (issu de la population) à un paramètre **observé** (issu d'un échantillon) (par exemple pour la comparaison de l'espérance de vie sans tt avec celle avec tt) ou bien deux paramètres observés entre eux (par exemple pour la comparaison de deux traitement entre eux).

Ainsi, on verra les tests statistiques pour les moyennes (quantitatif) et pour les pourcentages (qualitatif).

Pour réaliser un test, on commence toujours par poser des **hypothèses** :



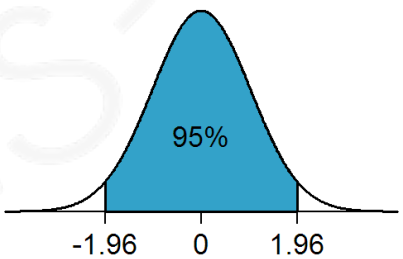
- On part toujours du postulat que **l'hypothèse nulle** est vraie à priori : H_0 . Elle correspond à une égalité entre les deux variables $\mu_A = \mu_B$. Par exemple, cela signifie que la moyenne de guérison avec le traitement A est la même qu'avec le traitement B. Il n'y a pas d'association entre les deux variables.



- On cherche à savoir si on peut accepter **l'hypothèse alternative : H_1 .** Cette hypothèse correspond à une différence entre les deux variables $\mu_1 \neq \mu_2$. Elle correspond à ce que l'on cherche à montrer. Elle est opposée ou contraire à l'hypothèse nulle. Elle affirme que le paramètre utilisé pour l'hypothèse nulle est soit supérieur, soit inférieur, soit différent. Par exemple : elle permet de conclure que la moyenne de guérison entre les deux traitements est différente (sous H_1 bilatérale) ou que le traitement B donne une moyenne de guérison inférieure à celle du traitement A (sous H_1 unilatérale).

Ainsi les tests permettent de conclure de la réussite ou non d'un nouveau traitement en médecine.

Pour réaliser le test, il faut également définir **le risque alpha** (donné dans les énoncés pour les QCMs) :



Lors d'un test statistique, il faut bien comprendre qu'on n'est jamais sûr à 100% du résultat. **Il y a toujours un risque de se tromper.** Il y a un risque d'observer qqch sur l'échantillon qui n'existe pas en population.

Il est défini à priori (c'est-à-dire avant de faire le test).

Voir définition en dessous.

Par défaut, on fixe le paramètre alpha à 5% : c'est-à-dire que la probabilité maximale de rejeter HO si elle est vraie, donc de se tromper, est de 5%.

Il s'agit du seuil que l'on fixe pour savoir s'il y a une différence ou non. Cette courbe de distribution de la loi normale, indique que les valeurs t<-1.96 et t>1.96 ont moins de 5% de chances d'être observée si H0 est vraie. Donc si on obtient une de ces valeurs c'est qu'il est peu probable que H0 soit vraie.

En bleu, on a la zone où H0 a de grandes chances d'être vraie selon les valeurs statistiques obenues (et mise en abscisse) et en blanc les zones où H0 est peu probable d'être vraie selon ces valeurs. Donc par exemple si on obtient 3,84, on se situe dans la zone blanche, on peut donc rejeter H0, en revanche si on obtient 0,7 on se trouve dans la zone bleue et on ne peut pas rejeter H0.

Pour réaliser un test, il faut également définir s'il est **bilatéral ou unilatéral** (idem précisé dans l'énoncé),

- Un test **bilatéral** est un test où on conclura que le paramètre de l'hypothèse alternative est différent sans pour autant pouvoir en donner le sens $\mu_1 \neq \mu_2$.
- Un test **unilatéral** est un test où on conclura que le paramètre de l'hypothèse alternative est plus petit ou plus grand que l'autre valeur $\mu_1 < \mu_2$ ou $\mu_1 > \mu_2$.

Ensuite, il faut **choisir le test** que l'on va utiliser. Pour cela, il faut se référer aux arbres décisionnels (suite du livret) en prenant en compte les différents paramètres de notre échantillon : observé ou théorique, sa taille, sa normalité, l'indépendance entre les groupes, les types de variables (qualitative/quantitative)…

Les deux grands types sont paramétriques ou non :

- Les tests **paramétriques** sont des tests dont l'échantillon que nous étudions suit une certaine loi (loi normale par exemple) ou vérifie un certain nombre d'hypothèses (même variance entre les deux échantillons donnés). Ils sont **plus puissants** mais nécessitent un certain nombre d'hypothèses à vérifier.
- Si l'échantillon que nous avons à disposition ne vérifie pas les hypothèses données, alors on applique un test **non paramétrique**. Ils ont tout de même certaines conditions à vérifier comme la taille de l'échantillon par exemple.

Choix d'un test

Type de variable		Echantillons indépendants		Echantillons dépendants	
		Paramétrique	Non Paramétrique	Paramétrique	Non paramétrique
qual vs qual			test du χ^2		Test de McNemar
			test du corrigé de yates χ^2 et test exact de Fisher et		
quant vs qual	2 échantillons	Test Z Test de student non apparié	Test U de Mann-Whitney	Test de Student apparié	Test de Wilcoxon
	> 2 échantillons	ANOVA	Test de Kruskall-Wallis		
quant vs quant				r de Pearson	r de Spearman

Paramétrique	Non Paramétrique
- Test Z de l'écart-réduit - Test de Student - ANOVA - Test de Pearson	- Test de Chi2 - Test exact de Fisher - Test du corrigé de Yates Chi2 - Test de Mann et Whitney - Test de Kruskall-Wallis - Test de McNemar - Test de Wilcoxon - Test de Spearman

Quant vs Quant Test de **Corrélation**	Quant vs Qual Test de comparaison de **moyenne**	Qual vs Qual Test de comparaison de **pourcentages** *Tous sont non paramétriques*
- Test de Spearman - Test de Pearson	- Test Z de l'écart-réduit - Test de Student - Test de Mann et Whitney - Test de Wilcoxon	- Test du Chi2 - Test du corrigé de Yates - Test de Fisher - Test de Mc Nemar

Vert = échantillons indépendants // **Rouge** = échantillons dépendants // **Test de Student** = les deux

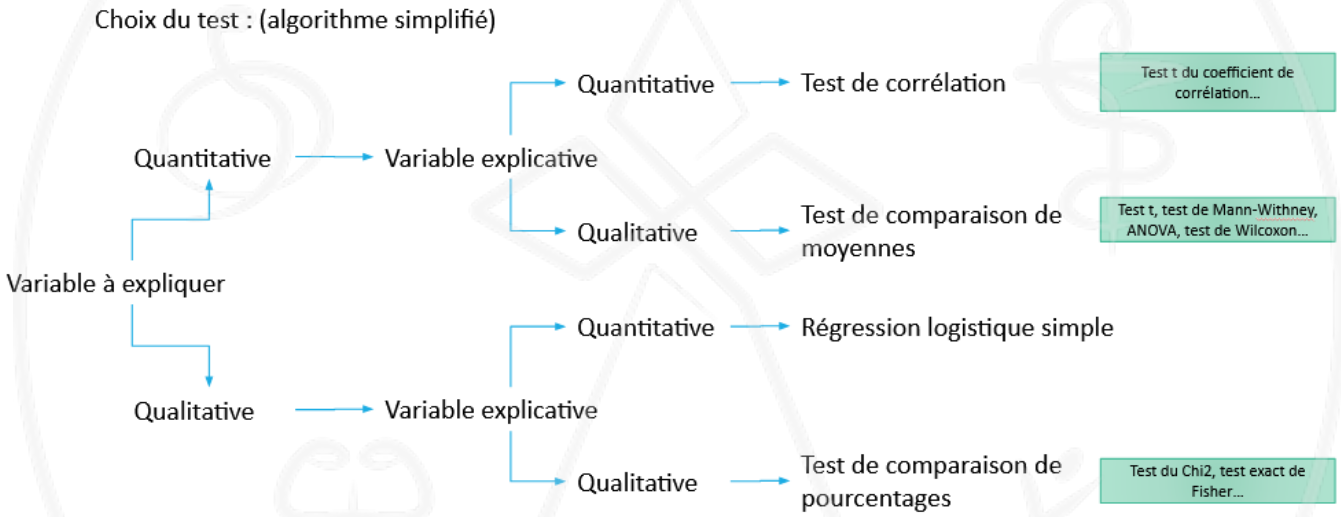


Schéma du Pr Bouzille

Ensuite, avec la valeur de t obtenue : on va pouvoir **conclure** sur l'association des deux variables :

Si :

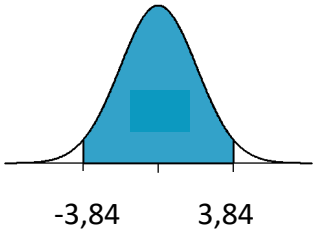
- très peu probable (généralement moins de 5%) : on rejette notre hypothèse H0 et conclue à H1 vraie.
- si probable (généralement au moins 5%) : on ne rejette pas notre hypothèse H0.

Pour cela, on utilise la table statistique (qui change en fonction du test utilié), et on prend la valeur tabulé

Pour finir, on calcule la **P-value** :

Il s'agit « d'un équivalent du risque alpha déterminé à postériori », c'est-à-dire après la réalisation du test.

Elle correspond au risque d'observer la valeur calculée avec le test statistique sous l'hypothèse nulle. Plus concrètement, c'est le risque qu'avec la valeur obtenue H0 soit vraie, c'est le pourcentage de probabilité que cette valeur puisse permettre de conclure à une hypothèse nulle vraie.



Elle permet de vérifier que le risque est bien plus petit que le seuil qu'on s'est fixé et donc que le résultat est significatif.

Si on reprend la courbe, on place la valeur obtenue au test, par exemple 3,84 ici, et on regarde qu'elle est la probabilité correspondante.

On cherche dans la table de loi du test (ici table de la loi normale) la probabilité d'observer 3,84 sous H0.

On a moins de 1% de chance d'observer cette valeur. C'est la p-value.

En pratique, on dit que :

- **Si p-value < alpha : on peut rejeter H0**
- **Si p-value ≥ alpha : on ne peut pas rejeter H0**

C'est-à-dire que si la p-value est supérieure au risque alpha, on « ne pas prendre le risque de rejeter H0 », H0 a une trop grande probabilté d'être vraie et le risque est plus grand que celui qu'on a décidé d'accepter en début du test.

Pour trouver la p-value, on va chercher dans la table le risque alpha correspondant à la valeur calculée, on utilise donc la table dans « l'autre sens ». On utilise bien la table qui correspond au test, et pour celle comprenant un degre de liberté, on se remplace sur la ligne correspondante. (voir exercice type)

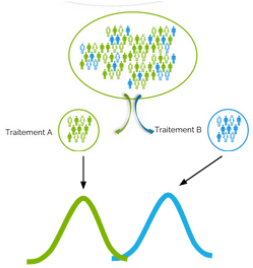
Résumé : Plusieurs étapes pour la réalisation d'un test :

1. Définir les **hypothèses** nulle et alternative
2. **Le risque alpha**
3. Test **unilatéral** ou **bilatéral**
4. Choix du **test** adapté et calcul de la statistique à partir de l'échantillon
5. On estime dans quelle mesure cette statistique peut être observée sous H0
6. Analyse des résultats avec la **p-value** : on calcule la probabilité d'observer notre statistique sous H0

PRINCIPES DES TESTS STATISTIQUES

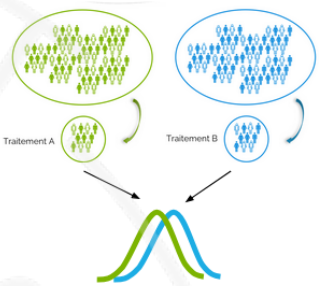
		Réalité	
		Hypothèse H_0 vraie	Hypothèse H_0 fausse
Statistique	Non rejet de H_0 = accepter H_0	Pas d'erreur	Risque β (ou risque de seconde espèce)
	Rejet de H_0 = accepter H_1	Risque α (ou risque de première espèce)	Pas d'erreur = H_1 vraie $1 - \beta$ puissance du test

La puissance correspond à la probabilité de rejeter H_0 lorsque H_0 est fausse donc lorsque H_1 est vraie.



- Le **risque** α correspond à la probabilité de rejeter H_0 lorsque H_0 est vraie (donc H_1 fausse), c'est-à-dire il s'agit du risque de faire un « faux positif ».

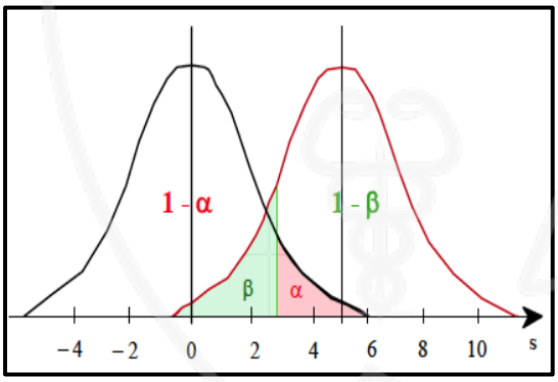
  Il s'agit du risque d'observer une différence entre les groupes alors que celle-ci n'existe pas en population.

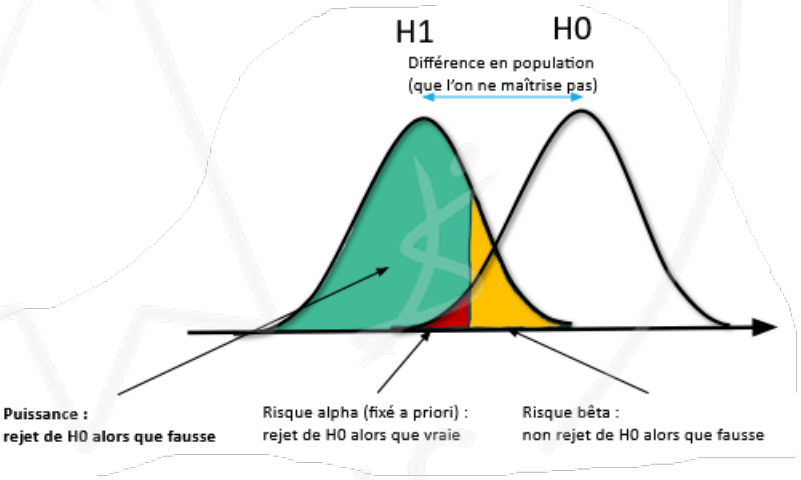


- Le **risque** β correspond à la probabilité de ne pas rejeter H_0 alors que H_0 est fausse (et H_1 vraie), c'est-à-dire faire un « faux négatif ».

Il s'agit du risque de ne pas observer de différence entre les groupes alors que celle-ci n'existe pas en population.

- **La puissance du test** (« bonne décision ») dépend de plusieurs paramètres
  - Le risque alpha
  - La différence qui existe entre les groupes en population
  - La variabilité des individus
  - La taille de l'échantillon => seul ce paramètre peut être modifié facilement
  - Plus on diminue α, plus β est grand, plus la puissance du test diminue, mais si on augmente trop alpha on prend le risque de conclure à une différence qui n'existe pas.
  - Quand n augmente, la puissance augmente

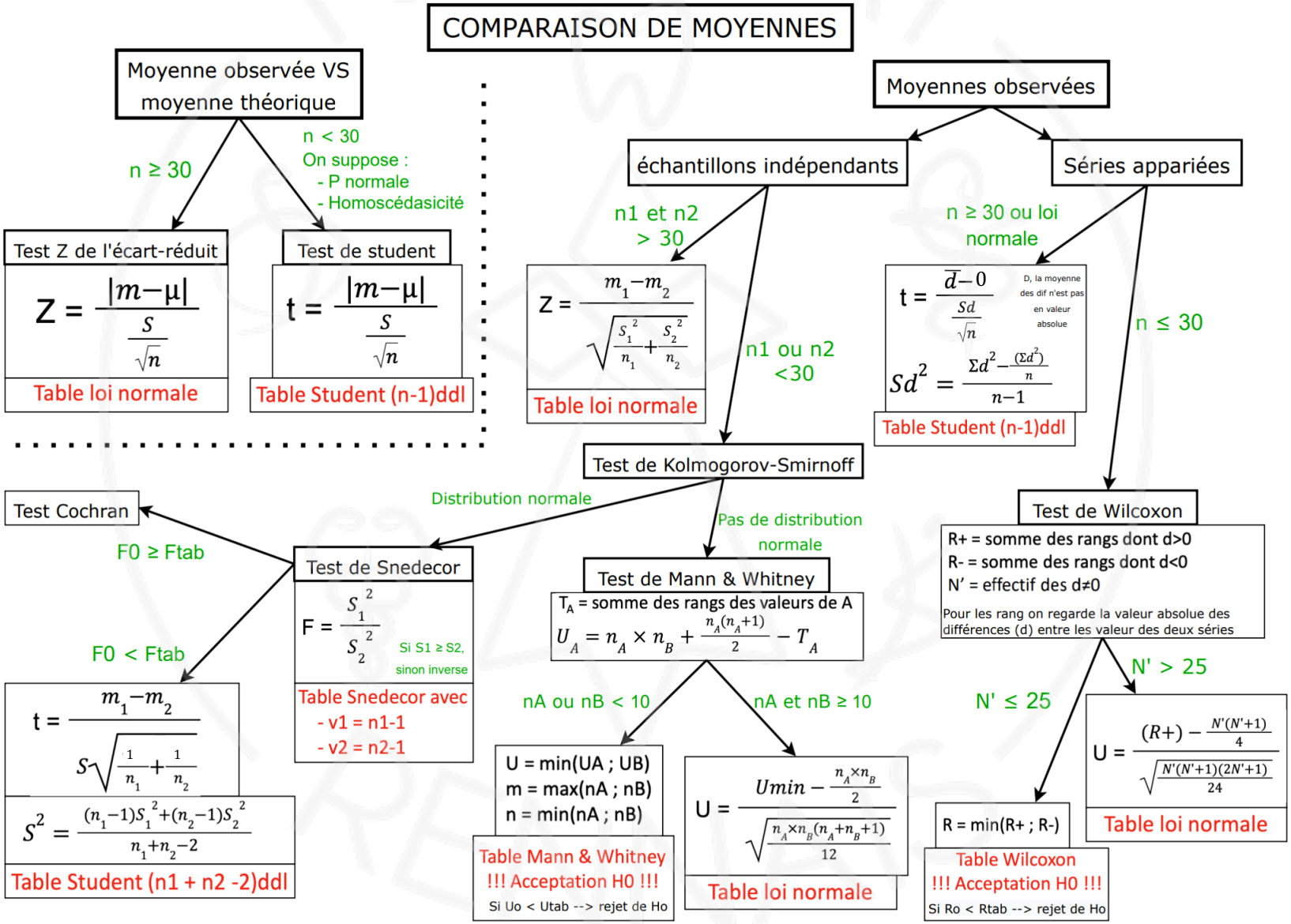




Pour déterminer la **p-value**, on cherche dans la table de **l'écart réduit** la valeur directement **inférieure à Z_0**

NOTES PERSONNELLES

TESTS DE COMPARAISON DE MOYENNES



Pour mieux comprendre comment s'attaquer aux QCMs proposés au concours pour ce chapitre nous allons voir en détail les étapes de résolution de quelques exercices de concours.

Par chance les étapes de résolution de ses exercices sont toujours les mêmes et sont guidées par les différents QCMs qui composent chaque exercice de ce chapitre.

Bon à savoir, le tout est globalement entièrement résumé dans la fiche "comparaison de moyennes".

Bon courage :)

Exercice du concours 2017/2018

On cherche à savoir si l'acuité visuelle des enfants atteints de tumeurs hypophysaires est différente en fonction de la localisation de la tumeur *(intra-sellaire ou supra-sellaire).*

Pour cela, on a mesuré *(sur 10)* l'acuité visuelle de 14 enfants tirés au sort et qui présentent l'un des deux types de tumeurs.

Les résultats sont présentés dans le tableau suivant :

Numéro sujet	Type de tumeur	Acuité visuelle
1	Intra-sellaire	9
2	Intra-sellaire	7
3	Intra-sellaire	8
4	Intra-sellaire	4
5	Intra-sellaire	2
6	Intra-sellaire	9
7	Supra-sellaire	10
8	Supra-sellaire	3
9	Supra-sellaire	10
10	Supra-sellaire	10
11	Supra-sellaire	2
12	Supra-sellaire	8
13	Supra-sellaire	6
14	Supra-sellaire	10

QCM 1 : Sachant que les conditions de normalité ne sont pas respectées, quel test réalisez-vous :

A. Un test de Z de l'écart-réduit
B. Un test de Mc Nemar
C. Un test de Wilcoxon pour séries appariées
D. Un test de Student
E. Toutes les propositions précédentes sont inexactes.

QCM 2 : On réalise le test adapté :

A. La statistique calculée du test vaut 38.
B. La statistique calculée du test vaut 17.
C. Au risque α de 5% et dans une situation bilatérale, l'acuité visuelle est significativement différente en fonction du type de tumeur.
D. Au risque α de 5% et dans une situation bilatérale, on ne peut pas rejeter l'hypothèse nulle.
E. Toutes les propositions précédentes sont inexactes.

Etape 1 *(et résolution du QCM1)* : Déterminer le test adapté à la situation

Pour cela, il suffit de suivre l'arbre de la fiche "comparaison de moyennes" et de s'orienter selon les réponses.

- Dans cette situation, on comprend que le test a pour but de comparer les moyennes d'acuité visuelle observées chez des enfants atteints de tumeurs hypophysaires en fonction de la localisation de la tumeur (intra-sellaire ou supra-sellaire).

On se trouve donc du côté "**moyennes observées**" de l'arbre.

Nb : La partie « Moyenne observée VS moyenne théorique » aurait pu correspondre à un énoncé du type *« On cherche à déterminer si la consommation de gravier à un impact sur la fréquence des vomissements ; dans la population, la moyenne des vomissement est de 3 par semaines et dans notre groupe d'individus consommants du gravier elle est représentée par ce tableau ».*

→ On remarque dans cet énoncé que le but est bien de comparer une moyenne théorique (3/semaines) avec une moyenne observée (chez les consommateurs de gravier)

Après avoir déterminé de quel côté de l'arbre on se trouve on peut continuer avec l'embranchement suivant, ici il s'agit de déterminer si les échantillons sont indépendants ou bien appariés.

- Cela revient à définir s'il s'agit de deux groupes distinct que l'on compare ou bien d'un seul et même groupe qui est étudié avec des paramètres différents. *Par exemple, on compare la moyenne de poids des élèves d'une classe avant et après avoir suivi un régime 100% Mcdonald pendant 30 jours.*

En l'occurrence il s'agit de deux "**échantillons indépendants**" dont les moyennes sont comparées, ceux avec une tumeur intra-sellaire et ceux avec une tumeur supra-sellaire.

Il faut ensuite voir si on peut utiliser la table de la loi normale ou non. Pour cela on regarde "$n1$" et "$n2$" le nombre de composants de nos deux groupes indépendants, comme au moins l'un des deux (les deux en l'occurrence) sont inférieurs à 30, le test de **Kolmogorov-Smirnoff** sera retenu.

Puis, selon si les conditions de normalité sont respectées ou non on s'orientera vers le test de Snedecor ou celui de Mann&Whitney, dans notre cas l'énoncé du QCM1 nous indique que les conditions de normalité ne sont pas respectées, on adopte donc le test de **Mann&Whitney**.

La réponse au QCM1 est donc la E.

Cette étape à été détaillée au mieux pour essayer de tout expliquer mais suivre le tableau en se posant ses questions là rend cette étape très facile et rapide, beaucoup plus qu'avec toutes mes explications :').

Il faut cependant rester vigilant pour ne pas s'y tromper car la résolution du QCM2 en découlera.

Etape 2 (et résolution du QCM2) : Utiliser le test adapté

Maintenant que nous savons quel test utiliser, il ne nous reste plus qu'à le faire, on va donc commencer par calculer **T_A** comme indiqué sur la fiche, pour cela, il faut tout d'abord classer nos valeurs par rangs, le tableau devient ainsi comme suit :

Groupe	Numéro sujet	Type de tumeur	Acuité visuelle	Rangs
A	1	Intra-sellaire	9	9,5
	2	Intra-sellaire	7	6
	3	Intra-sellaire	8	7,5
	4	Intra-sellaire	4	4
	5	Intra-sellaire	2	1,5
	6	Intra-sellaire	9	9,5
B	7	Supra-sellaire	10	12,5
	8	Supra-sellaire	3	3
	9	Supra-sellaire	10	12,5
	10	Supra-sellaire	10	12,5
	11	Supra-sellaire	2	1,5
	12	Supra-sellaire	8	7,5
	13	Supra-sellaire	6	5
	14	Supra-sellaire	10	12,5

Nb : Si deux (ou plus) valeurs ou plus sont identiques, on leur attribue toutes la moyenne des rangs qu'elles auraient obtenus (ex ; trois fois 4 aux rangs 6,7 et 8, la moyenne est 7 donc tous les 4 auront le rang "7").

On fait les rangs sur les 2 groupes en même temps A et B pour les tests de comparaison de moyenne. Ici les deux premiers rangs correspondent aux deux 2,

C'est-à-dire qu'en rang n°1 et en rang n°2 on aura 2 comme valeur.

Pour déterminer le rang quand on a des valeurs identiques, on fait la moyenne des rangs, ici on obtient donc (1+2)/2 = 1,5 pour les deux valeurs 2.

On poursuit ensuite avec la valeur suivante qui est ici 3, elle est unique elle obtient donc le rang 3. Puis on attribut à la valeur 4, le rang 4. Ensuite à la valeur 6, le rang 5 et ainsi de suite...

Comme T_A est la somme des rangs des valeurs de A, on a :

- **T_A**= 9,5 + 6 + 7,5 + 4 + 1,5 + 9,5 = **38**
- **T_B** = la somme des rangs des valeurs du groupe B = **67**

Avec ça, on peut maintenant calculer U_A et U_B avec la formule :

$$U_A = n_A \times n_B + \frac{n_A(n_A+1)}{2} - T_A$$ *Sachant que* ***n_A = 6*** *et* ***n_B = 8***

On a donc $U_A = 6 \times 8 + \frac{6 \times 7}{2} - 38 = 31$

On a aussi $U_B = n_B \times n_A + \frac{n_B(n_B+1)}{2} - T_B = 8 \times 6 + \frac{8 \times 9}{2} - 67 = 17$

Étape 3 (et résolution du QCM2) : Lire le résultat dans la table correspondante

Ensuite, si l'on observe notre fiche de comparaison de moyenne, on remarque qu'après avoir calculer T_A et U_A, deux voies sont possibles, étant donné que n_A (et n_B) sont inférieures à 10, c'est la table de Mann&Whitney qui sera utilisée.

- $\boldsymbol{U_0 = min\,(U_A, U_B) = U_B = 17}$,
- m sera égal à 8, car $\boldsymbol{n_B > n_A}$ et n sera égal à **6** car $\boldsymbol{n_A < n_B}$.

Avec cela nous pouvons enfin lire le résultat dans la table correspondante, pour le seuil demandé, **α de 5%,** la table est la suivante, on a n = 6 et m - n = 2, la statistique tabulée sera donc 8.

TEST DE MANN-WHITNEY

m − n	n = 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	—	—	—	0	2	5	8	13	17	23	30	37	45	55	64	75	87	99	113	127
1	—	—	—	1	3	6	10	15	20	26	33	41	50	59	70	81	93	106	119	
2	—	—	0	2	5	8	12	17	23	29	37	45	54	64	75	86	99	112		
3	—	—	1	3	6	10	14	19	26	33	40	49	59	69	80	92	105			
4	—	—	1	4	7	11	16	22	28	36	44	53	63	74	85	98				
5	—	—	2	4	8	13	18	24	31	39	47	57	67	78	90					
6	—	0	2	5	9	14	20	26	34	42	51	61	72	83						
7	—	0	3	6	11	16	22	29	37	45	55	65	76							
8	—	0	3	7	12	17	24	31	39	48	58	69								
9	—	0	4	8	13	19	26	34	42	52	62									
10	—	1	4	9	14	21	28	36	45	55										
11	—	1	5	10	15	22	30	38	48											
12	—	1	5	11	17	24	32	41	50											
13	—	1	6	11	18	25	34	43												
14	—	1	6	12	19	27	36	45												
15	—	2	7	13	20	29	38													
16	—	2	7	14	22	30	40													
17	—	2	8	15	23	32														
18	—	2	8	16	24	33														
19	—	3	9	17	25															
20	—	3	9	17	27															
21	—	3	10	18																
22	—	3	10	19																
23	—	3	11																	
24	—	4	11																	
25	—	4																		
26	—	4																		

Table V.a

(Seuil α = 0,05)

Nb : nous sommes dans le cas d'un test **non paramétrique**, donc, si **Uo < Utab → rejet de Ho**.

En l'occurrence $U_0 = 17$ et $U_{tab} = 8$, donc $U_0 > U_{tab}$.

On accepte H_0, au risque alpha de 5%, dans une situation bilatérale, l'acuité visuelle n'est pas significativement différente en fonction du type de tumeur.

La réponse au QCM 2 est la réponse D.

NOTES PERSONNELLES

TEST DE COMPARAISON DE FREQUENCES

Comparaison d'effectifs
Tests du X² (Chi²)

Fréquence observée VS Fréquence théorique

Tous les Ci ≥5

X² d'ajustement

$$x^2 = \Sigma \frac{(O_i - C_i)^2}{C_i}$$

Table X² (i-1)ddl

Fréquences observées

Ci <5 : Regrouper les effectifs

Ci ≥5

X² d'homogénéité

$$x^2 = \Sigma \frac{(O_i - C_i)^2}{C_i}$$

Table X² ((c-1)*(l-1))ddl

X² corrigé de Yates

$$x^2 = \Sigma \frac{(|O_i - C_i| - 0{,}5)^2}{C_i}$$

Table X² ((c-1)*(l-1))ddl

Comparaison de pourcentages
Tests de l'écart réduit

Fréquence observée VS Fréquence théorique

n*p et n*q ≥5

$$U = \frac{|p_0 - p|}{\sqrt{\frac{pq}{n}}}$$

Table loi normale

Test de l'écart réduit

Fréquences observées

nA*p, nB*p, nA*q et nB*q ≥5

$$U = \frac{p_A - p_B}{\sqrt{\frac{pq}{n_A} + \frac{pq}{n_B}}}$$

$$p = \frac{n_A \times p_A + n_B \times p_B}{n_A + n_B}$$

q = 1 - p

Table loi normale

Comparaison de séries appariées

Test de Mac Nemar

$$U = \frac{n_2 - n_3}{\sqrt{n_2 + n_3}}$$

Table loi normale

Après \ Avant	+	-
+	n1	n2
-	n3	n4

n2 + n3 ≥10

$$x^2 = \frac{(n_2 - n_3)^2}{n_2 + n_3}$$

Table X² 1ddl

Légende

i : nombre de colonnes

Oi : effectif observé (donné)

Ci : effectif théorique $C_i = \frac{\text{total ligne} \times \text{total colonne}}{\text{Total général}}$

	G+		G-		
A	45	38,3	55	61,7	100
B	70	76,7	130	123,3	200
	115		185		300

EXERCICE TYPE CORRIGE : COMPARAISON DE FREQUENCE

Annale 2017-2018 :

On souhaiterait savoir si le taux de récidive du cancer est dépendant du traitement proposé aux patients (chimiothérapie, chirurgie et radiothérapie). On constitue au hasard 3 groupes de patients en fonction du traitement qu'ils ont reçu et on regarde au bout de 2 ans combien de patients ont récidivé.

Les résultats sont résumés dans le tableau ci-dessous :

	Chimiothérapie	Chirurgie	Radiothérapie
Pas de récidive	400	350	450
Récidive	120	90	130

QCM 26 : On choisit de réaliser un test du Chi^2 :

A. Il s'agit d'un test sans condition d'application particulière.
B. Son calcul repose sur les effectifs observés des différents groupes.
C. Son calcul repose sur les différences entre les effectifs observés et les effectifs théoriques sous l'hypothèse nulle.
D. Son interprétation peut se faire avec la table du Chi^2.
E. Toutes les propositions précédentes sont inexactes.

QCM 27 : On réalise maintenant le test du Chi^2 :

A. Au risque a de 5 % et dans une situation bilatérale, la statistique calculée du test vaut 7,29 (valeur absolue à 0,01 près)
B. Au risque a de 5 % et dans une situation bilatérale, la statistique tabulée (ou théorique) du test vaut 5,99
C. On peut rejeter l'hypothèse nulle H_0 et conclure que le taux de récidive est significativement différent en fonction du traitement.
D. En rejetant l'hypothèse nulle, on a plus d'une chance sur 10 de se tromper de conclusion
E. Toutes les propositions précédentes sont inexactes.

Les QCMs de ce chapitre vont souvent par pair avec un **premier QCM de cours** et un **second de calculs**.

Dans le cas de **valeurs non appariées**, c'est-à-dire que les données ne sont pas liées puisqu'il s'agit d'un groupe différent pour chaque traitement ou test (on ne réalise pas un test avant/après) :

- Si on a déjà les **valeurs théoriques (Ci)** données dans l'énoncé, on utilise la formule du **Chi^2 d'ajustement** *(cf schéma précédent).* Ces valeurs théoriques sont alors souvent identifiées comme une "moyenne nationale" ou celle que l'on devrait obtenir selon une loi.
- Dans le cas où l'énoncé dispose de deux populations (échantillons) testées différemment, il n'y a aucune donnée préalable, elles sont toutes observées. C'est le cas ici. On utilise la formule du **Chi^2 d'homogénéité.**

Correction QCM 26 : BD

A. FAUX. Le test du Chi2 a bien des conditions d'applications particulières. Il faut que les Ci (valeurs théoriques) soient supérieures ou égales à 5.

B. VRAI. On est dans le cas de fréquences observées uniquement.

C. FAUX. On n'a pas de fréquences théoriques.

D. VRAI. On utilise la table du Chi^2 (*voir fiche précédente*).

CORRECTION 27 : BD

Pour résoudre les exercices : procéder par étape et être bien rigoureux.

	Chimiothérapie	Chirurgie	Radiothérapie	Total
Pas de récidive	**400** **405,19**	**350** **342,86**	**450** **451,95**	1200
Récidive	**120** **114,81**	**90** **97,14**	**130** **128,05**	340
Total	520	440	580	1540

1. Calculer les Ci

Pour cela, il faut d'abord *faire le tableau avec les données* de l'exercice s'il n'est pas donné dans l'énoncé, et *ajouter les totaux*.

On calcule ensuite les effectifs théoriques Ci avec $Ci = \frac{(Total\ ligne) \times (total\ colonne)}{Total\ général}$.

- Exemple : Case "chimiothérapie - pas de récidive" : $\frac{1200 \times 520}{1540}$ = 405,19 (à 0,01 près)
- Exemple : Case "chimiothérapie - récidive" : $\frac{340 \times 520}{1540} = 114{,}81$ (à 0,01 près)

Étape qu'on n'a pas besoin de faire dans le cas du Chi^2 d'ajustement.

2. Tous les $Ci \geq 5$, on applique donc la formule du Chi² d'homogénéité :

(Cf fiche précédente)

$$X^2\ calculé\ = \Sigma \frac{(Oi - Ci)^2}{Ci}$$

- **Oi : valeurs observées**
- **Ci : valeurs théoriques**

Soit $X^2\ calculé$ = $\frac{(400-405{,}19)^2}{405{,}19} + \frac{(120-114{,}81)^2}{114{,}81} + \frac{(350-342{,}86)^2}{342{,}86} + \frac{(90-97{,}14)^2}{97{,}14} + \frac{(450-451{,}95)^2}{451{,}95} + \frac{(130-128{,}05)^2}{128{,}05}$

Soit $X^2\ calculé$ = 0,07 + 0,23 + 0,15 + 0,52 + 0,01 + 0,03 = **1,01**

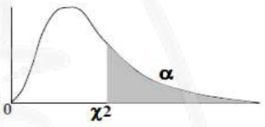
3. On cherche ensuite le $X^2\ tabulé$ dans la table du Chi².

On calcule le nombre de **degré de liberté** (**ddl**).

Nombre de ddl = (nombre de colonnes - 1) × (nombre de lignes - 1) = 2 × 1 = 2

Table de la loi du χ^2

La table donne la probabilité α pour que χ_2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



ddl \ α	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,12
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	39,25
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82

Le **risque** α que l'on prend est celui donné dans l'item *(très souvent 5% mais parfois 1%)* et le ddl est calculé selon la formule précédente.

Il ne reste plus qu'à lire dans le tableau.

On trouve ainsi $X^2\ tabulé$ = **5,99**.

- *Avec* $\alpha = 5\%$ *et 2ddl*

4. Interprétation

Dans une situation **bilatérale** :

- H_0 = **homogénéité** entre les différents traitements
- H_1 = **différence** entre les différents traitements (pas de notion de + ou - entre les traitements).

On trouve : $X^2\ calculé < X^2\ tabulé$ (1,01 < 5,99).

Donc au risque $\alpha = 5\%$, on accepte Ho et on rejette H1. On ne montre pas de différence significative entre les différents traitements proposés concernant le taux de récidive.

5. P-value

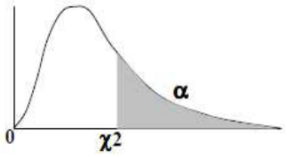
On cherche la **p-value** dans la table du Chi^2.

On cherche **dans la ligne du ddl** calculé précédemment *(soit ici **2**ddl)* la valeur qui se rapproche le plus du X^2 *calculé* = 1,01.

Elle se trouve entre deux valeurs : 0,21 < **1,01** < 4,61

Table de la loi du χ^2

La table donne la probabilité α pour que χ_2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



ddl \ α	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,12
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12

On lit dans la ligne de α les valeurs de la p-value correspondante :

- La p-value se trouve entre 0,900 et 0,100 soit 90% et 10%.
- 0,900 > p-value > 0,100
- 90% > p-value > 10% (1/10)

On a donc plus d'une chance sur dix de se tromper.

Plus exactement, la p-value vaut **50%**, on a donc une chance sur deux de se tromper.

1,01 - 0,21 = 0,8

Or, on constate dans le début de la ligne qu'une différence de 5% d'alpha correspond à une différence de 0,10 dans la valeur de X^2

$$\frac{0,8 \times 0,05}{0,1} = 0,4$$

0,9 - 0,4 = 0,5 = **50%**.

RAPPEL :

- *Lorsque X^2 calculé > X^2 tabulé , on rejette Ho.*
- *Lorsque X^2 calculé < X^2 tabulé , on accepte Ho.*

Exception : test de Mann & Whitney et Wilcoxon : c'est l'inverse !

NOTES PERSONNELLES

CORRELATIONS et REGRESSIONS

CORRELATION : EVALUER LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

<table>
<tr><th colspan="2">Test paramétrique de PEARSON</th><th>Test non paramétrique de SPEARMAN</th></tr>
<tr><td colspan="2">Estimation de ρ (coef de corrélation linéaire) par r :
● Si ρ < 0 : Les variables sont liées et évoluent en sens opposé
● Si ρ > 0 : Les variables sont liées et évoluent dans le même sens
● Si ρ = 0 : Les variables sont indépendantes (-1 ≤ ρ ≤ 1)</td><td>Coef de corrélation r_s sur les rangs

Intensité du lien monotone croissant ou décroissant (suivant le signe du coef).</td></tr>
<tr><td colspan="2">Bi-normalité Linéaire</td><td>Test de rangs
Monotone croissant ou décroissant</td></tr>
<tr><td colspan="2">Calcul de r (qui est une estimation de ρ)</td><td>Calcul de r_s</td></tr>
<tr><td>Calculatrice*[1]

Data L1 pour la première ligne et L2 pour la deuxième
*Pour supprimer vos données appuyer 2 fois sur Data Clear L1.
● Faire stat-reg/distr puis 2-Var Stats
● Choisir L1 pour xDATA, L2 pour yDATA et ONE pour FRQ puis CALC
● Descendre : on trouve r</td><td>$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{(\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2)(\sum_{i=1}^{n} {y_i}^2 - n\bar{y}^2)}}$$</td><td>Etablir des rangs pour les 2 colonnes. Ce sont les rangs de chaque valeur qu'il faudra désormais rentrer.
* Si on trouve plusieurs valeurs identiques : Faire la moyenne de leurs rangs

Calculatrice

Rentrer ces rangs comme données dans la calculatrice :
● Suivre la même méthode*[1]
Descendre : on trouve r_s</td></tr>
<tr><td colspan="2" rowspan="2">Test Statistique : Est-ce que r est significativement ≠ de 0 ?
H_0 : ρ = 0 et H_1 : ρ ≠ 0
Statistique de test $t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$
t_α sur la table de Student à (n-2) ddl au risque α
● Si $|t_0| \geq t_\alpha$: On <u>rejette</u> H_0. Il existe un <u>lien linéaire</u>
● Si $|t_0| < t_\alpha$: On ne rejette pas H_0. On ne peut pas affirmer de lien linéaire.</td><td>$n \leq 10$
On compare directement r_s à la valeur du coef de Sperman à n ddl</td><td>$n > 10$
$t_0 = \frac{r_s\sqrt{n-2}}{\sqrt{1-{r_s}^2}}$
Table de Student à n-2 ddl.</td></tr>
<tr><td colspan="2">Pour lire la table en fonction de si le test se fait sous une hypothèse bilatérale ou unilatérale :

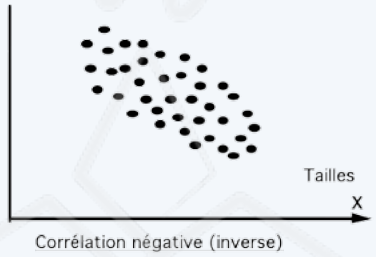
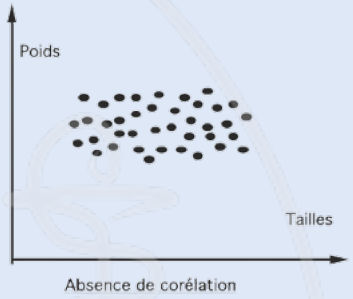
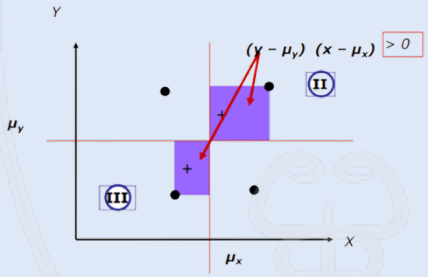
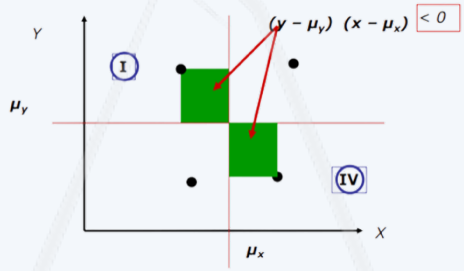
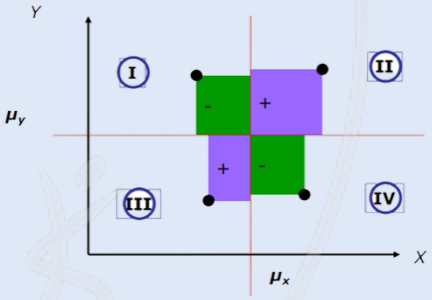
1α Bilatéral | Faire ×2 sur α Unilatéral</td></tr>
</table>

REGRESSION LINEAIRE : voir si la courbe des variations de y en fonction de x est une droite

$\boldsymbol{b = \bar{y} - a\bar{x}}$	$\boldsymbol{a = r\frac{Sy}{Sx}}$
Test : est-ce que a est différent de 0 ? $H_0 : a = 0$; $H_1 : a \neq 0$	
$$\boldsymbol{t_0 = \frac{a\sqrt{n-2}}{\sqrt{\frac{S_y^{\,2}}{S_x^{\,2}} - a^2}}}$$	
Table de Student à $n-2$ ddl	

NOTION DE COVARIANCE :

- Indicateur de mesure de **liaison entre 2 variables X et Y**. (= mesure de la variation simultanée)
- Moyenne des produits des écarts entre X et Y divisée par leurs moyennes respectives µx et µy
- La covariance peut être illustrée par la somme des aires des rectangles

Diagramme de dispersion (nuage de points de coordonnées (x ; y)		
Corrélation positive (directe)	Corrélation négative (inverse)	Absence de corrélation
		 *Nuage de points diffus <=> Absence de l° entre les caractères étudiés*
 *Les rectangles Mauves dans les zones II et III représentent les produits positifs.*	 *Les rectangles verts dans les zones I et IV représentent les produits négatifs*	
$Cov > 0 \Leftrightarrow \sum aires > 0$	$Cov < 0 \Leftrightarrow \sum aires < 0$	$Cov = 0 \Leftrightarrow \sum aires = 0$
X et Y évoluent dans le même sens **=> plus** $X \nearrow$ *plus* $Y \nearrow$	X et Y évoluent dans le sens opposé **=> plus** $X \nearrow$ *plus* $Y \searrow$	X et Y indépendantes *=> les aires se compensent : les points sont répartis uniformément dans les quadrants*
Cas du test unilatéral à **droite** : Test $H_0 : \rho = 0$ **contre** H_1 **: ρ > 0** *Rejeter* H_0 *revient alors à mettre en évidence une corrélation linéaire positive.*	Cas du test unilatéral à **gauche** : Test $H_0 : \rho = 0$ **contre** H_1 **: ρ < 0** *Rejeter* H_0 *revient alors à mettre en évidence une corrélation linéaire négative*	

POINTS IMPORTANTS A MAITRISER :

→ Partie corrélation

1. **Analyser un graphique** pour identifier le lien entre les variables et savoir quel calcul de coefficient (Pearson / Spearman) est suggéré.
2. **Calculer les coefficients :**
    - **À partir de la formule**, c'est ce que l'on retrouve désormais le plus souvent à l'examen.
        - *(Cf annales 2020-2021 et 2021-2022)*
    - /!\ Sachez aussi le faire **à partir du tableau** grâce à la calculatrice (comme dans l'exemple ici)
        - *(Cf méthodo calculatrice)*
3. **Effectuer le test**
    - /!\ Être très vigilant avec les ddl et les risques /!\
4. **Interpréter le test**

→ Partie régression

1. **Calculer le coefficient de régression**
2. **Calculer l'ordonnée à l'origine**
3. **Faire le test**

NB : le test de la **pente** α de la droite de régression est strictement le même que le test du coefficient de Pearson.

4. **Interpréter le test**
5. **Identifier si un point appartient ou non à la droite de régression.**

EXERCICE TYPE CORRIGE : CORRELATIONS REGRESSIONS

Reprise d'un exercice de l'annale 2014-2015 :

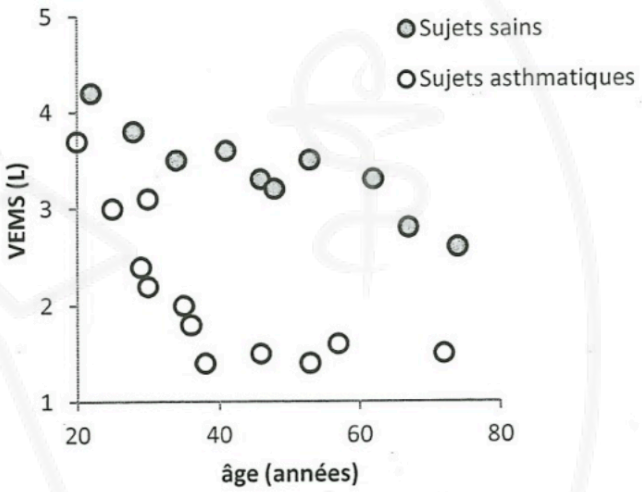
On a mesuré le **VEMS** (*Volume Expiratoire Maximal Seconde*) chez 10 sujets sains et 12 sujets souffrant d'asthme. On s'intéresse au lien entre le VEMS et **l'âge** dans chacune des populations (sujets sains et sujets asthmatiques). Les mesures sont présentées dans le tableau et le graphique ci-dessous.

Le VEMS est distribué selon une loi normale dans chaque population (sujets sains et sujets asthmatiques). Chez les sujets sains les conditions d'application de l'analyse par régression linéaire sont satisfaites. Chez les sujets asthmatiques le diagramme de dispersion fait douter de la linéarité du lien entre VEMS et âge.

Dans la suite toutes les valeurs numériques proposées sont telles que les âges sont exprimés en années et les VEMS en litres.

Sujets sains	
Âge (années)	VEMS (L)
22	4,2
28	3,8
34	3,5
41	3,6
46	3,3
48	3,2
53	3,5
62	3,3
67	2,8
74	2,6

Sujets asthmatiques	
Âge (années)	VEMS (L)
20	3,7
25	3
29	2,4
30	2,2
30	3,1
35	2
36	1,8
38	1,4
46	1,5
53	1,4
57	1,6
70	1,5



QCM 1 : Corrélation Chez les sujets asthmatiques.

A. Le doute émis sur la linéarité du lien entre le VEMS et l'âge impose le choix d'un test paramétrique pour évaluer l'existence d'un lien entre ces variables.
B. Le coefficient r_s de Spearman vaut 0,865 (à 0,001 près).
C. La statistique du test sur le coefficient de Spearman vaut 4,87 en valeur absolue (à 0,01 près).
D. Au seuil de risque 5%, on conclut qu'il existe un lien monotone décroissant entre les variables. La valeur de la table de Student utilisée pour faire cette conclusion est 1,812.
E. Au seuil de risque 5%, on conclut qu'il n'existe pas de lien monotone entre les variables. La valeur de la table de Student utilisée pour faire cette conclusion est 2,228.
F. Toutes les réponses sont inexactes.

Correction QCM1 : D

A. FAUX, un test NON paramétrique.
B. FAUX, le coefficient r_s de Spearman vaut -0,865 (à 0,001 près). **/!\ SIGNES /!**
C. FAUX : $|t_0| = 5,44$.
D. VRAI.
E. FAUX.

QCM 2 : Régression : Chez les sujets sains, concernant la droite de régression du VEMS en fonction de l'âge :

A. La pente de la droite vaut -0,025.
B. L'ordonnée à l'origine de la droite vaut 161,46 (à 0,01 près).
C. La droite de régression linéaire permet de prédire le VEMS pour un âge donné.
D. La droite de régression linéaire passe par le pont de coordonnées (47,5 ; 3,38).
E. La droite de régression linéaire passe par le point de coordonnées (22 ; 4,2).
F. Toutes les réponses sont inexactes.

METHODE DE RESOLUTION :

Choix du test

Conditions d'application des tests des coefficients :

- **Pearson** (*paramétrique*) : au moins une de ces 2 conditions doit être remplie :
  - La distribution du couple de variables aléatoires (X ; Y) doit être **bi-normale**, c'est-à-dire que pour chaque valeur de X, la distribution de Y est normale et vice-versa.
  - La distribution de **Y est normale** et de **variance constante** pour chaque valeur de X ou vice versa ET la relation entre **X et Y est linéaire**.

*NB : Cela est difficile à vérifier en pratique**, le plus souvent on vérifie la normalité de X ou de Y et le caractère monotone de leur relation** (examen visuel des données).*

Le test est robuste c'est à dire qu'il accepte quelques écarts à ces conditions d'applications.

- **Spearman** (*non-paramétrique*) : Aucune
  - On réalise donc le test de Spearman quand il est impossible de remplir les conditions du test de Pearson ou qu'on ne peut pas les vérifier.

Méthode pour calculer le coefficient de Spearman

Pour réaliser le test de **Spearman**, on n'utilise pas les paramètres des 2 variables contrairement à celui de Pearson, mais on **attribue des rangs** à chaque valeur en les classant par ordre croissant ce qui donne le tableau suivant.

Nb : *si on a 2 valeurs identiques, on fait la moyenne.*

- Ici, le 4ème et le 5ème ont 30, donc les 30 ont un rang de $\frac{(4+5)}{2} = \mathbf{4,5}$

On rentre ensuite dans la **calculatrice**, non pas les valeurs, mais les rangs attribués (cf méthode du livret).

La calculatrice nous donne alors les différents paramètres qui nous intéressent, dont le **coefficient** r_s de Spearman.

Age (années)	Rang	VEMS(L)	Rang
20	1	3,7	12
25	2	3	10
29	3	2,4	9
30	**4,5**	2,2	8
30	**4,5**	3,1	11
35	6	2	7
36	7	1,8	6
38	8	1,4	1,5
46	9	1,5	3,5
53	10	1,4	1,5
57	11	1,6	5
70	12	1,5	3,5

/!\ Bien prendre le tableau concernant les asthmatiques /!

Effectuer le test de Spearman

$n > 10$, donc on applique la formule vue ci-dessus avec $t_0 = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}}$ Table **de student** $(n - 2ddl)$

- $r_s = -0{,}865$ et $n = 12$
- Au risque $2 * \alpha$ Unilatéral

Alors, $|t_0|$ **= 5,44**

Interpréter le test

$n > 10$, on réalise donc un **test unilatéral**.

Pour trouver la valeur dans la table, comme $n > 10$, on prend :

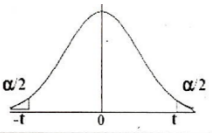
- **n-2ddl**, donc 12-2 = **10ddl**
- Au risque **2α**, soit **2 x 0,05 = 0,1**

- On a donc $t_{2\alpha} = 1{,}812$
- $|t_0| > t_{2\alpha}$ 5,44 > **1,812**

Table du *t* de Student

La table donne la probabilité α pour que *t* égale ou dépasse, en valeur absolue, une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).

2α



ddl \ α	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,599
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
n-2ddl → 10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437

On rejette H_0, il existe bien un lien **monotone** (croissant / décroissant).

De plus, le coefficient de Spearman étant négatif, c'est donc un lien **monotone <u>décroissant</u>**

Correction QCM2 : ACE

A. **VRAI**. Utiliser la calculatrice, il faut rentrer dans un tableau les 2 variables concernant les sujets sains, de la même manière que pour les coefficients de Pearson et Spearman. Puis, il faut identifier *a* dans les valeurs proposées.
B. **FAUX**, $b = \bar{y} - a\bar{x}$
On prend les valeurs moyennes de x et y qu'on obtient en rentrant le tableau dans la calculatrice de la même manière que vu précédemment.
Donc b = 3,38 - (-0,025) x 47,5 = 4,569
C. **VRAI**, grâce à l'équation y = ax + b avec a pente de la droite et b l'ordonnée à l'origine, soit y = -0,025x + 4,569 avec x l'âge et y le VEMS.
Par exemple, pour une personne âgée de 32 ans, on peut évaluer que le VEMS sera de 3,77 (à 0,01 près). y = ax + b = -0,025 x 32 + 4,569 = 3769
D. **FAUX**, pour x = 22, on a y = 4,02 car y = -0,025 x 22 + 4,569
E. **VRAI**, pour x = 32, on a y = 3,77 car y = -0,025 x 32 + 4,569

QCM BONUS : Questions de cours les plus fréquentes

A. Le coefficient directeur de la droite de régression est inversement proportionnel au coefficient de corrélation linéaire (Pearson)
B. La droite de régression a pour particularité d'avoir une ordonnée à l'origine nulle.
C. La corrélation permet d'évaluer 2 variables qualitatives.
D. Le coefficient de corrélation de Spearman se situe toujours dans l'intervalle [-1 ; 1].
E. Le coefficient de corrélation de Spearman est toujours égal au coefficient de corrélation de Pearson.
F. Le calcul du coefficient de corrélation de Pearson repose sur la notion de covariance.

Correction QCM Bonus : ADF

A. VRAI. $a = \frac{s_y}{s_x}$ ou $r = a\frac{s_x}{s_y}$

B. FAUX. $b = \bar{y} - a\bar{x}$
C. FAUX, quantitatives.
D. VRAI, comme celui de Pearson.
E. FAUX, il prend en compte les rangs, donc le résultat change parfois.
F. VRAI, cette formule est à connaître par ❤️.

L'estimation du coefficient de corrélation **est donnée par :**

$$r = \frac{cov(x,y)}{\sqrt{{s_x}^2 {s_y}^2}}$$

$$r = \frac{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$r = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\,\bar{y}}{\sqrt{(\sum_{i=1}^{n} {x_i}^2 - n\bar{x}^2)(\sum_{i=1}^{n} {y_i}^2 - n\bar{y}^2)}}$$

NOTES PERSONNELLES

EPIDEMIOLOGIE

Attention : les QCM sur ce chapitre peuvent porter sur le cours, ne vous contentez pas d'apprendre les formules.

MESURES EN EPIDEMIOLOGIE :

	Définition	Numérateur & Dénominateur	Résultat	Exemple
Proportion	Similaire à la fréquence. *Soit A = population et B une partie de A,* La proportion **de** B **par rapport à** A est définie par : $= \frac{b}{a}$ $= \frac{effectif\ de\ B\ (nb\ d'éléments\ de\ B)}{effectif\ de\ A\ (nb\ d'éléments\ de\ A)}$	Le numérateur est <u>une partie</u> du dénominateur. Ils sont de MÊME nature.	Pourcentage Compris entre 0 et 1 $(a > b)$.	*Dans une population de 490 élèves, 30 sont en antenne. Quelle est la* **proportion** *d'élèves en antenne ?* *30/490 = 0,061 = 6,1%*
Ratio	Rapport entre les effectifs de **2 classes** d'<u>une même variable</u>. *Soit A, une population et B, une seconde population, toutes deux faisant partie d'un même groupe.* Le ratio de B par rapport à A est défini par : $\frac{b}{a} = \frac{effectif\ de\ B}{effectif\ de\ A}$	Le numérateur et le dénominateur sont <u>exclusifs</u> l'un de l'autre. Ils ne font pas partie de la même classe. Ils sont de MÊME nature.	Sans unité	*Dans une population de 210 élèves, 150 sont des PASS et 60 sont des LAS.* *Quel est le* **ratio** *PASS/LAS ?* *150/60 = 2,5*

Cote	**Ratio de la probabilité** de **survenue** d'un évènement sur la probabilité de **non-survenue** de cet événement. $\simeq Pari$ La cote d'un événement est définie par $\frac{P(survenue\ de\ l'événement)}{P(non\ survenue\ de\ l'événement)}$	Le numérateur et le dénominateur sont **exclusifs** l'un de l'autre. Ils ne font pas partie de la même classe. Ils sont de MÊME nature.	Probabilité Compris entre 0 et 1 ($a > b$).	*Dans une population de 100 PASS, il y a 20 places pour passer en 2ème année.* *Quelle est la* **côte** *pour passer en 2ème année ?* *20/80 = 0,25*
Indice	**Rapport** de 2 effectifs qui sont **de natures différentes.** Soit A une population et B une seconde population d'objets distincts. L'indice de B par rapport à A est défini par : $\frac{b}{a} = \frac{effectif\ de\ B}{effectif\ de\ A}$	Le numérateur et le dénominateur sont exclusifs. Ils ne sont pas de la même nature.	Sans unité	*Dans un hôpital, on a 5 médecins et 20 lits. Quel est* **l'indice** *du nombre de lits par médecin ?* *20/5 = 4*
Taux	**Rapport** qui prend en compte la notion du **temps**, c'est-à-dire la probabilité qu'un évènement se réalise au cours du temps. Le taux se définit par : $\frac{Effectif^{\ qui\ a\ connu\ un\ événement}}{Effectif^{\ susceptible\ de\ connaître\ l'événement}}$	Le <u>dénominateur</u> est l'ensemble des individus **susceptible de connaître** cet événement au cours du temps. Le <u>numérateur</u> sont les individus **qui ont connu** cet événement **au cours du temps**.	Sans unité	**Taux** *brut de natalité =* $\frac{Nb^{\text{ naissances viables de l'année}}}{Pop^{\text{au milieu de l'année}}}$ **Taux** *global de fécondité =* $\frac{Nb^{\text{annuel de naissances}}}{Nb^{\text{femmes en âge de procréer}}}$

INDICATEURS EPIDEMIOLOGIQUES :

	Morbidité = fréquence de la maladie	Mortalité = fréquence des décès
À l'instant t	**Prévalence** : $$\frac{\textbf{nombre de cas}_{\textbf{observé à l'instant t}}}{\textbf{population à risque}_{\textbf{à l'instant t}}}$$ Elle ne prend pas en compte la durée de la maladie (malade depuis 3 jours ou 3 mois, pas de différence). Elle est utilisée pour les maladies chroniques ++.	**Létalité :** $$\frac{\textbf{nombre de décès}_{\textbf{dû à une maladie X}}}{\textbf{population}_{\textbf{atteinte de la maladie X}}}$$ Elle décrit la gravité de la maladie et la qualité de sa prise en charge.
Sur une période	**Incidence cumulée :** $$\frac{\textbf{nb de nouveaux cas}_{\textbf{sur une période } \Delta t}}{\textbf{population à risque}_{\textbf{sur une période } \Delta t}}$$ Plus la maladie est contagieuse, plus l'IC est grand.	**Mortalité globale :** $$\frac{\textbf{nb de décès}_{\textbf{sur une période } \Delta t}}{\textbf{population}}$$ Correspond à l'incidence cumulée mais pour les décès. **Mortalité spécifique** : $$\frac{\textbf{nb de décès}_{\textbf{dû à la cause X sur une période } \Delta t}}{\textbf{population}_{\textbf{étudiée sur une période } \Delta t}}$$ Elle peut être également sur la mortalité d'une certaine classe d'âge (mortalité infantile). **Mortalité proportionnelle** : $$\frac{\textbf{nb de décès dû à la cause X}_{\textbf{sur une période } \Delta t}}{\textbf{nb de décès}_{\textbf{sur une période } \Delta t}}$$ Elle permet de rendre compte du poids d'une maladie dans le nombre de décès.

NOTION D'EPIDEMIOLOGIE :

On a deux types d'enquêtes :

- Les **enquêtes descriptives**
- Les **enquêtes étiologiques**

Les enquêtes descriptives

Objectifs : avoir des données sur la maladie à un instant T. Donc on va soit utiliser **l'incidence** soit la **prévalence**.

Les enquêtes étiologiques

Objectifs : établir des liens entre les maladies et les facteurs de risques (*est-ce que fumer provoque plus de cancer du poumon).*

Donc dans ces enquêtes ont va surtout comparer des groupes de personnes (non-fumeur et fumeur).

ATTENTION : on n'établit pas de relation causale mais une présomption c'est-à-dire que l'on n'est jamais certain à 100% mais que l'on le pense fortement.

Il existe 3 types d'enquêtes étiologiques établies selon la chronologie : **prospective, rétrospective, transversale.**

Prospective (cohorte) :

- Mesure de l'exposition avant l'apparition de la maladie :
- Exposition → maladie ?
- Non exposition → maladie ?

Rétrospective (cas-témoins ou cohorte) :

- Recherche de l'exposition après constat (ou non) de la maladie
- Malades → exposition ?
- Non malades → exposition ?

Transversale :

- Mesure de l'exposition et de la maladie simultanément

Enquête de cohorte (prospective ou rétrospective) ou enquête exposé non-exposé

Dans les enquêtes de cohorte, on va suivre les sujets sur le temps avec des **sujets non-exposés** et des **sujets exposés**. On aura donc à la fin de l'enquête, les **incidences** de la maladies parmi les exposés et les non-exposés.

Présentation des données :

	Malade	Non Malade
Exposés	a	b
Non exposés	c	d

Mesures :

- **Incidence** : risque de survenue de maladie dans chaque groupe

$$I_e = \frac{a}{(a+b)}$$

$$I_{ne} = \frac{c}{(c+d)}$$

- **Différence de risque** entre exposés et non-exposés : $DR = I_e - I_{ne}$
- **Risque relatif (RR)** : rapport entre l'incidence chez les exposés sur l'incidence chez les non-exposés : **RR** = $\frac{I_e}{I_{ne}}$

Le **RR** est un nombre sans unité compris entre 0 et l'infini. Un RR "nul" est égal à 1. Plus le RR est éloigné de 1, plus l'association entre la survenue de la maladie et la présence du facteur étudié est forte.

Interprétation de l'intervalle de confiance du RR à 95%

(L'intervalle sera donné pour ce type d'exercice) :

- S'il contient 1 : pas de relation démontrée entre la maladie et l'exposition
- S'il est > 1 : excès de risque dans le groupe exposé
- S'il est < 1 : moins de risque de contracter la maladie s'il y a exposition (facteur protecteur)

Enquête cas-témoins (rétrospective)

Les études cas-témoins sont des études **rétrospectives** : c'est-à-dire que l'on fait l'enquête une fois que les sujets sont malades. On réalise l'enquête avec des sujets sains (sujet témoins) et des sujets malades et on regarde leur **fréquence d'exposition** à la possible cause de leur maladie. Le but final étant de savoir si l'exposition est à l'origine potentielle de la maladie.

Présentation des données :

	Cas *(malade)*	**Témoins *(non malade)***
Exposés	a	b
Non exposés	c	d

Mesures :

- On a choisi le nombre de témoins : on ne peut pas calculer l'incidence
- On calcule les fréquences d'exposition au facteur :

$$Exp_c = \frac{a}{(a+c)}$$

$$Exp_t = \frac{b}{(b+d)}$$

- On peut comparer les fréquences avec le χ^2 (*voir fiche comparaison de fréquence*).

Les cotes d'exposition :

- La cote d'exposition chez les cas : $\frac{a}{c}$
- La cote d'exposition chez les témoins : $\frac{b}{d}$

Rapport des cotes : **Odd ratio** (**OR**) : permet d'évaluer la liaison entre l'exposition et la maladie.

$$OR = \frac{\left(\frac{a}{c}\right)}{\left(\frac{b}{d}\right)}$$

OR : nombre sans unité entre 0 et l'infini :

- *OR "nul" : 1*
- *Plus OR est éloigné de 1 : plus l'association entre la maladie et la présence du facteur étudié est forte*

Interprétation de l'intervalle de confiance du RR à 95% :

- 1 est compris dans l'IC : pas de relation démontrée entre la maladie et l'exposition
- OR significativement > 1 (borne inf de l'IC >1) : fréquence d'exposition supérieure chez les cas
- OR significativement < 1 (borne sup de l'IC <1) : Exposition : facteur protecteur

RESUME

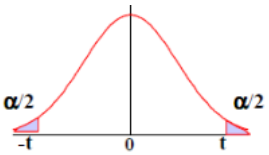
	Enquête Cohorte	Enquête Cas-témoins
Avantages	Bien adaptée pour étudier : • Les risques (incidences) • Des expositions rares • Plusieurs maladies • La séquence exposition-maladie	Bien adaptée pour étudier : • Des maladies rares • Plusieurs facteurs de risque
	Peu de biais de sélection	Etude courte, coût faible
	Peu de biais de mémorisation	Rapidité d'exécution Échantillon de taille modérée
Inconvénients	Non adaptée pour étudier : • Des maladies rares • Plusieurs expositions	Non adaptée pour étudier : • Des expositions rares • Plusieurs maladies • La séquence temporelle exposition-maladie
	Coût élevé	Pas de calcul de taux d'incidence
	Longue période de latence	OR biaisé si prévalence élevée

NOTES PERSONNELLES

TABLES DE LOIS

Table du *t* de Student

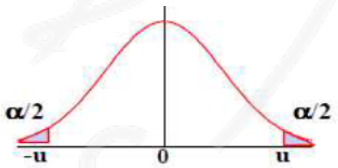
La table donne la probabilité α pour que *t* égale ou dépasse, en valeur absolue, une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



ddl \ α	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,599
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,768
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,659
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,646
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,526	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639	3,416
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Table de l'écart réduit (loi Normale centrée réduite)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée u, c'est à dire la probabilité extérieure à l'intervalle (-u, +u).

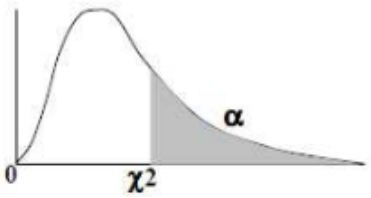


α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,575 829	2,326 348	2,170 090	2,053 749	1,959 964	1,880 794	1,811 911	1,750 686	1,695 398
0,1	1,644 854	1,598 193	1,554 774	1,514 102	1,475 791	1,439 531	1,405 072	1,372 204	1,340 755	1,310 579
0,2	1,281 552	1,253 565	1,226 528	1,200 359	1,174 987	1,150 349	1,126 391	1,103 063	1,080 319	1,058 122
0,3	1,036 433	1,015 222	0,994 458	0,974 114	0,954 165	0,934 589	0,915 365	0,896 473	0,877 896	0,859 617
0,4	0,841 621	0,823 894	0,806 421	0,789 192	0,772 193	0,755 415	0,738 847	0,722 479	0,706 303	0,690 309
0,5	0,674 490	0,658 838	0,643 345	0,628 006	0,612 813	0,597 760	0,582 842	0,568 051	0,553 385	0,538 836
0,6	0,524 401	0,510 073	0,495 850	0,481 727	0,467 699	0,453 762	0,439 913	0,426 148	0,412 463	0,398 855
0,7	0,385 320	0,371 856	0,358 459	0,345 126	0,331 853	0,318 639	0,305 481	0,292 375	0,279 319	0,266 311
0,8	0,253 347	0,240 426	0,227 545	0,214 702	0,201 893	0,189 118	0,176 374	0,163 658	0,150 969	0,138 304
0,9	0,125 661	0,113 039	0,100 434	0,087 845	0,075 270	0,062 707	0,050 154	0,037 608	0,025 069	0,012 533

α	0,002	0,001	0,0 001	0,00 001	0,000 001	0,0 000 001	0,00 000 001	0,000 000 001
u	3,090 23	3,290 53	3,890 59	4,417 17	4,891 64	5,326 72	5,730 73	6,109 41

Table de la loi du χ^2

La table donne la probabilité α pour que χ^2 égale ou dépasse une valeur donnée, en fonction du nombre de degrés de liberté (d.d.l.).



ddl \ α	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,001
1	0,0002	0,0010	0,0039	0,0158	2,71	3,84	5,02	6,63	10,83
2	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21	13,82
3	0,11	0,22	0,35	0,58	6,25	7,81	9,35	11,34	16,27
4	0,30	0,48	0,71	1,06	7,78	9,49	11,14	13,28	18,47
5	0,55	0,83	1,15	1,61	9,24	11,07	12,83	15,09	20,52
6	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81	22,46
7	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48	24,32
8	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09	26,12
9	2,09	2,70	3,33	4,17	14,68	16,92	19,02	21,67	27,88
10	2,56	3,25	3,94	4,87	15,99	18,31	20,48	23,21	29,59
11	3,05	3,82	4,57	5,58	17,28	19,68	21,92	24,72	31,26
12	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22	32,91
13	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69	34,53
14	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14	36,12
15	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58	37,70
16	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00	39,25
17	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41	40,79
18	7,01	8,23	9,39	10,86	25,99	28,87	31,53	34,81	42,31
19	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19	43,82
20	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57	45,31
21	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93	46,80
22	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29	48,27
23	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64	49,73
24	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,98	51,18
25	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31	52,62
26	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64	54,05
27	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96	55,48
28	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,28	56,89
29	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59	58,30
30	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89	59,70

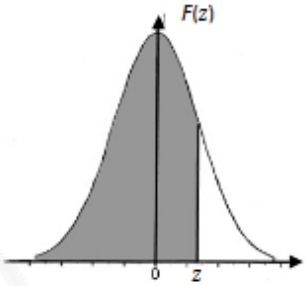
Test de Wilcoxon

N	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
α=0,05	0	2	4	6	8	11	14	17	21	25	30	35	40	46	52	59	66	73	81
α=0,01	-	-	0	2	3	5	7	10	13	16	20	23	28	32	38	43	49	55	61

Table destinée à un test bilatéral

Loi normale : fonction de répartition

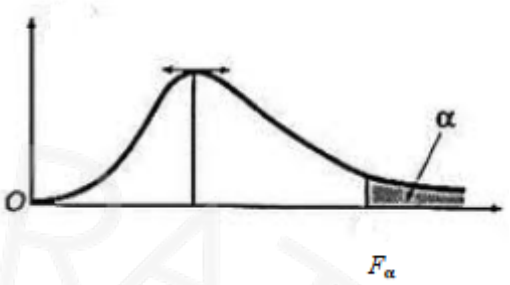
Pour une valeur $z \geq 0$, la table ci-dessous renvoie la valeur de la fonction de répartition $F(z)$ d'une loi normale centrée réduite au point z.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

Table de la loi de Fisher-Snedecor pour $\alpha = 2{,}5\%$

La table donne la limite supérieure F_α de $F = \frac{S_1^2}{S_2^2}$, telle que $P(F \geq F_\alpha) = \alpha$, en fonction des nombres de degrés de liberté ν_1 et ν_2.



ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	∞
1	648	799	864	900	922	937	948	957	963	969	985	993	1001	1008	1013	1016	1017	1018
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.2	14.1	14.0	14.0	13.9	13.9	13.9
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.38	8.32	8.29	8.27	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.23	6.14	6.08	6.05	6.03	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	4.98	4.92	4.88	4.86	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.28	4.21	4.18	4.16	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.81	3.74	3.70	3.68	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.77	3.67	3.56	3.47	3.40	3.37	3.35	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.31	3.22	3.15	3.12	3.09	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.33	3.23	3.12	3.03	2.96	2.92	2.90	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.18	3.07	2.96	2.87	2.80	2.76	2.74	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.05	2.95	2.84	2.74	2.67	2.63	2.61	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	2.95	2.84	2.73	2.64	2.56	2.53	2.50	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.55	2.47	2.44	2.41	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.79	2.68	2.57	2.47	2.40	2.36	2.33	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.72	2.62	2.50	2.41	2.33	2.29	2.26	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.67	2.56	2.44	2.35	2.27	2.23	2.20	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.62	2.51	2.39	2.30	2.22	2.18	2.15	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.25	2.17	2.13	2.10	2.09
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.50	2.39	2.27	2.17	2.09	2.05	2.02	2.00
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.44	2.33	2.21	2.11	2.02	1.98	1.95	1.94
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.39	2.28	2.16	2.05	1.97	1.92	1.90	1.88
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.34	2.23	2.11	2.01	1.92	1.88	1.85	1.83
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	1.97	1.88	1.84	1.81	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.83	1.74	1.69	1.66	1.64
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.11	1.99	1.87	1.75	1.66	1.60	1.57	1.55
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.70	1.60	1.54	1.51	1.48
80	5.22	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.28	2.21	2.00	1.88	1.75	1.63	1.53	1.47	1.43	1.40
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	1.97	1.85	1.71	1.59	1.48	1.42	1.38	1.35
200	5.10	3.76	3.18	2.85	2.63	2.47	2.35	2.26	2.18	2.11	1.90	1.78	1.64	1.51	1.39	1.32	1.27	1.23
500	5.05	3.72	3.14	2.81	2.59	2.43	2.31	2.22	2.14	2.07	1.86	1.74	1.60	1.46	1.34	1.25	1.19	1.14
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.43	1.30	1.21	1.13	1.01

Test de Mann-Whitney

m - n	n																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	-	-	-	0	2	5	8	13	17	23	30	37	45	55	64	75	87	99	113	127
1	-	-	-	1	3	6	10	15	20	26	33	41	50	59	70	81	93	106	119	
2	-	-	0	2	5	8	12	17	23	29	37	45	54	64	75	86	99	112		
3	-	-	1	3	6	10	14	19	26	33	40	49	59	69	80	92	105			
4	-	-	1	4	7	11	16	22	28	36	44	53	63	74	85	98				
5	-	-	2	4	8	13	18	24	31	39	47	57	67	78	90					
6	-	0	2	5	9	14	20	26	34	42	51	61	72	83						
7	-	0	3	6	11	16	22	29	37	45	55	65	76							
8	-	0	3	7	12	17	24	31	39	48	58	69								
9	-	0	4	8	13	19	26	34	42	52	62									
10	-	1	4	9	14	21	28	36	45	55										
11	-	1	5	10	15	22	30	38	48											
12	-	1	5	11	17	24	32	41	50											
13	-	1	6	11	18	25	34	43												
14	-	1	6	12	19	27	36	45												
15	-	2	7	13	20	29	38													
16	-	2	7	14	22	30	40													
17	-	2	8	15	23	32														
18	-	2	8	16	24	33														
19	-	3	9	17	25															
20	-	3	9	17	27															
21	-	3	10	18																
22	-	3	10	19																
23	-	3	11																	
24	-	4	11																	
25	-	4																		
26	-	4																		

Seuil $\alpha =$ 0,05

m - n	n																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	-	-	-	-	0	2	4	7	11	16	21	27	34	42	51	60	70	81	93	105
1	-	-	-	-	1	3	6	9	13	18	24	31	38	46	55	65	75	87	99	
2	-	-	-	0	1	4	7	11	16	21	27	34	42	50	60	70	81	92		
3	-	-	-	0	2	5	9	13	18	24	30	37	45	54	64	74	86			
4	-	-	-	1	3	6	10	15	20	26	33	41	49	58	68	79				
5	-	-	-	1	4	7	12	17	22	29	36	44	53	63	73					
6	-	-	0	2	5	9	13	18	24	31	39	47	57	67						
7	-	-	0	2	6	10	15	20	27	34	42	51	60							
8	-	-	0	3	7	11	16	22	29	37	45	54								
9	-	-	1	3	7	12	18	24	31	39	48									
10	-	-	1	4	8	13	19	26	33	42										
11	-	-	1	5	9	15	21	28	36											
12	-	-	2	5	10	16	22	30	38											
13	-	-	2	6	11	17	24	32												
14	-	-	2	6	12	18	25	34												
15	-	-	2	7	13	19	27													
16	-	-	3	8	14	21	29													
17	-	0	3	8	14	22														
18	-	0	3	9	15	23														
19	-	0	4	9	16															
20	-	0	4	10	17															
21	-	0	4	10																
22	-	0	5	11																
23	-	0	5																	
24	-	0	5																	
25	-	1																		
26	-	1																		

Seuil $\alpha =$ 0,01

REMERCIEMENTS

Jules ONGHENA
Tuteur 2022-24

Brieuc BOUFFORT
Tuteur 2022-24

Emelyne ROBERT
VP Design 2023

Lola LE GUELL
VP Site 2023-24

Pierre GELGON
Pilote 2023-24

Merci au professeur Eric Hitti pour sa relecture de ce livret de Biostatistiques.

Très heureuse d'avoir pu à nouveau participer à l'amélioration de ce livret et de faire partie de cette nouvelle équipe particulièrement motivée et sympathique ! Bon courage à vous les pious, la team UE4 est là pour vous aider tout le long du semestre ! N'oubliez pas de faire des QCMs de Biostats régulièrement 😉

Léa JEGOREL
Tutrice 2023-25

Merci à toutes les personnes ayant pu permettre la réalisation de ce livret, et à ceux qui, dans le futur, l'entretiendront et le feront vivre. Et pour toi, cher PASS arrivé jusqu'ici, **BON COURAGE** pour cette année si *riche* !

Estelle RENARD
Tutrice 2023-25

Une seule chose à dire : merci ! Merci à vous, si vous êtes en PASS, d'avoir pris le temps de vous aider de ce livret durant cette année si compliquée. Merci à vous, mes anciens collègues d'avoir eu l'idée de ce livret tellement utile pour cette matière ; avoir eu la possibilité de le perfectionner pendant 2 ans a été un grand plaisir pour moi. Enfin, merci à vous, futurs tuteurs de l'UE4, de continuer à l'entretenir et à le mettre à jour ! En espérant que toi qui lit quelques lignes soit peut être l'un des prochains tuteurs... 😊 Biostatistiquement,

Julien ROUGER—LE PIERRES

Tuteur 2024-2026



Sélène BASTIDE
Tutrice 2025-2026

Hubert GUERIN
Tuteur 2024-2026

Sonny BELAN
Tuteur 2025-2026