République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 8 Mai 1945 – Guelma

Faculté de Mathématiques et de l'Informatique et Sciences de la Matière

Département de Mathématiques

Polycopié de cours :

Biostatistiques

3ème Année Licence LMD, Immunologie

Dr. MENACEUR Amor

Guelma 2017

Table des matières

Chapitre 1

Statistiques descriptives à 1 ou 2 variables

1.1 Statistique

Le terme statistique désigne à la fois un ensemble de données d'observations, et l'activité qui consiste en leur recueil, leur traitement et leur interprétation. Les termes statistiques, ou statistiques (au pluriel) englobent ainsi plusieurs notions distinctes :

a. D'une part le recensement de grandeurs d'intérêt comme le nombre d'habitants d'un pays, le revenu moyen par habitant, le nombre de séropositifs dans la population Algérienne. Nous voyons que la notion fondamentale qui se dégage de cette énumération est celle de population. Une population est un ensemble d'objets, d'êtres vivants ou d'objets abstraits (ensemble des mains de 5 cartes distribuées au bridge...) de même nature.

b. La statistique en tant que science s'intéresse aux propriétés des populations naturelles. Plus précisément elle traite de nombres obtenus en comptant ou en mesurant les propriétés d'une population. Cette population d'objets doit en outre être soumise à une variabilité, qui est due à de très nombreux facteurs inconnus (pour les populations d'objets biologiques qui nous intéressent ces facteurs sont les facteurs génétiques et les facteurs environnementaux).

c. A ces deux acceptions du terme statistiques (au pluriel) il faut ajouter le terme statistique (au singulier) qui définit toute grandeur calculée à partir d'observations. Ce peut être la plus grande valeur de la série statistique

d'intérêt, la différence entre la plus grande et la plus petite, la valeur de la moyenne arithmétique de ces valeurs, etc.

**Les statistiques descriptives visent à représenter des données dont on veut connaître les principales caractéristiques quantifiant leur variabilité.*

1.2 Population et échantillon

On appelle population P un ensemble généralement très grand, voire infini, d'individus ou d'objets de même nature. Tous les médecins d'Algérie constituent une population, de même que l'ensemble des résultats possibles du tirage du loto. Une population peut donc être réelle ou fictive. Il est le plus souvent impossible, ou trop coûteux, d'étudier l'ensemble des individus constituant une population ; on travaille alors sur une partie de la population que l'on appelle échantillon. Pour qu'un échantillon permette l'étude de la variabilité des caractéristiques d'intérêt de la population, il faut qu'il soit convenablement sélectionné. On parlera d'échantillon représentatif si les individus le constituant ont été tirés au sorti dans la population. Si par exemple on souhaite déterminer les caractéristiques « moyennes » du poids et de la taille des prématurés masculins on tirera au hasard un certain nombre de sujets parmi les naissances de prématurés de l'année.

Chaque individu, ou unité statistique, appartenant à une population est décrit par un ensemble de caractéristiques appelées variables ou caractères. Ces variables peuvent être quantitatives (numériques) ou qualitatives (non numériques) :

Quantitatives : pouvant être classées en variables continues (taille, poids) ou discrètes (nombre d'enfants dans une famille).

Qualitatives : pouvant être classées en variables catégorielles (couleurs des yeux) ou ordinales (intensité d'une douleur classée en nulle, faible, moyenne, importante).

Le but d'une étude statistique est généralement de déterminer certaines caractéristiques moyennes d'une population qu'on appelle aussi un univers. Les éléments de cette population peuvent être des individus, des objets réels, ou des éléments abstraits.

Exemples

1. On souhaite déterminer l'âge moyen des habitants d'une ville.

2. On s'intéresse à la consommation moyenne (par Km) de la "population" des voitures qui circulent dans un pays.

1.3 Séries statistiques à une variable

1.3.1 Définitions-Tableaux statistiques

Un ensemble fini Ω est dit population. Les éléments de Ω sont appelés individus. Une application de Ω dans $\mathbb{R}$ est dit caractère. Le caractère détermine une partition de Ω suivant ses modalités.

Il est souvent difficile, voire impossible, d'observer toutes les données. On étudiera alors une partie de la population qu'on appelle échantillon. Une variable X peut-être discrète ou continue.

♦ Variables discrètes

a) Tableau

Soit n_i l'effectif de la valeur x_i de la variable X.

On a $\sum_{i=1}^{p} n_i = n$ et $f_i = \frac{n_i}{n}$ la fréquence correspondante.

Un tableau statistique est présenté sous la forme :

x_i	x_1 x_2 .....x_p	total
n_i	n_1 n_2 .....n_p	n
f_i	f_1 f_2 .....f_p	1

b) Représentation graphique

Diagramme en batons : On porte f_i (ou n_i) en ordonnée en fonctionde x_i

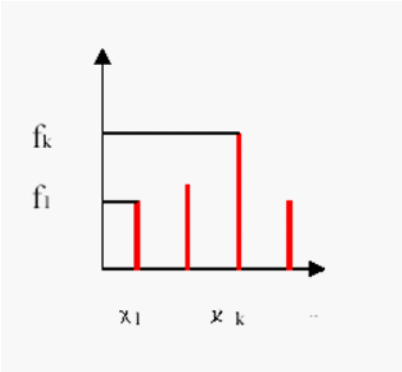


Diagramme en batons

♦ **Variables continues - Données groupées**

a) Tableau

Les valeurs de la variable X sont regroupées en classes $[x_i, x_{i+1}[$, $(i = 1, 2, ..., p)$. Une centre de classe c_i est choisi pour la classe i (moyenne arithmétique des deux extrémités). L'effectif et la fréquence de la classe i sont n_i et f_i.

On note par $F_k = \sum_{i=1}^{k} f_i$ les fréquences cumulées.

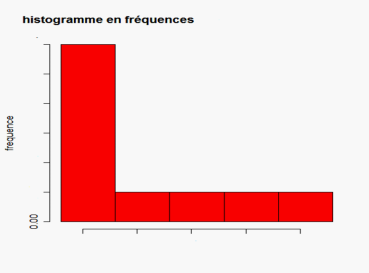
On représente ces données sous la forme du tableau

classes	centre de classe	n_i	f_i	F_i
$[x_1, x_2[$	c_1	n_1	f_1	F_1
$[x_2, x_3[$	c_2	n_2	f_2	F_2
...	...	...	...	...
$[x_p, x_{p+1}[$	c_p	n_p	f_p	F_p

L'amplitude de la classe i est $a_i = x_{i+1} - x_i$. Les classes ne sont pas toujours d'égale amplitude.

b) Représentation graphique

Histogramme : Le rectangle pour chaque classe a pour longueur l'axe des abscisses, l'amplitude de cette classe et une surface proportionnelle à la fréquence de la classe.



1.3.2 Paramètre de position et valeurs centrales

Le but des valeurs centrales est de résumer en une seule valeur l'ensemble des valeurs d'une distribution statistique. Il existe quatre valeurs de positions :

1- Le mode (Mo),

2- La moyenne (X ou m)
3- La médiane ou le médian (M_e ou Md)
4- Les fractiles (Quantiles) (Q_n)
Parmi ces valeurs les trois premières sont des valeurs de position centrales :

a) Le mode

Le mode : est la valeur la plus fréquente d'une distribution. Cette valeur se calcule toujours à partir d'un dénombrement des modalités du caractère. Il faut donc distinguer le cas des caractères discrets et des caractères continus.

- Caractère qualitatif et caractère discret : Pour un caractère qualitatif, ou pour un caractère quantitatif discret ayant un nombre de modalités inférieur au nombre d'éléments, le mode est la modalité ou la valeur qui a la fréquence simple la plus élevée (ou l'effectif le plus élevé, ce qui revient au même).

- Caractère quantitatif continu : Les modalités étant en nombre infini, il est peu probable que deux éléments aient la même valeur. Dans ce cas, le mode ne peut pas être défini directement, il faut au préalable établir une partition en classes. Le mode est alors le centre de la classe modale, c'est à dire de la classe qui a la fréquence moyenne la plus élevée.

Le mode correspond à la valeur lue en abscisse du sommet de l'histogramme. Lorsque celui-ci présente deux pics séparés par un creux, on dit que la distribution est bimodale.

Application : Cas de calcul des modes :

- Cas 1 : Données rangées : le mode est la valeur de la donnée qui apparaît le plus fréquemment (celle qui a le plus d'occurrences) :

$$140; 141; 144; 144; 148; 148; 152; 152;$$
$$152; 154; 155; 158; 158; 161; 170; 172$$

Le mode est 152 car il possède le plus grand nombre d'occurrences (il est référencé 3 fois).

- Cas 2 : Données condensées : le mode est la valeur de la donnée qui possède la fréquence la plus élevée (relative ou absolue).

Modalités x_i (age en années)	14	**16**	18	21	22	24	25	*total*
Fréquences absolues	5	**12**	10	8	11	7	3	56
Fréquences relatives	0.089	**0.214**	0.179	0.143	0.196	0.125	0.054	1

Dans cette série statistique, le mode est égal à $Mo = 16$ ans

- Cas 3 : Données groupées en classes : la classe modale est la classe ayant la plus haute fréquence (relative ou absolue). Il est possible de calculer de façon plus précise le mode en appliquant la formule suivante :

$$M_o = a_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} L$$

où

Δ_1 : différence entre l'effectif de la classe modale et l'effectif de la classe précédente.

Δ_2 : différence entre l'effectif de la classe modale et l'effectif de la classe qui suit.

a_i : Borne inférieur de la classe modale

L : largeur de la classe modale

b) La moyenne arithmétique

Formalisation mathématique de la moyenne arithmétique, noté $\bar{X}$ ou m, est la mesure la plus commune de tendance centrale, elle se définit comme la somme des scores divisée par le nombre de scores. Par exemple, en biologie la moyenne peut être résumée par la somme des observations divisée par l'effectif de l'échantillon étudié :

$$m = \frac{1}{n} \sum_{i=1}^{n} x_i$$

On effectue une moyenne pondérée en assimilant chaque classe j à son centre x_j et en pondérant par l'effectif n_j de la classe.

$$m = \frac{1}{n} \sum_{j=1}^{p} n_j x_j$$

Exemples

1-Soit les valeurs de quatre notes : $10, 12, 13$ et 16, la moyenne arithmétique est :

$$(11 + 12 + 13 + 16)/4 = 13$$

2-Soit la série statistique suivante :

valeurs	0	1	2	3	4
effectifs	1	2	1	4	2

$$m = \frac{0 + 2 \times 1 + 2 + 4 \times 3 + 2 \times 4}{1 + 2 + 1 + 4 + 2} = 2.4$$

Remarque 1 Si les données ont été regroupées en classes, on ne peut calculer la valeur exacte de la moyenne. On peut toutefois en déterminer une bonne approximation en remplaçant chaque classe par son milieu.

Dans les séries statistiques suivantes déterminer les moyennes

a) Tableau de fréquences

$valeurs$	12	13	14	15	16
$fréquences$	0.05	0.17	0.43	0.30	0.05

b) Données réparties en classes

Classes	$[0; 5[$	$[5; 10[$	$[10; 15[$	$[15; 20]$
Effectifs	7	12	14	2

Autres indicateurs de moyenne :

Il existe des indicateurs de la moyenne autre que la moyenne arithmétique. Néanmoins, ils sont moins utilisés en biostatistique car ils ne présentent d'intérêt que dans des cas très particuliers. Ils ne feront pas l'objet de ces modules : la moyenne géométrique, la moyenne harmonique, la moyenne quadratique, la moyenne arithmético-géométrique.

c) La médiane et la classe médiane

Définition générale

On appelle médiane la valeur "du milieu". On dit qu'elle partage la série statistique en deux moitiés : il y a autant de valeurs en dessous qu'au dessus. (C'est la donnée qui permet de diviser une série ordonnée d'une façon croissante en 2 parties égales $(50\%, 50\%)$. La médiane ne peut être calculée que pour les caractères quantitatifs.

-Médiane, pour les données rangées

Les valeurs du caractère X étant classées par ordre croissant, la médiane est la valeur du caractère qui partage l'ensemble décrit par X en deux sous ensembles d'effectifs égaux : 50% des éléments ont des valeurs de X supérieures à $M_é$ et 50%prennent des valeurs inférieures.

Méthode : Soit une série statistique d'effectif total n, rangée par ordre croissant $(x_1, x_2, ..., x_n)$. Pour déterminer son rang, il y a 2 cas :

-Si $n = 2p$ est pair, $M_é$ est le centre de l'intervalle $[x_p, x_{p+1}]$.

-Si $n = 2p + 1$ est impaire, $M_é$ est le nombre x_{p+1}.

Exemple 1

Cas de données discrètes "en vrac" $10, 7, 12, 18, 16, 15, 5, 11, 11, 20, 15, 11, 18, 14$

Ordonnons la série par ordre croissant : $5, 7, 10, 11, 11, 11, 12, 14, 15, 15, 16, 18, 18, 20$

$n = 14$ est pair, $M_é$ est le centre de l'intervalle $[12, 14]$ (La médiane est donc la demi somme des 7ème et 8ème termes) alors $M_é = 13$.

-Médiane, pour les données condensées

La définition est la même, elle correspond dans ce cas à la première modalité ou valeur dont la fréquence relative cumulée dépasse 0.5 ou l'effectif cumulé dépasse les 50%.

Il faut calculer les fréquences ou les effectifs cumulés dès que celle-ci atteint respectivement 0.5 ou 50% il suffit de choisir le nombre à mi chemin entre la modalité ou valeur concernée et la suivante.

Cas d'un tableau d'effectifs

On ordonne le tableau, et on cherche l'élément qui partage la distribution en deux parties égales : on repère l'élément qui a le rang $(N + 1)/2$ pour le caractère X. Si la distribution a un nombre impair d'éléments on trouve une valeur unique qui est la médiane, si la distribution a un nombre pair d'éléments, on trouve deux valeurs qui déterminent un intervalle médian : on prend alors pour médiane le centre de cet intervalle médian.

valeurs	effectifs	effectifs cumulées
1	6	6
2	11	17
3	25	**42**
4	19	61
5	15	76
6	5	81

L'effectif total est de 81 or la valeur de rang $\frac{81+1}{2} = 41$

La médiane est donc le $41^{ème}$ terme : médiane $= 3$

Médiane d'une série statistique continue

Si les données ont été regroupées en classes, on ne peut déterminer la valeur exacte de la médiane. En revanche, on appellera *classe médiane*, la classe qui la contient (et permet donc d'en donner un encadrement).

La classe médiane est la première classe où la fréquence cumulée est supérieure à $0,50$

Exemple 2

valeurs	effectifs	effectifs cumulées	fréquence(f_i)	fréquence cumulée(f_i^c)
[0.5[	10	10	0.2	0.2
[5.8[	8	18	0.16	0.36
[8.12[	12	**30**	0.24	0.6
[12.15[	11	41	0.22	0.82
[15.20[	9	50	0.18	1
/	50	/	1	/

Utilisons la colonne des effectifs cumulés pour déterminer la médiane, il ya 50 notes 50% de l'effectif cumulée 25, la médiane se trouve donc dans l'intervalle [8.12[

Pour préciser la valeur de la médiane, il faut supposer que toutes les données sont réparties uniformément (c'est-à-dire que les données sont réparties sur un continuum). On repère la classe qui contient la médiane, puis on réalise une interpolation linéaire pour estimer la valeur de celle-ci selon la formule suivante :

$$M_{\acute{e}} = a_i + \frac{(0.5 - f_{i-1}^c)}{f_i} L$$

a_i : Borne inférieure de la classe médiane

f_{i-1}^c : Fréquence relative cumulée de la classe qui précède la classe médiane.

f_i : Fréquence relative de la classe médiane.

L : largeur, amplitude des classes.

Application pour l'exemple précédent :

$$M_{\acute{e}} = 8 + \frac{(0.5 - 0.36)}{0.24} 4 = 10.333$$

Remarque 2 Autre méthode de calcul de la médiane : il est aussi possible de déterminer la médiane à l'aide des polygones des effectifs cumulés.

d) Quantiles

Il a été vu précédemment que la médiane partage la distribution des fréquences en 2 parties égales. Il est possible de partager une distribution de fréquence en 4 parties égales (quartiles), en 10 parties égales (déciles), en 100 parties égales (centiles), en n parties égales....

Définition 1 on appelle quantiles les valeurs du caractère qui définissent les bornes d'une partition en classes d'effectifs égaux.

1-Les quartiles : sont les trois valeurs qui permettent de découper la distribution en quatre classes d'effectifs égaux. On les notes Q_1, Q_2 et Q_3.

- Q_1 : quartile inférieur, 25% des valeurs de la variable lui sont inférieures et 75% lui sont supérieures

- Q_2 : médiane, 50% des valeurs de la variable lui sont inférieures et 50% lui sont supérieures

- Q_3 : quartile supérieur, 75% des valeurs de la variable lui sont inférieures et 25% lui sont supérieures.

Remarque 3 Q_2 est égal à la médiane.

Les déciles sont les 9 valeurs de X qui permettent de découper la distribution en dix classes d'effectifs égaux. 0n les note $X_{d_1}...X_{d_9}$.

2-Détermination des valeurs de la variable à partir d'un rang centile données.

a) Cas des données rangées :

C_k : rang du centile : Il correspond à la donnée dont le rang est l'entier qui suit : $\frac{Nk}{100}$ si $\frac{Nk}{100}$ n'est pas un entier. Dans le cas contraire si $\frac{Nk}{100}$ est un entier, C_k correspond à la données dont la position (le rang) est a mis chemin entre le rang donnée par : $\frac{Nk}{100}$ et la position suivante :

N : nombre total de valeurs dans la série statistique

k : le rang du centile

b) Cas des données groupées en classes :

La classe contenant C_k correspond à la première classe où la fréquence cumulée atteint ou dépasse $\frac{\alpha}{100}$, par référence à la formule du calcul de la médiane (vue précédemment) il est possible d'écrire la formule suivante de C_k

$$C_k = a_i + \frac{\left(\frac{\alpha}{100} - f_{i-1}^c\right)}{f_i} L_i$$

où : a_i : Borne inférieure de la classe contenant C_k

f_{i-1}^c : Fréquence relative cumulé de la classe qui précède la classe contenant C_k

f_i : Fréquence relative de la classe contenant C_k.

L_i : largeur, amplitude de la classe contenant C_k

Exemple 3 Soit la série statistique suivante :

$$58; 59; 64; 64; 64; 68; 71; 71;$$
$$79; 82; 82; 85; 92; 92; 92; 95$$

- Trouver les centiles suivants : C_{15}
- Trouver les quartiles : Q_2 et Q_3

◦ Pour centile C_{15} : $k = 15$, le rang de la donnée est déterminé par la formule

$$\frac{Nk}{100} = \frac{16 \times 15}{100} = 2.4$$

La valeur n'est pas un entier, le rang est donc le premier entier suivant $2,4$ ainsi C_{15} correspond au rang 3, ce dernier correspond à la valeur : 64

◦ Pour centile C_{50} ou quartile Q_2 ou la médiane : $k = 50$ le rang de la donnée est déterminé par la formule

$$\frac{Nk}{100} = \frac{16 \times 50}{100} = 8$$

La valeur est un entier, C_k correspond à la données dont la position (le rang) est a mis chemin entre le rang 8 et le rang 9, ainsi Q_2 correspond à la moyenne des valeurs du au rang 8 (qui correspond à la valeur 71) et le rang 9 (qui correspond à la valeur 79) :

$$Q_2 = \frac{71 + 79}{2} = 75$$

◦ Pour centile C_{75} ou quartile Q_3 ou la médiane : $k = 75$ le rang de la donnée est déterminé par la formule

$$\frac{Nk}{100} = \frac{16 \times 75}{100} = 12$$

La valeur est un entier, C_k correspond à la données dont la position (le rang) est a mis chemin entre le rang 12 et le rang 13, ainsi Q_3 correspond à la moyenne des valeurs du au rang 12 (qui correspond à la valeur 85) et le rang 13 (qui correspond à la valeur 92) :

$$Q_3 = \frac{85 + 92}{2} = 88.5$$

1.3.3 Paramètre de dispersion

Dispersion statistique : On appelle dispersion statistique, la tendance qu'ont les valeurs de la distribution d'un caractère à s'étaler, à se disperser, de part et d'autre d'une valeur centrale. On distingue la dispersion absolue (mesurée dans l'unité de mesure du caractère) et la dispersion relative (mesurée par un nombre sans dimension).

a) L'étendue de la variation

L'étendue d'une distribution est égale à la différence entre la plus grande et la plus petite valeur de la distribution :

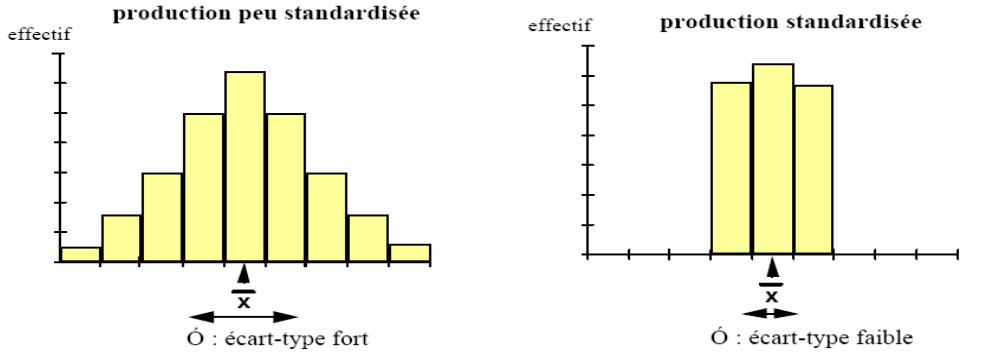
$$\text{Etendue de } X = X_{max} - X_{min}$$

plus l'étendu est grande plus les valeurs sont dispersées.

b) Variance et écart-type :

La variance et écart-type servent à évaluer la dispersion d'une distribution autour d'une valeur centrale, la moyenne. Soit deux séries de microscopes produits dans deux usines différentes. Nous désirons juger de la standardisation de chacune des deux séries. Je choisis de comparer le poids maximal de chaque microscope.

- Si les écarts à la moyenne sont faibles la production est standardisée.
- Si les écarts à la moyenne sont élevés, la production est peu standardisée.



a - Variance : La variance, notée S^2 est la moyenne du carré des écarts à la moyenne.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

-La variance n'est pas un paramètre de dispersion absolue mais plutôt une mesure globale de la variation d'un caractère de part et d'autre de la moyenne arithmétique (quantité d'information). Pour obtenir un paramètre de dispersion absolue, on effectue la racine carrée de la variance S^2, appelé écart-type et que l'on note S.

-La variance pour des données rangées ou groupées en classe devient :

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - m\right)^2$$

où n_i désigne les effectifs de chaque donnée ou de chaque classe.

b - Ecart-type : L'écart type, noté S est la racine carré de la moyenne du carré des écarts à la moyenne, c'est à dire la racine carrée de la variance.

c- Simplification des écritures des variances

La formule de la variance peut être remplacée par une formule plus facile à utiliser (formule pratique de calcul) à savoir :

$$\begin{aligned} S^2 &= \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - m\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - m^2 \end{aligned}$$

1.3.4 Coefficient d'asymétrie

Le coefficient d'asymétrie renseigne sur l'asymétrie et éventuellement la dérive par rapport à une valeur centrale choisie. La distribution d'une variable est symétrique si les observations sont également dispersées de part et d'autre d'une valeur centrale. Ainsi, dans le cas de distributions symétriques, moyenne et médiane sont confondues, sinon elles sont distinctes.

Ce coefficient mesure l'asymétrie d'une distribution, il renseigne sur une asymétrie négative (dissymétrie à gauche), ou une asymétrie positive (dissymétrie à droite), c'est-à-dire il précise si la répartition "penche" d'un côté ou de l'autre. Selon la valeur centrale choisie (mode, médiane ou moyenne arithmétique), il existe différentes manières de caractériser et de mesurer une dissymétrie.

a) Les coefficients d'asymétrie de Yule

Le coefficient de Yule est basé sur les écarts de quartiles, tel que :

$$Y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)}$$

Si : $Y = 0 \Rightarrow$ symétrie parfaite

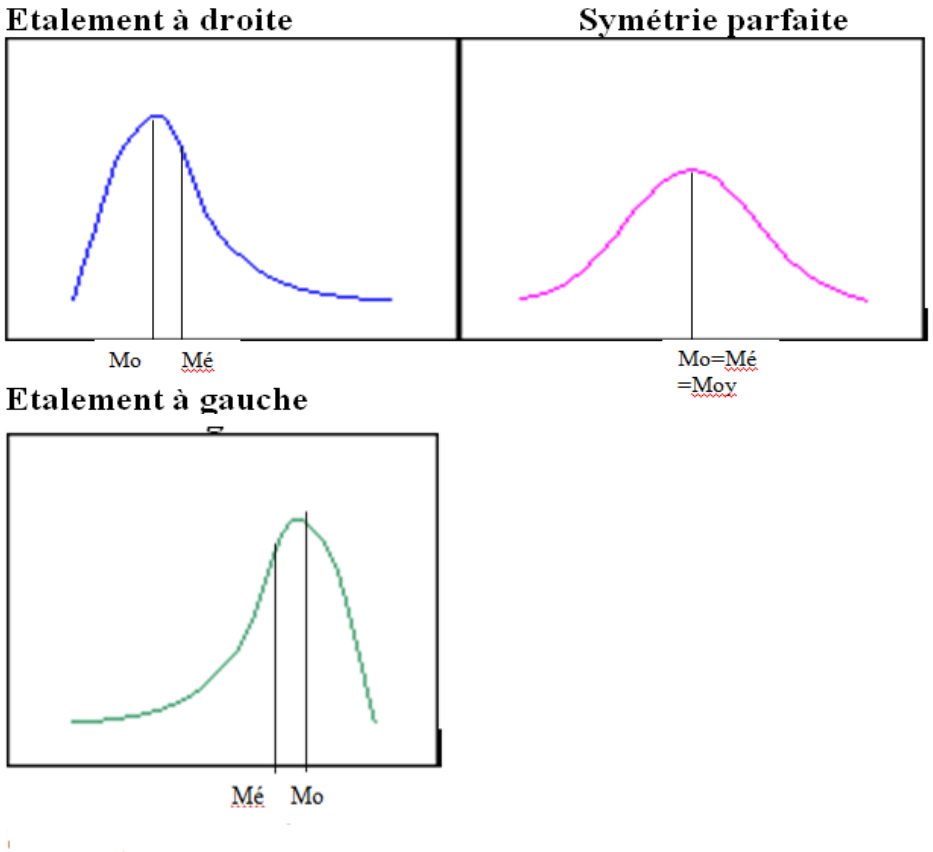
$Y > 0 \Rightarrow$ la courbe de fréquence étalée à gouche.
$Y < 0 \Rightarrow$ la courbe de fréquence étalée à droite.

b) Les coefficients d'asymétrie de Pearson

Analyse la position de deux valeurs centrales (le mode et la moyenne arithmétique) relativisée par la dispersion de la série :

$$P = \frac{m - M_o}{S}$$

Si : $p = 0 \Rightarrow$symétrie parfaite.
$p > 0 \Rightarrow$oblique à gauche (ou étalement à droite)= dissymétrie à droite.
$p < 0 \Rightarrow$oblique à droite (ou étalement à gauche)= dissymétrie à gauche.



c) Les coefficients d'asymétrie de Fisher

Ce coefficient a été défini par Fisher comme suit :

$$\gamma = \frac{m_3(x)}{[S]^3}$$

où

$$m_l(x) = \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - m\right)^l$$

on a

$\gamma = 0$: la distridution est symétrique.

$\gamma < 0$: la distribution est symetrique à droite.

$\gamma > 0$: la distribution est symetrique à gouche.

Exemple 4

classes	n_i	x_i	$n_i x_i$	$n_i x_i^2$	$n_i x_i^3$
$50-60$	8	55	440	24200	1331000
$60-70$	10	65	650	42250	2746250
$70-80$	16	75	1200	90000	6750000
$80-90$	14	85	1190	101150	8597750
$90-100$	10	95	950	90250	8573750
$100-110$	5	105	525	55125	5788125
$110-120$	2	115	230	26450	3041750
total	65	/	5185	429425	36828625

on trouve

$$M_0 = 75; M_{\acute{e}} = Q_2 = 79.1; Q_1 = 68.2; Q_3 = 90.7 \text{ et } m = 79.8$$

D'ou

$$Y = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_2) + (Q_2 - Q_1)} = 0.03,$$

$$P = \frac{\bar{X} - M_o}{S} = \frac{79.8 - 75}{15.44} = 0.31$$

et

$$\gamma = \frac{m_3(x)}{[S]^3} = \frac{1337.31}{(15.44)^3} = 0.36$$

La distribution est donc légèrement oblique à gouche.

1.4 Séries statistiques à 2 variables

L'objectif de cette étude statistique est d'étudier sur une même population de N individus, deux caractères différents (ou modalités différentes) et de rechercher s'il existe un lien ou corrélation entre ces deux variables.

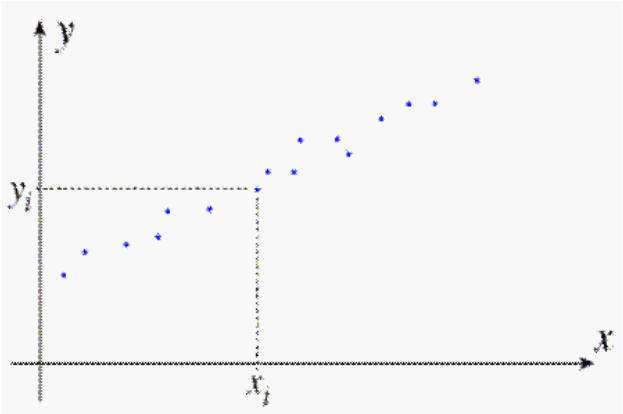
Exemple de relations possibles entre les variables suivantes : taille et âge ; diabète et poids ; taux de cholestérol et régime alimentaire ; niche écologique et population ; ensoleillement et croissance végétale ; toxine et réaction métabolique ; survie et pollution ; effets et doses ; organe 1 et 2 ; organe et fonction biologique ; ...

Tableaux statistiques

Lorsqu'il n'y a qu'une observation pour un couple (x_i, y_i) on décrit la série statistique par le tableau

X	x_1	x_2	...	x_i	...	x_n
Y	y_1	y_2	...	y_i	...	y_n

Le couple (x_i, y_i); $i = 1, ..., n$; représente la valeur prise par (X, Y) dans la i^{eme} observation. On représente la distribution sous forme d'un nuage de points dans $\mathbb{R}^2$



Dans certaines distributions statistiques bidimensionnelles il est possible de calculer les moyennes, les variances et les écart-types marginaux.

Pour les moyennes

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ et } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Pour les variances

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

1.4.1 Covariance

Une première approche entre de la relation éventuelle des valeurs d'une variable X avec des valeurs d'une variable Y est donnée par le calcul de la covariance

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

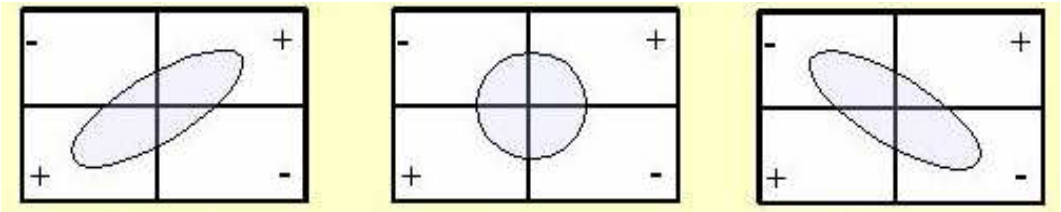
Dans cette formule la « co-variance » apparaît bien comme une combinaison de la variance de X et celle de Y.

Par analogie aux formules précédentes les formules pratiques de calculs de la covariance peuvent aussi s'écrire :

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

Propriétés de la covariance

- $Cov(x,x) = S_x^2$
- $Cov(x,y) \leq S_x S_y$
- Le signe de la Cov est un indicateur de la tendance de la relation sens positif ou négatif (direction d'étirement du nuage de point). Une covariance positive indique une tendance « croissante » des valeurs de Y en fonction de X, une covariance négative une tendance « décroissante »



$cov(x,y) > 0$ $cov(x,y) \approx 0$ $cov(x,y) < 0$

1.4.2 Coefficient de corrélation

La covariance n'est pas un indicateur indépendant de l'ordre de grandeur des variables impliquées (de l'unité employée, par exemple). Le coefficient de corrélation, noté r, permet de résoudre cette difficulté. Ce coefficient pour le couple (X, Y) s'écrit selon la formule suivante :

$$r = \frac{cov(x, y)}{S_x S_y}$$

où S_x et S_y désignent respectivement l'écart-type de la série statistique X et celui de la séries statistique Y.

Propriété de r

- Pour $r = 1$, la droite de la pente est croissante.
- Si $0 < r < 1$, la corrélation est positive, X et Y varient dans le même sens.
- Si $-1 < r < 0$, la corrélation est négative, X et Y varient dans le sens contraire.
- Si $r = -1$, la droite de la pente est décroissante.
- Quand $(r = 0)$, aucune tendance ne peut être déterminée.

1.4.3 Droite de régression linéaire

Une droite de régression linéaire s'écrit selon l'équation :

$$y = ax + b$$

Cette approche de corrélation repose sur l'hypothèse que la relation entre deux variables est de nature linéaire. En faite, il est possible de soupçonner une relation différente entre ces variables :

- courbe de puissance
- courbe exponentielle
- courbe logarithmique,...etc...

Cependant, il existe de nombreuses méthodes permettant de « linéariser » un grand nombre de ces courbes. Ainsi, on se retrouve souvent dans des situations où il est alors possible de tester l'existence d'une relation linéaire entre les variables auxiliaires.

En partant de l'équation $y = ax+b$, a et b doivent être choisis convenablement de sorte que la droite passe au plus proche (ou par le plus possible) des

points expérimentaux. Pour ce faire, on utilise la méthode des moindres carrés : On cherche les coefficients a et b de la droite qui minimise la somme des carrés des distances entre les points expérimentaux et la droite de régression (les points théoriques).

- le coefficient a se détermine comme suit :

$$a = \frac{cov(x,y)}{S_x^2}$$

- le coefficient b se détermine comme suit :

$$\bar{y} = a\bar{x} + b$$

Remarques

1-La droit de régréssion $D^{'}$ de x en y a pour $x = a^{'}y + b^{'}$ où

$$\begin{cases} a^{'} = \frac{cov(x,y)}{S_y^2} \\ \bar{x} = a^{'}\bar{y} + b^{'} \end{cases}$$

2-Si $r = 1$, la droite de régréssion passe par touts les points du nuage.

3-$r^2 = aa^{'}$ tel que $(D^{'})$ est la droite de régréssion de y en x et $(D^{'})$ est la droite de régréssion de x en y, avec (D) :$y = ax + b$ et $(D^{'})$: $x = a^{'}y + b^{'}$.

Exemple 5 La tableau ci-dessous donne la consommation en milliers de calories de douze familles en moyenne par jour. Chaque homme adute est compté pour une "unité de consommation" ; un enfant est compté pour une part d'unité, dépendant de son age et de son sexe.

n^0 de famille	unité de consommation x_i	calories par jour y_i
1	5.3	13
2	7.2	18
3	5.6	9.4
4	7.1	15.4
5	5	7.8
6	3.3	9.3
7	5.2	10.1
8	4.5	7.1
9	4	8.9
10	2	4.4
11	5.7	12.1
12	4.7	11.5
Total	59.6	127

Calculons le coefficient de corrélation linéaire. A l'iade de ce tableau, on peut effectuer les calcule suivants :

n^0 de famille	unité de consommation x_i	calories par jour y_i	x_i^2	y_i^2	$x_i y_i$
1	5.3	13	28.09	169	68.9
2	7.2	18	51.84	324	129.6
3	5.6	9.4	31.36	88.36	52.64
4	7.1	15.4	50.41	237.16	109.34
5	5	7.8	25	60.84	39
6	3.3	9.3	10.89	86.49	30.69
7	5.2	10.1	27.04	102.01	52.52
8	4.5	7.1	20.25	50.41	31.95
9	4	8.9	16	79.21	35.6
10	2	4.4	4	19.36	8.8
11	5.7	12.1	32.49	146.41	68.97
12	4.7	11.5	22.09	132.25	54.05
Total	59.6	127	319.46	1495.5	682.06

$\bar{x} = \frac{59.6}{12} = 4.97 \approx 5$ unité de consommation
$\bar{y} = \frac{127}{12} = 10.583 \Rightarrow \bar{y} \approx 10.6 \times 10^3$ calories.

$$\begin{aligned} S_x^2 &= \frac{1}{12}\sum_{i=1}^{12} x_i^2 - \bar{x}^2 \\ &= \frac{319.46}{12} - 4.97^2 = 1.95 \\ &\Rightarrow S_x = 1.4 \end{aligned}$$

$$\begin{aligned} S_y^2 &= \frac{1}{12}\sum_{i=1}^{12} y_i^2 - \bar{y}^2 \\ &= \frac{146.41}{12} - 10.58^2 = 12.62 \\ \Rightarrow S_y &= 3.55 \qquad (3.55 \ \times 10^3 calories) \end{aligned}$$

$$\begin{aligned} Cov(x,y) &= \frac{1}{12}\sum_{i=1}^{12} x_i y_i - \bar{x}\bar{y} \\ &= \frac{682.06}{12} - 4.97 \times 10.58 = 4.26 \end{aligned}$$

Le coefficient de corrélation est alors :

$$\begin{aligned} r &= \frac{cov(x,y)}{S_x S_y} \\ &= \frac{4.26}{1.4 \times 3.55} = 0.857 \end{aligned}$$

La droit de régréssion D de y en x a pour $y = ax + b$

$$a = \frac{4.\,26}{1.95} = 2.\,18$$

et

$$b = 10.\,583 - 2.\,18 \times 4.97 = -0.25$$

on a $y = 2.18x - 0.25$ droite d'estimation de y en x

1.5 Exercices sur le chapitre 1

Exercice 1

D'un échantillon d'étudiants de sexe masculin, on a mesuré la masse de chacun. Les masses ont été arrondies à l'entier. Les données ont été groupées en 5 classes :

Masses en kg	$[55;59[$	$[60;64[$	$[65;69[$	$[70;74[$	$[75;79[$
Nombre d'étudiants	14	33	47	26	13

-Calculer les caractéristiques : classe modale, médiane, quartiles, le coeffcients de symétrie de Yule ?

Exercice 2

On donne la série suivante indiquant le nombre de réglettes fabriqueés dans une usine

Longueur x_i (en cm)	5	15	25	35	45
Effectifs n_i (en milliers)	5	7	8	6	4

-Trouver le mode, la moyenne et l'écart type ?

-Calculer le conffcient de symétrie de Pearson ?

Exercice 3

On a relevé l'âga et la pression systolique de 5 patients qui se sont présentés dans laboratoire :

x : âge	56	42	72	36	63
y : tension	14.7	12.5	16	11.8	14.9

-Tracer le nuage de points dans un repère orthogonal ?

-Déterminer par la méthode des moindres carrés la droite de régression de y en x ?

-calculer le coefficient de corrélation linéaire. conclution ?

Exercice 4

Cinq personnes soufrant d'obésité suivent un régime d'amincissement. Le tableau suivant donne le nombre de Kgs perdus par chacune d'elle pendant la période de cure suivie

Durée X (en mois)	3	1	2	4	5
Nombre Y de Kg perdus	6	4	5	9	11

1-Calculer la moyenne arithmétique de la variable X et celle de la variable Y.

2-Calculer la variance de la variable X et celle de la variable Y.

3- Calculer la covariance des variables statistiques X et Y, donner la droite de régression de Y en fonction X.

1.6 Série de TD N^{0}1 (2015-2016)

Université 08Mai 1945 Guelma **2015-2016**
3ème année Licence : Immunologie **Biostatistiques**

Série 1

Exercice 1

Soit le tableau statistique donnant le nombre d'enfants dans 116 familles

Nombre d'enfants	0	1	2	3	4	6
Nombre de familles n_i	6	18	25	33	21	13

a) Calculer les fréquences correspondantes ainsi que les fréquences cumulées, tracer la courbe des fréquences cumulées.

b) Trouver le mode, la médiane et les quartiles de cette distribution.

c) Trouver la moyenne et l'écart type de cette distribution.

d) Calculer le coefficients de symétrie de Pearson, puis celui de Yule. Conclusion.

Exercice 2

Soit le tableau donnant le poids de 133 étudiants :

Poids (Klogramme)	Nombre n_i d'étudiants
de 56 à moins de 58	5
$[58, 60[$	12
$[60, 62[$	18
$[62, 64[$	39
$[64, 66[$	36
$[66, 69[$	15
$[69, 72[$	8
$Total$	133

1) Construire l'histogramme de la distribution ainsi que la courbe des fréquences cumulées.

2) Calculer les caractéristique : classe modale, médiane, quartiles, moyenne et écart-type.

Exercice 3

Un pharmacien observe, durant les six (6) premier mois de l'ouverture de son officine, le chiffre d'affaire en million de Fcfa. Le résultat de l'observation

est résumé dans le tableau suivant où x désigne le numéro du mois et y le chiffre d'affaire correspondant.

x	1	2	3	4	5	6
y	12	13	15	19	21	22

1-Calculer les moyennes $\bar{x}$ et $\bar{y}$

2-Construire le nuage de points.

3-Calculer la variance S^2 et la covariance $COV(x, y)$

4-Démontrer que la droite de régression de y en fonction de x est

$$y = \frac{78}{35}x + 9.2$$

5-Calculer une estimation du chiffre d'affaire à la fin du 7ème mois.

Exercice 4

Le tableau suivant donne l'âge x et la moyenne y des maxima de tension artérielle en fonction de l'âge d'une population féminine.

x	36	42	48	54	60	66
y	11.8	14	12.6	15	15.5	15.1

1) Représenter graphiquement le nuage de points dans le plan muni d'un repère orthogonal.

2) Calculer la moyenne et la variance des séries statistiques aux caractères x et y.

3) a- Trouver une équation de la droite de régression de y en fonction de x.

b- Trouver une équation de la droite de régression de x en fonction y.

c- Représenter les deux droites sur le même graphique que celui utilisé pour le nuage de points.

4) Calculer le coefficient de corrélation linéaire.

5) Une personne de 70 ans a une tension artérielle de $16,2$. Cela vous paraît-il normal ?

Exercice 5

On donne la série statistique double de quatre éléments.

x	α	1.3	β	1.6
y	4	5	5	6

Trouver α et β sachant que la droite de régression de y en fonction de x a pour équation $y = 5x$.

***Exercice 6**

Montrer que le coefficient de corrélation r vérifie $-1 \leq r \leq 1$.

Chapitre 2

Variables Aléatoires

Après avoir réalisé une expérience, on ne s'intéresse bien souvent à une certaine fonction du résultat et non au résultat en lui-même. Lorsqu'on regarde une portion d'ADN, au lieu de vouloir connaitre tout la suite de nucléotides, on peut vouloir juste connaitre le nombre d'apparition d'un “”mot“”. Ces grandeurs (ou fonctions) auxquelles on s'intéresse sont en fait des fonctions réelles définies sur l'ensemble fondamental et sont appelées *variables aléatoires*.

On considère un ensemble Ω muni d'une probabilité P.

Définition 1 Une variable aléatoire X est une fonction de l'ensemble fondamental Ω à valeurs dans $\mathbb{R}$, $X : \Omega \to \mathbb{R}$.

Lorsque la variable X ne prend que des valeurs discrètes, on parle de *variable aléatoire discrète.*

Exemple 1 On jette deux dés distincts et on s'intéresse à la somme des points. On note X cette variable aléatoire, elle est définie par

$$X : \Omega \to \mathbb{R}.$$

$$(x, y) \to x + y$$

avec $\Omega = \{(1,1), (1,2), ..., (6,5), (6,6)\}$

L'ensemble des valeurs possibles de X est $\{2, 3, ..., 12\}$.

2.1 Loi de probabilité, Fonction de répartition

La loi de probabilité d'une variable aléatoire permet de connaitre les chances d'apparition des différentes valeurs de cette variable. On se place sur l'espace de probabilité (Ω, P).

Définition 2 Soit X une variable aléatoire. La loi de probabilité de X est définie par la fonction F, appelée fonction de répartition de la variable X, définie par

$$F(x) = P(X \leq x)$$

Remarque 1 On a $P(X \in \mathbb{R}) = 1$, car $P(X \in \mathbb{R}) = P(\Omega) = 1$.

2.2 Loi d'une variable aléatoire discrète

Une variable aléatoire est dite **discrète** si elle ne prend que des valeurs discontinues dans un intervalle donné (borné ou non borné). L'ensemble des nombres entiers est discret. En règle générale, toutes les variables qui résultent d'un dénombrement ou d'une numération sont de types discrètes.

Exemples Les variables aléatoires

-le nombre de petits par porté pour une espèce animale donnée (chat, marmotte, ect...)

-le nombre de bactéries dans 100 ml de préparation.

-le nombre de mutations dans une séquence d'ADN de $10kb$.

etc.... sont des variables aléatoires discrètes.

La fonction de répartition d'une variable discrète est constante par morceaux. Si X est une variable discrète à valeurs dans $\{x_1, ..., x_n\}$ avec

$x_1 < ... < x_n$ alors pour $x \in \mathbb{R}$

$$F(x) = \sum_{i=1}^{k} P(X = x_i)$$

avec k tel que $x_k < x < x_{k+1}$.

Exemple 2 On considère l'évènement w (lancer de 3 pièces). On introduit une variable aléatoire X définie par $X(w)$ (nombre de piles de l'évènement

w). La loi de probabilité de X est :

nombre de piles	$P(X = x_i)$	F_X
0	1/8	1/8
1	3/8	1/2
2	3/8	7/8
3	1/8	1

Dans le cas d'une variable aléatoire discrète, on utilise un diagramme en bâtons pour visualiser la distribution de probabilités et une fonction en escalier pour la fonction de répartition.

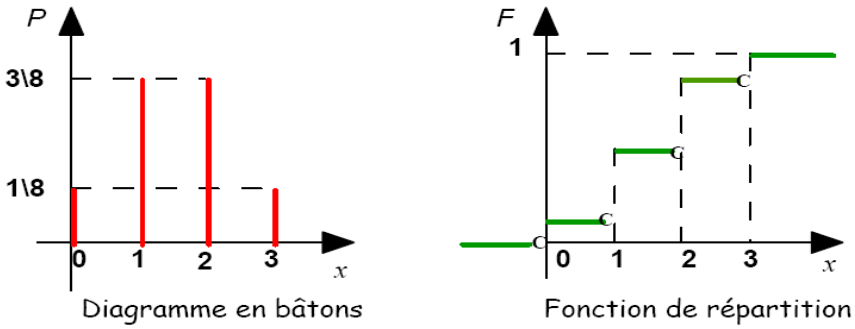


Diagramme en bâtons — Fonction de répartition

Exercice
Soit X une variable aléatoire discrète tel que

$$\Omega = \{3, 4, 5, 6\}$$

Déterminer la loi de probabilité X tel que

$$P(X = 3) = P(X = 4), P(X \leq 4) = \frac{1}{3} \text{ et } P(X > 5) = \frac{1}{2}$$

2.3 Loi d'une variable aléatoire à densité

Considérons la durée de vie d'une bactérie. On conçoit facilement que la probabilité que cette durée de vie vaille exactement une certaine valeur est nulle. Par exemple, il est quasiment impossible qu'une bactérie vive exactement 1 an 0 mois, 0 heure, 0 minute. La fonction de répartition d'une telle variable est par conséquent continue. On peut par contre s'intéresser à la probabilité que la bactérie vive moins d'un an.

On ne verra dans ce cours que des variables qui sont soit discrètes soit continues même s'il existe des variables plus complexes

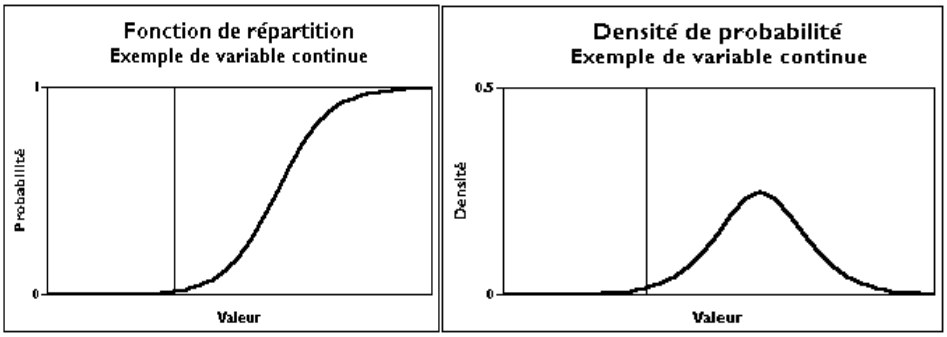
Définition 3 Une variable aléatoire X est à densité, ou continue, s'il existe une fonction f définie sur $\mathbb{R}$ telle que la fonction de répartition de X s'écrit

$$F_X(x) = \int_{-\infty}^{x} f(t)dt$$

où f est une fonction intégrable sur $\mathbb{R}$ satisfaisant les conditions suivantes :

1. $f(t) \geq 0$ pour tout $t \in \mathbb{R}$,
2. $\int_{-\infty}^{+\infty} f(t)dt = 1$.

Une fonction qui vérifie les conditions 1 et 2 est appelée *densité de probabilité*.



2.4 Espérance et variance d'une variable aléatoire

L'idée intuitive de l'espérance puise son origine dans les jeux de hasard. Considérons le jeu suivant : on lance un dé plusieurs fois de suite. Supposons que pour une mise de 1 dinar, on gagne 1 dinar si le résultat obtenu est pair, 2 dinars si le résultat est 1 ou 3, et on perd 3 dinars si le résultat est 5. Est-il intéressant de jouer à ce jeu ? Quel peut-être le gain moyen ?

Soit X la variable aléatoire correspondant au nombre dinars gagnés ou perdus. La loi de X est

k	-3	1	2
P	1/6	1/2	1/3

L'espérance de gain, noté $E[X]$, est alors

$$E[X] = -3 \times 1/6 + 1 \times 1/2 + 2 \times 1/3 = 2/3$$

Le joueur gagne donc en moyenne 2/3 de dinars pour une mise de 1 dinar.

Définition 4 L'espérance d'une variable aléatoire X est notée $E[X]$. Elle représente la valeur moyenne prise par la variable X :

1-Si X est une variable discrète à valeurs dans l'ensemble $D = \{x_1, x_2, ..., x_n\}$, lorsque la somme est bien définie, son espérance est

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

2-Si X est une variable à densité f, lorsque l'intégrale est bien définie, son espérance est

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx$$

Lorsqu'une variable X vérifie $E[X] = 0$, on dit que la variable est *centrée*.

2.5 Variance et écart type

On a vu que l'espérance correspondait à la valeur moyenne d'une variable aléatoire. L'écart type représente l'écart moyen (la distance moyenne) entre la variable et sa moyenne. Elle mesure la dispersion d'une variable, plus l'écart-type est grand plus la variable prend des valeurs qui peuvent être éloignées les unes des autres, plus l'écart-type est petit plus la variable prend des valeurs proches de sa moyenne.

Définition 5 La variance d'une variable aléatoire X, notée S^2 (ou $Var(X)$), est définie par

$$S^2 = E\left[X^2\right] - E[X]^2$$

L'écart type (notée S ou $\sigma(x)$) est la racine carrée de la variance.

Exemple 3 Supposons que la durée de vie T d'une bactérie est modélisée par (la loi exponentielle) de densité $f(t) = \lambda \exp(-\lambda t)$ pour $t \geq 0$ pour une certaine valeur de λ. La variance de la durée de vie de la bactérie étudiée est $S_x^2 = 1/\lambda^2$.

2.6 Exemples de variables aléatoires discrètes

1-Loi de Bernoulli de paramètre p notée $b(p)$. Une v.a. X suit une loi de Bernoulli de paramètre $p \in [0;1]$ si elle ne prend que les deux valeurs 0 et 1 avec :

$$P(X=1) = p; P(X=0) = 1-p = q$$

Son espérance est $E[X] = 0 \times (1-p) + 1 \times p = p$. Sa variance est

$$\begin{aligned} S^2 &= E[X^2] - E[X]^2 \\ &= p - p^2 \\ &= p(1-p). \end{aligned}$$

Exemple 4 Pile ou face avec $p = 1/2$ si la pièce est équilibrée, $p \neq 1/2$ si elle est truquée.

2-Loi de Poisson Cette loi intervient dans les processus aléatoires dont les éventualités sont faiblement probables et survenant indépendamment les unes des aoutes : cas de phénomènes accidentels, d'anomalies diverses, de problèmes d'encombrement (files d'attente), de rupture de stocks, etc...

On dit qu'une v.a. discrète X suit une loi de Poisson de paramètre $\lambda > 0$ si elle prend des valeurs entières positives ou nulles et

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, k \in \mathbb{N}$$

La loi de Poisson de paramètre $\lambda > 0$ est notée $P(\lambda)$. Son espérance est $E[X] = \lambda$, Sa variance est $S^2 = \lambda$.

2.7 Exemples de variables aléatoires continues

1-**Loi normale (ou gaussienne)**

a) Définition 6 On dit que la v.a. X suit une loi normale $N(m, \sigma)$ si elle a pour densité la fonction

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), x \in \mathbb{R}$$

Son espérance est $E[X] = m$. Sa variance est $S^2 = \sigma^2$.

b) La distribution normale centrée réduite

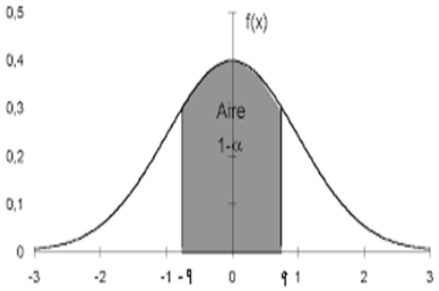
On dit que la distribution est centrée si son espérance m est nulle ; elle est dite réduite si sa variance σ^2 (et son écart-type σ) est égale à 1. La distribution normale centrée réduite $N(0,1)$ est donc définie par la formule $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), x \in \mathbb{R}$

Les probabilités correspondant aux divers intervalles ont été calculées et regroupées dans une table numérique. Ainsi la table de la loi normale centrée réduite permet, à partir d'une probabilité α donnée, de trouver les bornes $-q, +q$ d'un intervalle symétrique autour de 0, tel que

$$P(-q \leq X \leq q) = 1 - \alpha$$

et

$$P(X < q) = P(X > q) = \frac{\alpha}{2}$$



Loi normale centre rduite
$N(0,1)$

c)Transformation d'une loi normale quelconque en loi normale centrée réduite

Soit F la fonction de répartition de loi normal $N(m, \sigma)$, pour calcul $F(a) = P(X \leq a)$

on pose $X = m + \sigma Z$ alors

$$Z = \frac{X - m}{\sigma}$$

où Z suit une loi normal $N(0,1)$.on a

$$\begin{aligned} F(a) &= P(X \leq a) \\ &= P(\frac{X-m}{\sigma} \leq \frac{a-m}{\sigma}) \\ &= P(Z \leq \frac{a-m}{\sigma}) \\ &= \Phi\left(\frac{a-m}{\sigma}\right) \end{aligned}$$

est les valeurs de Φ sont donnés par la table de la loi $N(0,1)$.

C'est une loi très importante pour plusieurs raisons :

-Elle apparait dans de nombreux problèmes courants (pour les modéliser),

-Bien souvent, on peut approcher une loi par une loi normale.

-De plus, on dispose de la table de ses valeurs à laquelle on se réfere pour des calculs approchés.

Remarque 2 Soit Z suit une loi normale $N(0,1)$ et Φ la fonction de répartition, comme la fonction Φ est symétrique par rapport à l'axe $(x=0)$ alors

$$\Phi(-x) = 1 - \Phi(x)$$

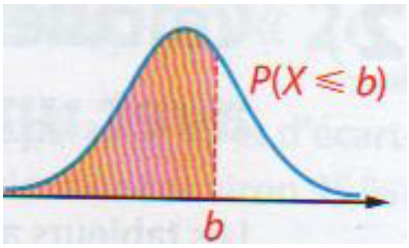
Exemple 5 Soit X une v.a de loi normale $N(20,5)$, calculer les probabilités

$$P(X \leq 28), P(X > 28) \text{ et } P(12 \leq X \leq 28)$$

$m = 20$ et $\sigma = 5$

Transformation de X en Z où Z suit une loi normale $N(0,1)$.

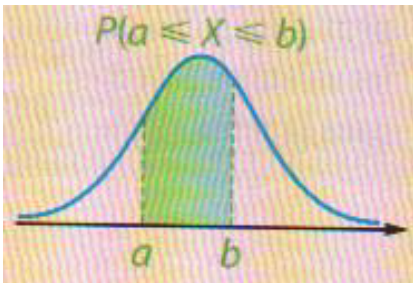
$$\begin{aligned} P(X &\leq 28) = P(\frac{X-20}{5} \leq \frac{28-20}{5}) \\ &= P(Z \leq 1.6) \\ &= \Phi(1.6) \\ &= 0.9452 \end{aligned}$$



$b = 28$

$$\begin{aligned} P(X &> 28) = 1 - P(X \leq 28) \\ &= 1 - 0.9452 \\ &= 0.054\,8 \end{aligned}$$

$$\begin{aligned} P(12 \leq X \leq 28) &= P(\frac{12-20}{5} \leq \frac{X-20}{5} \leq \frac{28-20}{5}) \\ &= P(-1.6 \leq Z \leq 1.6) \\ &= \Phi(1.6) - \Phi(-1.6) \\ &= \Phi(1.6) - (1 - \Phi(1.6)) \\ &= 2\Phi(1.6) - 1 \\ &= 0.8904 \end{aligned}$$



$a = 12$ et $b = 28$

Exemple 6 La température T dans une chambre Froide suit une loi $N(0, 1)$ où T est en degrés Celsius, probabilité que la température soit com-

prise entre -1.5 et 1.5 degrés

$$\begin{aligned} P(-1.5 &\leq X \leq 1.5) = \Phi(1.5) - \Phi(-1.5) \\ &= 2\Phi(1.5) - 1 \\ &= 0.8664 \end{aligned}$$

Remarque 3 La distribution Gaussienne est une des distributions les plus utilisées en statistique. Beaucoup de variables biologiques ont un comportement approximativement Gaussien.

2-Loi du khi-deux :

Soient $X_1, ..., X_n$ des v.a. indépendantes de même loi $N(0;1)$. Posons $\chi^2 = \sum_{i=1}^{n} X_i^2$. Par définition, la v.a.χ^2 suit une loi du khi-deux à n degrés de liberté (abréviation d.d.l.). On note cette loi $\chi^2(n)$. On a

$$E[\chi^2] = n \text{ et } S^2 = 2n.$$

Cette loi et sur tout utile dans les testes statistique.

3-Loi de Student

Soient deux v.a indépendantes X et Y distribution suivant une loi normale et khi-deux respectivement. La variable

$$T = \frac{X}{\sqrt{\frac{Y}{X}}}$$

suit une loi de student à n degrés de liberté.

On a

$$E[T] = 0, \qquad n > 1 \qquad \text{et} \qquad S^2 = \frac{n}{n-2}, \qquad n > 2$$

2.8 Exercices sur le chapitre 2

Exercice 1

Une urne contient 3 sortes de boules de poids différents : 7 boules de poids $1kg$, 5 boules de poids $3kg$ et 3 boules de poids $5kg$. On tire au hasard une boule de l'urne et on note X son poids.

1. Déterminer la loi de la variable X.

2. Calculer l'espérance et la variance de X.

Exercice 2

Considérons deux parents hétérozygotes de génétopye Aa tels que leur enfants peuvent avoir les génotypes AA, Aa ou aa avec probabilité

$$P(AA) = 1/4; P(Aa) = 1/2; P(aa) = 1/4.$$

Supposons qu'ils aient 4 enfants.

1. Quelle est la probabilité qu'exactement l'un deux aient le génotype aa ?

2. Quelle est la probabilité qu'au moins l'un deux ait le génotype aa ?

Exercice 3

La proportion des groupes sanguins en Algeria est environ :

$$A = 44\%, B = 13\%, AB = 3\%, O = 40\%$$

On considère la répartition de ces différents groupes sur 50 étudiants.

1. Donner la loi de la variable X égale au nombre d'étudiants de groupe O.

2. Donner la loi de la variable Y égale au nombre d'étudiants de groupe AB.

3. Calculer $P(Y \leq 5)$, l'espérance et la variance de Y .

Exercice 4.

Dans un pays donné, le taux de cholestérol sérique d'un individu pris au hasard est modélisé par une loi normale avec une moyenne de 200 $mg/100mL$ et un écart-type de 20 $mg/100mL$.

1. Quelle est la probabilité qu'un individu pris au hasard dans ce pays ait un taux de cholestérol inférieur à 160 $mg/100mL$?

2. Quelle proportion de la population a un taux de cholestérol compris entre 170 et 230 $mg/100mL$?

3. Dans un autre pays, le taux moyen de cholestérol sérique est de 190 $mg/100mL$, pour le même écart-type. Reprendre les questions précédentes.

4. On choisit un individu au hasard dans le premier pays, puis dans le second. Quelle est la probabilité que le premier individu ait un taux supérieur au second ?

Exercice 5

Un chercheur a étudié l'âge moyen auquel les premiers mots du vocabulaire apparaissent chez les jeunes enfants. Une étude effectuée auprès d'un millier de jeunes enfants montre que les premiers mots apparaissent, en

moyenne, à 11.5 mois avec un écart-type de 3.2 mois. La distribution des âges étant normale, on souhaite

-Évaluer la proportion d'enfants ayant acquis leurs premiers mots avant 10 mois.

-Évaluer la proportion d'enfants ayant acquis leurs premiers mots entre 8 mois et 12 mois.

2.9 Série de TD N^{0}2 (2015-2016)

Université 08Mai 1945 Guelma **Biostatistiques, 2015-2016**
3ème année Licence : Immunologie

Série 2

Exercice 1

On jette deux dés réguliers à quatre fases et on fait la somme X des points obtenus

1- Donner la loi de la variable aléatoire X obtenue ?

2- Quelle est sa moyenne ? Sa variance ?

3-Calculer : $P[X \leqslant 5]$,$P[X > 5]$; $P[3 \leqslant X < 5]$.

4- Quelle est la fonction de répartition de X ?

Exercice 2 *(espérance de vie d'une population)*

On suppose que la durée de vie d'un individu dans une population donnée est modélisée par une v.a.continue X dont la fonction densité de probabilité est donnée par :

$$f(x) = \begin{cases} kx^2(100-x)^2 & \text{si } 0 \leq x \leq 100 \\ 0 & \text{si non} \end{cases}$$

où k est une contante positive.

1. Déterminer la valeur de k.

2. Calculer la probabilité qu'un individu meure entre 60 ans et 70 ans.

3. Quelle est l'espérance de vie d'un individu dans cette population ?

Exercice 3

Un magasin spécialisé reçoit en moyenne 4 clients par jour, le nombre de clients étant distribué selon une loi de *Poisson*. Calculer la probabilité que le magasin soit visité le mercredi par :

1- aucun client ;

2- 5 clients ;

3- au moins 6 clients.

Exercice 4

En 1955, *Wechler* a proposé de mesurer le QI (Quotient Intellectuel) des adultes grâce à deux échelles permettant de mesurer les compétences verbales et les compétences non verbales. On compare le score global de la personne testée avec la distribution des scores obtenu par un échantillon

représentatif de la population d'un âge donné, dont les performances suivent une loi normale ayant pour moyenne 100 et pour écart-type 15.

1/ Quel est le pourcentage de personnes dont le QI est inférieur à 80 ?

2/ Quelle chance a-t-on d'obtenir

un QI compris entre 100 et 110 ?

un QI compris entre 105 et 110 ?

3/ Un patient obtenant un score de 69 fait-il partie des 5% inférieur de la distribution ?

4/ En dessous de quel QI se trouve le tiers des individus ?

Exercice 5

Sur un grand nombre de personnes on a constaté que la répartition du taux de *cholestérol* suit une loi normale avec les résultats suivants :

- 56% ont un taux inférieur à $165cg$;
- 34% ont un taux compris entre $165cg$ et $180cg$;
- 10% ont un taux supérieur à $180cg$.

Quelle est le nombre de personnes qu'il faut prévoir de soigner dans une population de 10000 personnes, si le taux maximum toléré sans traitement est de $182cg$?

Chapitre 3

Théorie d'estimation

L'objectif de l'estimation statistique est le suivant : évaluer certaines grandeurs associées à une population à partir d'observations faites sur un échantillon. Bien souvent, ces grandeurs sont des moyennes ou des variances. On prendra soin de distinguer ces grandeurs théoriques (inconnues et à estimer) de celles observées sur un échantillon.

Exemples de problèmes :

-Quelle est la fréquence (probabilité) de survenue d'un certain cancer chez les souris ?

-Quelle est l'écart moyen de la glycémie d'un patient autour de sa glycémie moyenne ?

On apporte deux types de réponses à ces questions : à partir d'un échantillon,

1. On « calcule »une valeur qui semble être la meilleure possible : on parle d'estimation ponctuelle,

2. On « calcule »un intervalle de valeurs possibles : c'est la notion d'intervalle de confiance ou d'estimation par intervalle.

Définition 1 On appelle n-échantillon de loi P un suite $(X_1, X_2, ..., X_n)$ de n variables aléatoires indépendantes et de même loi P.

3.1 Estimation ponctuelle

Définition 2 On cherche à estimer une valeur θ inconnue liée à un certain phénomène aléatoire, en général, la moyenne m ou la variance S^2 ou encore l'écart-type de la loi du phénomène.

Pour ce faire, on dispose d'observations indépendantes du phénomènes, c.à.d de variables aléatoires $X_1; ...; X_n$ indépendantes et de même loi (celle du phénomène). On parle d'un échantillon. On définit à partir de l'échantillon une nouvelle variable aléatoire notée X dont les valeurs seront proches de celle de la grandeur θ à estimer.

a) Estimation de la moyenne (loi de grands nombres)

Soit $X_1, X_2, ..., X_n$ sont des variables aléatoires indépendantes de même moyenne m et de même S^2 (variance), l'estimation de la moyenne

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

on appelle la moyenne empirique ($\hat{m}$ est un estimateur m).

b) **Estimation de la variance**

-Dans le cas m (la moyenne) est connue

$$\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

où $\hat{S}^2$ est un estimateur S^2.

-Dans le cas m (la moyenne) est inconnue

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{m})^2$$

Exemple 1 m : moyenne des poids des nouveaux nés en Alegria. Ici, on prendra comme estimateur $\bar{X}$ la variable aléatoire donnée par la moyenne (arithmétique) observée sur un échantillon de 10 nouveaux nés. On note cet estimateur en général $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$.

3.1.1 Méthode de maximum de vraisemblance

Soit un paramètre θ (en générale m ou S) d'une population à estimer. Il faut trouver un estimateur T à partir d'un échantillon.

Définition 3 soit $x = (x_1, x_2, ..., x_n)$ une réalisation d'un échantillon $X = (X_1, X_2, ..., X_n)$ de n variables aléatoires, la fonction $L(x_1, x_2, ..., x_n, \theta)$ est donnée par

Dans le cas discrète

$$\begin{aligned} L(x_1, x_2, ..., x_n, \theta) &= L_n(x, \theta) \\ &= P(x, \theta) \\ &= P(x_1, \theta) \times P(x_2, \theta) \times ... \times P(x_n, \theta) \end{aligned}$$

Dans le cas continue

$$\begin{aligned} L(x_1, x_2, ..., x_n, \theta) &= L_n(x, \theta) \\ &= f(x, \theta) \\ &= f(x_1, \theta) \times f(x_2, \theta) \times ... \times f(x_n, \theta) \end{aligned}$$

tel que L fonction de θ pour x fixé, s'appelle la vraisemblance de x.

La méthode de maximum de vraisemblance (M.V) consiste à choisir comme estimateur de θ, la valeur particulière de θ qui maximise la fonction de vraisemblance $L(x_1, x_2, ..., x_n, \theta)$.

Cet estimateur T est solution de l'équation :

$$\frac{\partial L(x_1, x_2, ..., x_n, \theta)}{\partial \theta} = 0$$

ou

$$\frac{\partial l(x_1, x_2, ..., x_n, \theta)}{\partial \theta} = 0$$

où $l(x_1, x_2, ..., x_n, \theta) = \ln(L(x_1, x_2, ..., x_n, \theta))$

Remarque 1 Dans le cas discrète :

$$\begin{aligned} l(x_1, x_2, ..., x_n, \theta) &= \ln\left[L(x_1, x_2, ..., x_n, \theta)\right] \\ &= \ln\left[P(x_1, \theta) \times P(x_2, \theta) \times ... \times P(x_n, \theta)\right] \\ &= \sum_{i=1}^{n} \ln\left[P(x_i, \theta)\right] \end{aligned}$$

Dans le cas continue :

$$\begin{aligned} l(x_1, x_2, ..., x_n, \theta) &= \ln\left[L(x_1, x_2, ..., x_n, \theta)\right] \\ &= \ln\left[f(x_1, \theta) \times f(x_2, \theta) \times ... \times f(x_n, \theta)\right] \\ &= \sum_{i=1}^{n} \ln\left[f(x_i, \theta)\right] \end{aligned}$$

Exemple 2 Soit une variable aléatoire X suivant une loi de poisson, estimer le paramètre λ de la loi, en utilisant la méthode du M.V.

Pour une variable aléatoire X suivant une loi de poisson on a

$$P(x,\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}, x = 0, 1, 2, ...$$

où λ est le paramètre inconnue.

Calcul la fonction de vraisemblance

$$\begin{aligned} L(x_1, x_2, ..., x_n, \lambda) &= P(x_1,\lambda) \times P(x_2,\lambda) \times ... \times P(x_n,\lambda) \\ &= \frac{\lambda^{x_1}}{x_1!}e^{-\lambda} \times \frac{\lambda^{x_2}}{x_2!}e^{-\lambda} \times ... \times \frac{\lambda^{x_n}}{x_n!}e^{-\lambda} \end{aligned}$$

et

$$\begin{aligned} l(x_1, x_2, ..., x_n, \lambda) &= \sum_{i=1}^{n} \ln\left[P(x_i,\lambda)\right] \\ &= \sum_{i=1}^{n} \ln\left[\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right] \\ &= \sum_{i=1}^{n} x_i \ln(\lambda) - \sum_{i=1}^{n} \ln(x_i!) - n\lambda \end{aligned}$$

donc

$$\frac{\partial l(x_1, x_2, ..., x_n, \lambda)}{\partial \lambda} = \sum_{i=1}^{n} x_i.\frac{1}{\lambda} - n$$

$$\frac{\partial l(x_1, x_2, ..., x_n, \lambda)}{\partial \lambda} = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = m$$

alors estimateur de λ est m.

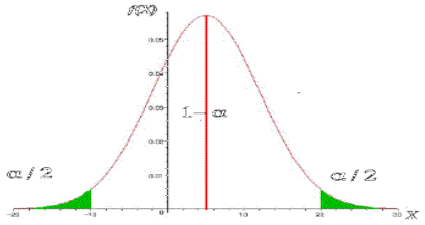
3.2 Estimation par intervalle

L'estimation est dit par intervalle si on estime un paramètre inconnu θ par une construction d'un intervalle $[a, b]$ on a :

$$P(a < \theta < b) = 1 - \alpha$$

telle que

- a et b dits limites de confiance
- $1-\alpha$: niveau de confiance
- α : risque d'erreur



3.2.1 Intervalle de confiance de la moyenne

On veut estimer la moyenne m d'une population normale à l'aide d'un échantillon aléatoire.

a) Estimation de la moyenne quand la variance est connue et (pour un grand échantillon)

Si $n \geq 30$ (n taille de l'échantillon) la distribution d'échantillonnage de la moyenne approche la distribution normale (un loi normale $N(m,\sigma)$) .

Soit X une variable aléatoire de loi $N(m,\sigma)$

Théorème 1 *Lorsque σ^2 est connu, un intervalle de confiance au niveau $1-\alpha$ de m est*

$$\left[\bar{x} - u\frac{\sigma}{\sqrt{n}}; \bar{x} + u\frac{\sigma}{\sqrt{n}}\right]$$

où la valeur u est lue dans la table normale centrée réduite $N(0,1)$ telle que $\Phi(u) = 1 - \frac{\alpha}{2}$.

Exemple 3 Soit $n = 100$; $\sigma = 2.5$ et $\bar{x} = 11.5$

Donner un intervalle de confiance de niveau 0.95 pour m

Ici, $\alpha = 0.05$ et $1 - \frac{\alpha}{2} = 0.975$. Le quantile d'ordre 0.975 de la loi $N(0,1)$ est $u = 1.96$. L'intervalle de confiance est :

$$\left[11.5 - 1.96\frac{2.5}{\sqrt{100}}; 11.5 + 1.96\frac{2.5}{\sqrt{100}}\right]$$

donc

$$m \in [11.01; 11.99]$$

b) Estimation de la moyenne quand la variance est inconnue (pour un grand échantillon)

Théorème 2 *Lorsque σ^2 est inconnu un intervalle de confiance au niveau $1-\alpha$ de m est*

$$\left[\bar{x} - u\frac{\hat{S}}{\sqrt{n}}; \bar{x} + u\frac{\hat{S}}{\sqrt{n}}\right]$$

où $\hat{S}$ est un estimateur de σ et la valeur u est lue dans la table normale centrée réduite $N(0,1)$ telle que $\Phi(u) = 1 - \frac{\alpha}{2}$.

Exemple 4 On a effectué 90 mesures de concentration d'une solution de de fluorescéine. On a observé une moyenne empirique $\bar{x} = 4.38\ mg/l$ et un écart-type empirique $\hat{S} = 0.08\ mg/l$. Donner un intervalle de confiance pour la concentration réelle de la solution, aux niveaux de confiance 0.95 et 0.99.

$\Phi(u) = 1 - \frac{0.05}{2} = 0.975$ on a $u = 1.96$. L'intervalle de confiance de niveau 0.95 est :

$$m \in \left[4.38 - 1.96\frac{0.08}{\sqrt{90}}; 4.38 + 1.96\frac{0.08}{\sqrt{90}}\right]$$

on a

$$m \in [4.363; 4.397]$$

Le quantile d'ordre 0.995 de la loi $N(0,1)$ est 1.96
$(\Phi(u) = 1 - \frac{0.01}{2} = 0.995 \Rightarrow u = 2.5758)$
L'intervalle de confiance de niveau 0.99 est :

$$m \in \left[4.38 - 2.5758\frac{0.08}{\sqrt{90}}; 4.38 + 2.5758\frac{0.08}{\sqrt{90}}\right]$$

on a

$$m \in [4.358; 4.402]$$

c) Intervalle de confiance avec la distribution t

Pour des échantillons de taille $n < 30$ extraits d'une population suivant une loi normale d'écart-type σ inconnu, on utilise la distribution t de Student pour déterminer l'intervalle de confiance de la moyenne.

Théorème 3 *Lorsque σ^2 est inconnu, un intervalle de confiance au niveau $1-\alpha$ de m est*

$$\left[\bar{x} - t_{n-1,\frac{\alpha}{2}}\frac{\hat{S}}{\sqrt{n}}; \bar{x} + t_{n-1,\frac{\alpha}{2}}\frac{\hat{S}}{\sqrt{n}}\right]$$

où $\hat{S}$ est un estimateur de σ et la valeur $t_{n-1,\frac{\alpha}{2}}$ est lue dans la table de Student à $k = n-1$ dégrés de liberté (ddl) et $\gamma = \frac{\alpha}{2}$.

Exemple 5 Pour $n = 10$, avec un niveau de confiance de 95% et un intervalle symétrique on obtient l'intervalle

$$\left[\bar{x} - 2,26\frac{\hat{S}}{\sqrt{10}}; \bar{x} + 2,26\frac{\hat{S}}{\sqrt{10}}\right]$$

Exemple 6 Un examen de probabilité est organisé pour promotion très nombreuse on extrait un échantillon de 4 notes

$$12.5; 10; 14.5, 14$$

Déterminer l'intervalle de confiance à 95% pour la moyenne de tout la promotion

$n = 4 < 30$, en utilisant la distribution t de student on a :

$$m \in \left[\bar{x} - t_{3,\frac{0.05}{2}}\frac{\hat{S}}{\sqrt{n}}; \bar{x} + t_{3,\frac{0.05}{2}}\frac{\hat{S}}{\sqrt{n}}\right]$$

et niveau de confiance $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$ et $ddl = k = 3$, on a

$$t_{3,\frac{0.05}{2}} = t_{3,0.025} = 3.182$$

et $\bar{x} = 12.75$, $\hat{S}^2 = 4.08$ est une estimation de la valeur inconnue σ^2 , donc

$$m \in \left[12.75 - 3.182\frac{\sqrt{4.08}}{\sqrt{4}}; \bar{x} + 3.182\frac{\sqrt{4.08}}{\sqrt{4}}\right]$$

$$m \in [9.535; 15.964]$$

Exemple 7 On suppose que le taux de cholestérol X d'un individu choisi au hasard dans une population donnée suit une loi normale. Sur un échantillon de 20 individus, on constate la moyenne des taux observés est $\bar{x}$= 1.55 (gr pour $millr$). On constate aussi une variance empirique $\hat{S}^2 = 0.25$.

Donner un intervalle de confiance pour la moyenne m au niveau de confiance 0.95 ?

3.2.2 Intervalle de confiance de la différence de deux moyennes

Soient $(X_1, X_2, ..., X_{n_1})$ un échantillon d'une population suivant la loi normale $N(m_1, \sigma_1)$ et $(Y_1, Y_2, ..., Y_{n_2})$ un échantillon d'une population suivant la loi normale $N(m_2, \sigma_2)$. On pose

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i; \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i \text{ et } D = \bar{X} - \bar{Y}$$

Théorème 4 *Si σ_1 et σ_2 sont connues, un intervalle de confiance de $m_1 - m_2$ au niveau de confiance $1 - \alpha$ est*

$$\left[D - u\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}; D + u\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}\right]$$

où la valeur u est lue dans la table normale centrée réduite $N(0, 1)$ telle que $\Phi(u) = 1 - \frac{\alpha}{2}$.

3.2.3 Intervalle de confiance d'une proportion

Dans une certaine population, la proportion d'individus ayant une propriété donnée est égale à p. Soit X le nombre d'individus d'un échantillon de taille n ayant la propriété.

On ne sait pas déterminer exactement un intervalle de confiance. On utilise des solutions approchées, qui fonctionnent lorsqu'on dispose d'échantillon de grande taille. Ainsi, lorsque n est grand ou/et p voisin de 0.5 on peut approcher la loi binomiale par une loi normale.

On considère une population (P) contenant deux types d'individus A et B en proportion p et $1 - p$. Soit $X_1, X_2, ..., X_n$ un n-échantillon de loi Bernoulli $B(p)$.

Théorème 5 *Un intervalle de confiance approché de p au niveau $1 - \alpha$ est donnée par*

$$\left[T - u\frac{\sqrt{T(1-T)}}{\sqrt{n}}; T + u\frac{\sqrt{T(1-T)}}{\sqrt{n}}\right]$$

où T la fréquence de type A (T estimateur sans biais de p)

$$T = \frac{card(A)}{n}$$

et la valeur u est lue dans la table normale centrée réduite $N(0,1)$ telle que $\Phi(u) = 1 - \frac{\alpha}{2}$.

Exemple 8 Douze des 75 arbres d'un échantillon aléatoire sont contaminés par une maladie. Déterminer un intervalle de confiance au niveau 95% pour p la proportion d'arbres malades.

$T = \frac{12}{75} = 0.16$ et niveau de confiance $1 - \alpha = 0.95$ et $n = 75$.

$$\Phi(u) = 1 - \frac{0.05}{2} = 0.975$$

Dans la table $N(0.1)$, on trouve $u = 1.96$.

Intervalle de confiance de p est

$$p \in \left[0.16 - 1.96\frac{\sqrt{0.16(1-0.16)}}{\sqrt{75}}; 0.16 + 1.96\frac{\sqrt{0.16(1-0.16)}}{\sqrt{75}}\right]$$

on a

$$p \in [0.077; 0.243]$$

Exercice. On a observé un échantillon de taille $n = 500$ d'adolescents de 15 ans, dans lequel 210 présentent un surpoids. Soit p la proportion d'adolescents de 15 ans qui présentent un surpoids. Donner un intervalle de confiance pour p, aux niveaux de confiance 0.95 et 0.99.

3.2.4 Intervalle de confiance de la variance

a) Estimation de la variance quand la moyenne est connue

Théorème 6 *Lorsque m est connu un intervalle de confiance au niveau $1 - \alpha$ de la variance σ^2 est*

$$\left[\frac{1}{\chi_{n-1,\frac{\alpha}{2}}}\sum_{i=1}^{n}(x_i - m)^2; \frac{1}{\chi_{n-1,1-\frac{\alpha}{2}}}\sum_{i=1}^{n}(x_i - m)^2\right]$$

où les valeurs $\chi_{n-1,\frac{\alpha}{2}}$ et $\chi_{n-1,1-\frac{\alpha}{2}}$ est leu dans la table du Khi-deux avec $(n-1)$ dégrées de liberté (ddl).

b) Estimation de la variance quand la moyenne est inconnue

A nouveau, comme m est inconnue, l'idée est de la remplacer par son estimation $\bar{X}$. L'intervalle de confiance de la variance σ^2 se calcule alors à partir de l'échantillon de taille n par

$$\left[\frac{(n-1)\,\hat{S}^2}{\chi_{n-1,\frac{\alpha}{2}}}; \frac{(n-1)\,\hat{S}^2}{\chi_{n-1,1-\frac{\alpha}{2}}}\right]$$

où $\hat{S}^2$ est un estimateur de σ^2 et les valeurs $\chi_{n-1,\frac{\alpha}{2}}$ et $\chi_{n-1,1-\frac{\alpha}{2}}$ est leu dans la table du Khi-deux avec $(n-1)$ dégrées de libertés (ddl).

Remarque 2 Intervalle de confiance au niveau $1-\alpha$ d'écart-type σ est

$$\left[\sqrt{\frac{(n-1)\,\hat{S}^2}{\chi_{n-1,\frac{\alpha}{2}}}}; \sqrt{\frac{(n-1)\,\hat{S}^2}{\chi_{n-1,1-\frac{\alpha}{2}}}}\right]$$

3.3 Exercices sur le chapitre 3

Exercice1

Pour étudier la pourriture des pommes de terre, un chercheur injecte à 13 pommes de terre des bactéries qui causent cette pourriture. Il mesure ensuite la surface pourrie (en mm^2) sur ces 13 pommes de terre. Il obtient une moyenne empirique de 7.84 mm^2 pour une variance empirique de 14.13. On modélise la surface pourrie d'une pomme de terre par une loi normale $N(m, \sigma)$.

1. Calculer un intervalle de confiance pour m au niveau 0.95 puis 0.99.

2. Calculer un intervalle de confiance pour σ^2 au niveau 0.95 puis 0.99.

Exercice 2

On a mesuré le poids de raisin produit par pied sur 10 pieds pris au hasard dans une vigne. On a obtenu les résultats suivants exprimés en kilogrammes :

2.4	3.4	3.6	4.1	4.3	4.7	5.4	5.9	6.5	6.9

On modélise le poids de raisin produit par une souche de cette vigne par une variable aléatoire de loi $N(m, \sigma)$.

1. Calculer la moyenne et la variance empiriques de l'échantillon ?

2. Donner un intervalle de confiance de niveau 0.95 pour m.

3. Donner un intervalle de confiance de niveau 0.95 pour σ^2.

4. On suppose désormais que l'écart-type des productions par pied est connu et égal à 1.4. Donner un intervalle de confiance de niveau 0.95 pour m ?

5. Quel nombre de pieds au minimum devrait-on observer pour estimer m au niveau de confiance 0.99 avec une précision de plus ou moins 500 grammes ?

Exercice 3. Une clinique a proposé une nouvelle opération chirurgicale, et a connu 40 échecs, sur 200 tentatives. On note p le pourcentage de réussite de cette nouvelle opération.

1. Quelle estimation de p proposez-vous ?

2. En utilisant l'approximation normale, donner un intervalle de confiance pour p de niveau de confiance 0.95.

3. Combien d'opérations la clinique devrait-elle réaliser pour connaître le pourcentage de réussite avec une précision de plus ou moins 1%, au niveau de confiance 0.95 ?

Exercice 4

Une expérience sur les alphaglobulines (composent les protéines du sérum) a donné les résultats suivants :

8; 12; 13.5; 16; 20 et 21

Déterminer l'intervalle de confiance de la moyenne au niveau 99%.

Exercice 5

On considère un lot de pelotes de laine dont, on suppose que les poids (en grammes) suivent une distribution normale de moyenne m et d'écart type . On extrait un échantillon de 8 pelotes, et on obtient :

53; 48; 52; 50; 47; 49; 55 et 52

–Déterminer les estimations ponctuelles de la moyenne et l'écart-type.

–Donner un intervalle de confiance au niveau 95%, pour la moyenne si on suppose que est inconnue.

Exercice 6

Soit une machine M qui fabrique des comprimés. On considère la population de taille très grande, formée de tous les comprimés fabriqués en une journée par M. Pour étudier le caractère "poids du comprimé" sur cette population, on prélève au hasard et de manière non exhaustive un échantillon de 6 comprimés que l'on pèse. On a obtenu les résultats suivants :

Poids en g	0.79	0.8	0.78	0.81	0.65	0.59

Donne une estimation ponctuelle de la moyenne et de l'ecart type du poids des comprimés dans la population.

Déterminer l'intervalle de confiance de la moyenne et la varaince au niveau 95%.

3.4 Série de TD N^{0}3 (2015-2016)

Université 08Mai 1945 Guelma **Biostatistiques, 2015-2016**
3ème année Licence : Immunologie

Série 3

Exercice 1

Soient $X_1, X_2, ..., X_n$ n variables aléatoires indépendantes suivant une loi géométrique de paramètre $p \in [0, 1]$ définie par :

$$P(x_i, p) = p(1-p)^{x_i - 1}$$

Estimer le paramètre p de la loi, en utilisant la méthode de maximum de vraisemblance.

Exercice 2

Un dosage de sucre dans une solution effectué sur 8 prélèvement provenant d'une même population a donné les résultats suivants exprimés en g/l.

19.5 19.7 19.8 20.2 20.2 20.3 20.4 20.8

1-Déterminer les estimations ponctuelles de la moyenne et l'écart-type de cette distribution ?

2-Quel est l'intervalle de confiance de la moyenne au niveau 95% ?

-Quel est l'intervalle de confiance de la moyenne au niveau 99% ?

3-Quel est l'intervalle de confiance de la variance au niveau 95% ?

Exercice 3

Une expérience sur les *bêta-globulines* (*) a donné les résultats suivants :

Ci	6	8	10	12	14	16	18	20	22	24	26
n_i	2	6	13	17	17	38	10	17	6	5	2

1-Calculer la moyenne de l'échantillon et l'écart type de cette moyenne.

2-Déterminer l'intervalle de confiance de la moyenne au niveau 95%.

3-Déterminer l'intervalle de confiance de la moyenne au niveau 90%.

(*) : Les alphaglobulines et les gammaglobulines, les bêta-globulines composent les protéines du sérum.

Exercice 4

Le staff médical d'une grande entreprise fait ses petites statistiques sur le taux de cholestérol de ses employés ; les observations sur 100 employés tirés au sort sont les suivantes.

taux de cholestérol en cg :	effectif d'employés :
120	9
160	22
200	25
240	21
280	16
320	7

1. Estimer la moyenne et l'écart-type pour le taux de cholestérol dans toute l'entreprise.

2. Déterminer un intervalle de confiance pour la moyenne au niveau de confiance 95%.

3. Déterminer la taille minimum d'échantillon pour que l'amplitude de l'intervalle de confiance soit inférieure 10.

Chapitre 4

Tests statistiques

Soit une hypothèse H_0 concernant une population. Sur la base des résultats d'échantillons extraits de cette population on est amené à accepter ou rejeter l'hypothèse H_0. Les règles de décision sont appelées tests statistiques. H_0 désigne l'hypothèse dite hypothèse nulle et par H_1 on note l'hypothèse dite hypothèse alternative.

On a H_0 vraie et H_1 fausse ou bien H_0 fausse et H_1 vraie.

Tests d'homogénéité

A partir d'un échantillon de taille n_1 extrait d'une population P_1 et d'un échantillon de taille n_2 extrait d'une population P_2, le test permet de décider :

$$\begin{cases} H_0 : \theta_0 = \theta_1 \\ H_1 : \theta_0 \neq \theta_1 \end{cases}$$

où θ_0 et θ_1 sont les deux valeurs d'un même paramètre des deux populations P_1 et P_2.

4.1 Test de Student (comparaison de deux moyennes)

Soient X et Y deux variables aléatoires indépendants de moyennes m_1 et m_2 et d'écart-type σ_1 et σ_2. On dépose de deux échantillons indépendants $\{X_1; X_2; ...; X_{n_1}\}$ tel que X_i suit la même loi $N(m_1, \sigma_1)$ et $\{Y_1; Y_2; ...; Y_{n_2}\}$ tel que Y_i suit la même loi $N(m_2, \sigma_2)$. On cherche à décider si les moyennes m_1 et m_2 sont significativement différentes ou non, on utilise le test de Student :

a- Si $n_1 \geq 30$ **,** $n_2 \geq 30$ **et** σ_1 **,** σ_2 **sont connus.**

On teste au seuil de signification α

$$\begin{cases} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{cases}$$

-On accepte H_0 (c.à.d il n'ya pas différance significative entre les moyennes de deux échantillons) si

$$z \in]-u; u[$$

où $z = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ et la valeur u est lue dans la table normale centrée réduite $N(0,1)$ telle que $\Phi(u) = 1 - \frac{\alpha}{2}$.

-On rejette H_0 si $z \notin]-u; u[$ (Il ya une différance significative).

Remarque 1 Si σ_1 et σ_2 sont inconnues, on les remplace par les estimateurs

$\hat{S}_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2$ et $\hat{S}_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(y_j - \bar{y})^2$ respectivement, c.à.d

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}}$$

Exemple 1 Une machine remplit des paquets de café, on prélève un échantillon de paquets de taille $n_1 = 120$ de poids moyen 48.53 g et d'écart type 2.8 g, le lendemain on prélève un échantillon de taille $n_2 = 270$ de moyen 50.08 g et l'écart type 3.1 g.

Au seuil de signification 5% (risque d'erreur), qu'il existe une différance significative entre les poids moyens des paquets ?

Echantillon 1	Echantillon 2
$n_1 = 120$	$n_2 = 270$
$\bar{x} = 48.53$	$\bar{y} = 50.08$
$\sigma_1 = 2.8$	$\sigma_2 = 3.1$

Il s'agit du test $H_0 : m_1 = m_2$

$$\begin{aligned} z &= \frac{48.53 - 50.08}{\sqrt{\frac{(2.8)^2}{120} + \frac{(3.1)^2}{270}}} \\ &= -4.88 \end{aligned}$$

$$\begin{aligned}\Phi(u) &= 1-\frac{0.05}{2}\\ &= 0.975\end{aligned}$$

Dans la table $N(0,1)$, on trouve $u=1.96$, $z\notin[-1.96;1.96]$ donc on rejette H_0, il ya une différance significative entre les poids moyens des paquets.

b- si $n_1<30$, $n_2<30$ et σ_1 , σ_2 égaux et inconnus ($\sigma_1=\sigma_2=\sigma$)

-On accepte H_0 (c.à.d il n'ya pas différance significative entre les moyennes de deux échantillons) si

$$z\in\left]-t_{n_1+n_2-2,\frac{\alpha}{2}};t_{n_1+n_2-2,\frac{\alpha}{2}}\right[$$

où

$$z=\frac{\bar{x}-\bar{y}}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

avec

$$S=\sqrt{\frac{(n_1-1)\,\hat{S}_1^2+(n_2-1)\,\hat{S}_2^2}{n_1+n_2-2}}$$

et la valeur $t_{n_1+n_2-2,\frac{\alpha}{2}}$ *est lue dans* la table de Student à $k=n_1+n_2-2$ dégrées de liberté (ddl) et $\gamma=\frac{\alpha}{2}$.

-On rejette H_0 si $z\notin\left]-t_{n_1+n_2-2,\frac{\alpha}{2}};t_{n_1+n_2-2,\frac{\alpha}{2}}\right[$ (Il ya une différance significative).

Exemple 2 Le poids d'un médicament conditionné en boites est réparti suivant une loi normale $N(m,\sigma)$. Deux échantillons de tailles respectives $n_1=12$ et $n_2=18$ ont pour moyennes $\bar{x}=22.235\ g$ et $\bar{y}=21.988\ g$ et écart type (estimateur) $\hat{S}_1=0.18\ g$ et $\hat{S}_2=0.23\ g$

Qu'il existe une différance significative entre les poids moyens des deux échantillons pour un seuil de signification de 5% ?

Echantillon 1	Echantillon 2
$n_1=12$	$n_2=18$
$\bar{x}=22.235$	$\bar{y}=21.988$
$\hat{S}_1=0.18$	$\hat{S}_2=0.23$

Il s'agit du test $H_0: m_1=m_2$

$$S=\sqrt{\frac{(12-1)\,(0.18)^2+(18-1)\,(0.23)^2}{12+18-2}}=0.211\,77$$

donc

$$z = \frac{(22.235 - 21.988)}{0.211\,77 \times \sqrt{\frac{1}{12} + \frac{1}{18}}} = 3.\,129$$

Dans la table de loi de Student, on trouve

$$t_{n_1+n_2-2,\frac{\alpha}{2}} = t_{28,0.025} = 2.048,$$

$z \notin [-2.048; 2.048]$ donc on rejette H_0, il ya une différance significative entre les moyennes des deux échantillons.

4.2 Comparaison de deux proportions

Soient deux population P_1 et P_2, on extrait un échantillon de population P_1 de taille n_1 et on extrait un échantillon de taille n_2 dans la population P_2.

On compare deux proportions inconnues p_1 et p_2. On souhaite tester si ce sont les mêmes. L'hypothèse nulle à tester est H_0 : «$p_1 = p_2$»contre H_1 : «$p_1 \neq p_2$».

On dispose de deux séries d'observations, de taille n_1 pour p_1 qu'on estime par f_1 et de taille n_2 pour p_2 qu'on estime par f_2.

-On accepte H_0 (c.à.d on admet alors l'égalité des proportions) si

$$z \in]-u; u[$$

où

$$z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

avec

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

et la valeur u est lue dans la table normale centrée réduite $N(0,1)$ telle que $\Phi(u) = 1 - \frac{\alpha}{2}$.

-On rejette H_0 si $z \notin]-u; u[$ (Il ya une différance significative entre les proportions des deux échantillons).

Exemple 3 On expérimente un vaccin contre une maladie M sur des animaux. Un échantillon aléatoire de taille $n_1 = 80$ animaux vaccinés montre que 42 d'entre eaux ont contracté la maladie. Un échantillon aléatoire de taille

$n_2 = 113$ animaux non vaccinés montre que 76 d'entre eaux ont contacté la maladie.

Peut-on dire au seuil de signification de 5% que le vaccin est inefficace ?

On décide : $H_0 : p_1 = p_2$

$n_1 = 80, n_2 = 113, f_1 = \frac{42}{80}$ et $f_2 = \frac{76}{113}$, donc on a :

$$f = \frac{80\left(\frac{42}{80}\right) + 113\left(\frac{76}{113}\right)}{80 + 113} = 0.611$$

alors

$$z = \frac{\frac{42}{80} - \frac{76}{113}}{\sqrt{0.611\,(1 - 0.611\,)\left(\frac{1}{80} + \frac{1}{113}\right)}} = -2.\,071\,6$$

Dans la table $N(0, 1)$, on trouve $u = 1.96$, $z \notin [-1.96; 1.96]$ donc on rejette H_0, au seuil de signification de 5% la différance entre les proportions est significative.

4.3 Test de Fisher (comparaison de deux variances)

Soient X et Y deux variables aléatoires indépendants de moyennes m_1 et m_2 et d'écart-type σ_1 et σ_2. On dépose de deux échantillons indépendants $\{X_1; X_2; ...; X_{n_1}\}$ tel que X_i suit la même loi $N(m_1, \sigma_1)$ et $\{Y_1; Y_2; ...; Y_{n_2}\}$ tel que Y_i suit la même loi $N(m_1, \sigma_1)$. On cherche à décider si les variances σ_1^2 et σ_2^2 sont significativement différentes ou non, on utilise le test de Fisher :

On pose l'hypothèse $H_0 : \sigma_1 = \sigma_2$ (les deux populations ont la même variance) et

$$F = \begin{cases} \frac{\hat{S}_1^2}{\hat{S}_2^2} & \text{si } \hat{S}_1^2 > \hat{S}_2^2 \\ \frac{\hat{S}_2^2}{\hat{S}_1^2} & \text{si } \hat{S}_1^2 < \hat{S}_2^2 \end{cases}$$

où $\hat{S}_1^2$ est un estimateur de σ_1^2 et $\hat{S}_2^2$ est un estimateur de σ_2^2 c.à.d

$$\hat{S}_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_i - \bar{x})^2;\ \hat{S}_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(y_j - \bar{y})^2$$

Si $F < F^{\alpha}_{n_1-1,n_2-1}$ on accepte H_0 (on admet alors l'égalité des variances)

Si $F > F^{\alpha}_{n_1-1,n_2-1}$ on rejette H_0 (il ya différance significative entre les variances des deux échantillons), avec la valeur $F^{\alpha}_{n_1-1,n_2-1}$ est lue dans la table de Fisher au risque d'erreur α et à $n_1 - 1$ et $n_2 - 1$ dégrées de liberté (ddl).

Exemple 4 Reprenons les données des 2 échantillons

Ech 1	7	18	9	9	18	27	12	10	32	6	37
Ech 2	12	15	14	16	22	17	25	9	18	/	/

Qu'il existe une différance significative entre les variances des deux échantillons pour un seuil de signification de 5%.

On pose l'hypothèse $H_0 : \sigma_1 = \sigma_2$

Ech 1	Ech 2
$n_1 = 11$	$n_2 = 9$
$\bar{x} = 16.82$	$\bar{y} = 16.44$
$S_1^2 = 114.96$	$S_2^2 = 23.78$

donc on a

$$F = \frac{S_1^2}{S_2^2} = \frac{114.96}{23.78} = 4.834$$

Dans la table de Fisher, on trouve :

$$F^{0.05}_{10,8} = 3.347$$

$F > F^{0.05}_{10,8}$ donc on rejette H_0 , il ya une différance significative entre les variances de deux échantillons.

4.4 Les Tests du Khi-deux

On peut distinguer trois types de test du Khi-deux χ^2 :

· Le test du χ^2 d'adéquation (H_0 : « le caractère X suit-il une loi particulière ? »),

· Le test du χ^2 d'homogénéité (H_0 : « le caractère X suit-il la meme loi dans deux populations données ? ») ,

· Le test du χ^2 d'indépendance (H_0 : « les caractères X et Y sont-ils indépendants ? »).

Ces trois tests ont un principe commun qui est le suivant : on répartit les observations dans k classes dont les effctifs sont notés $n_1 = N_1(w), ..., n_k =$

$N_k(w)$. L'hypothèse H_0 permet de calculer les effectifs théoriques, notés $n_{1,th}, ..., n_{k,th}$. On rejette H_0 si les effectifs observés sont trop differents des effectifs théoriques.

On accepte H_0 si

$$h \notin \left]\chi_{k-1-m,\alpha}; +\infty\right[$$

où

$$h = \sum_{i=1}^{k} \frac{(n_i - n_{i,th})^2}{n_{i,th}}$$

où la valeurs $\chi_{k-1-m,\alpha}$ est leu dans la table du Khi-deux avec $(k-1-m)$ dégrées de liberté (ddl)$(\gamma = \alpha)$ avec k est le nombre de classes et m est le nombre de paramètres estimés nécessaires au calcul des effectifs théoriques.

On rejette H_0 si

$$h \in \left]\chi_{k-1-m,\alpha}; +\infty\right[$$

Exemple 5 Un croisement entre roses rouges et blanches a donné en seconde génération des roses rouges, roses et blanches. Sur un échantillon de taille 600, on a troué les résultats suivants :

couleur	effectifs
rouges	141
roses	315
blanches	144

Peut-on affirmer que les résultats sont conformes aux lois de *Mendel* ?

Il s'agit donc de tester

$H_0 : p_{rouges} = 0.25, p_{roses} = 0.5, p_{blanches} = 0.25$ au risque disons $\alpha = 0,05$.

On dresse alors le tableau suivant :

couleur	effectifs observés n_i	effectifs théoriques $n_{i,th}$
rouges	141	$0.25 \times 600 = 150$
roses	315	$0.5 \times 600 = 300$
blanches	144	$0.25 \times 600 = 150$

Ici on a $k = 3$ classes et $m = 0$ (aucun paramètre a estimer pour pouvoir calculer les effectifs théoriques) donc $k - 1 - m = 2$; on calcule ensuite $]\chi^2_{k-1-m,\alpha}; +\infty[$ à l'aide de la table du Khi-deux et on obtient $\chi^2_{2,0.05} = 5.99$. Enffin, on calcule

$$\begin{aligned} h &= \sum_{i=1}^{k} \frac{(n_i - n_{i,th})^2}{n_{i,th}} \\ &= \frac{(141-150)^2}{150} + \frac{(315-300)^2}{300} + \frac{(144-150)^2}{150} \\ &= 1.53 \end{aligned}$$

donc $h \notin]5.99; +\infty[$.

On ne rejette pas H_0 au risque d'erreur $\alpha = 0,05$ (On accepte H_0), on ne peut pas dire que les observations contredisent la loi de *Mendel*.

4.5 Test de Kruskal-wallis (Test sur échantillons indépendants)

Le test de Kruskal-Wallis est un test à utiliser lorsque vous êtes en présence de k échantillons indépendant, afin de déterminer si les échantillons proviennent d'une même population ou si au mois un échantillon provient d'une population différente des autres. Il permet de tester si k échantillons ($k > 2$) proviennent de la même population, ou de population ayant des caractéristiques identiques, au sens d'un paramètre de position.

Principe du test de Kruskal-wallis

Si on désigne par M_i le paramètre de position l'échantillon i, les hypothèses nulle H_0 et alternative H_1 du test de Kruskal-wallis sont les suivantes :

-H_0 : $M_1 = M_2 = ... = M_k$

-H_1 : il existe au moins un couple (i, j) tel que $M_i \neq M_j$

1/Classer les données sous forme de tableau

Noter l'effectif de chaque série

Exemple pratique :

On veut comparer 3 milieux de culture différents A, B et C, pour cela on compte le nombre de colonies bactériennes dans chaque milieu sur plusieurs jours.

Milieu	$J1$	$J2$	$J3$	$J4$	$J5$	$J6$
A	7	4	3	2	4	–
B	5	4	4	1	3	5
C	6	7	6	5	7	6

$n_a = 5, n_b = 6, n_c = 6$

2/Ranger les données en fonction de leur fréquence dans chaque série

Dans notre série

Nombre de colonies	1	2	3	4	5	6	7
Fréquence dans A	0	1	1	2	0	0	1
Fréquence dans B	1	0	1	2	2	0	0
Fréquence dans C	0	0	0	0	1	3	2

3/Calculer la somme des fréquences

Dans notre exemple :

Nombre de colonies	1	2	3	4	5	6	7
Fréquence dans A	0	1	1	2	0	0	1
Fréquence dans B	1	0	1	2	2	0	0
Fréquence dans C	0	0	0	0	1	3	2
Somme des fréquences	1	1	2	4	3	3	3

4/Classer les données en rang par ordre

Dans notre exemple :

Nombre de colonies	1	2	3	4	5	6	7
Fréquence dans A	0	1	1	2	0	0	1
Fréquence dans B	1	0	1	2	2	0	0
Fréquence dans C	0	0	0	0	1	3	2
Somme des fréquences	1	1	2	4	3	3	3
RANG	1	1-2	3-4	5-6-7-8	9-10-11	12-13-14	15-16-17

5/Calculer le rang corrigé qui est la moyenne des rangs pour chaque fréquence Rc

Dans notre exemple :

RANG	1	1-2	3-4	5-6-7-8	9-10-11	12-13-14	15-16-17
Rang corrigé	1	2	3.5	6.5	10	13	16

6/Calculer les fréquences corrigées fc

$fc = f.Rc$

Dans notre exemple :

Rang corrigé	1	2	3.5	6.5	10	13	16
Fc pour A	0	2	3.5	13	0	0	16
Fc pour B	1	0	3.5	13	20	0	0
Fc pour C	0	0	0	0	10	39	32

7/Calculer le total des rangs : $R_i = \sum Fc$

Dans notre exemple :

$R_a = 0 + 2 + 3.5 + 13 + 0 + 0 + 16 = 34.5$

$R_b = 1 + 0 + 3.5 + 13 + 20 + 0 + 0 = 37.5$

$R_c = 0 + 0 + 0 + 0 + 10 + 39 + 32 = 81$

8/Calcul de H

$$H = \frac{12}{N \times (N+1)} \sum \frac{R_i}{n_i} - 3\,(N+1)$$

N étant l'effectif total

R_i étant le total des rangs corriges

n_i étant l'effectif de chaque série

Dans notre exemple :

$n_a = 5; R_a = 34.5; n_b = 6; R_b = 37.5; n_c = 6; R_c = 81$

$$\begin{aligned} H &= \left(\frac{12}{17 \times 18}\right) \times \left(\frac{(34.5)^2}{5} + \frac{(37.5)^2}{6} + \frac{(81)^2}{6}\right) - 3 \times 18 \\ &= 7.40 \end{aligned}$$

9/ Comparer H avec la valeur du χ^2 pour $(k-1)$ degré de liberté k étant le nombre d'échantillons

Si H est supérieur au χ^2 de la table, il existe donc une différence significative entre les séries

Si H est inférieur au de χ^2 la table, il n'existe pas de différence significative entre les séries

Dans notre exemple :

Pour$k - 1 = 2$, degré de liberté la table du χ^2 montre 5.99 et $H = 7.418$

Donc H est supérieur à la valeur du χ^2 lue, il existe ainsi une différence significative entre les 3 milieux de culture.

4.6 Exercices sur le chapitre 4

Exercice 1

Le pH (degré d'acidité) a été mesuré dans deux types de solutions chimiques A et B. Dans la solution A, 6 mesures ont été faites, avec un PH moyen de $7,52$ et un écart-type estimé de $S_1 = 0,024$. Dans la solution B, 5 mesures ont été faites, avec un PH moyen de $7,49$ et un écart-type estimé de $S_2 = 0,032$.

Déterminer si, au seuil de signification de 5%(risque) , les deux solutions ont des PH différents (l'hypothèse selon laquelle les moyennes et les variances des dégrées d'acidités).

Valeur théorique : $F_{4,5}^{0.05} = 5.192$.

Exercice 2

Les QI de 9 enfants d'un quartier d'une grande ville ont pour moyenne empirique 107 et écart-type empirique 10. Les QI de 12 enfants d'un autre quartier ont pour moyenne empirique 112 et écart-type empirique 9.

-Tester l'égalité des variances au seuil de 5%.

Exercice 3

Les tensions maximales des muscles gastrocnémiens (exprimées en g) de la grenouille varient selon que ces muscles sont normaux ou dénervés. Lors d'une expérience faite sur 10 grenouilles, on a relevé les mesures suivantes :

Muscles normaux	75	96	32	41	50	39	59	45	30
Muscles dénervés	53	67	32	29	35	27	37	30	21

1. Préciser les hypothèses de modélisation.
2. Tester l'hypothèse d'égalité des variances au seuil de 5%.

Exercice 4

Le pourcentage des femmes de 35 ans présentant des rides est de 25%. Sur 200 femmes de 35 ans ayant suivi un traitement antirides, on a observé que 40 avaient des rides. Au risque de 5%, peut-on dire que le traitement est efficace ?

Exercice 5

Pour une certaine maladie, on dispose d'un traitement satisfaisant dans 70% des cas. Un laboratoire propose un nouveau traitement et affirme qu'il donne satisfaction plus souvent que l'ancien traitement. Sur 200 malades ayant suivi ce nouveau traitement, on a observé une guérison pour 148 d'entre eux. En tant qu'expert chargé d'autoriser la mise sur le marché de ce nouveau traitement, que concluez-vous ?

4.7 Série de TD N^{0}4 (2015-2016)

Université 08Mai 1945 Guelma **Biostatistiques, 2015-2016**
3ème année Licence : Immunologie

Série 4

Exercice 1

Dans une maternité pour deux échantillons de nouveau-nés de sexes différents on a obtenu les résultats suivants :

51 garçons : taille moyenne 51 cm et écart-type des tailles $3cm$

59 filles : taille moyenne 49 cm et écart-type des tailles $3.2cm$

Au risque de 5% peut-on déduire de ces indications une différence significative entre les moyennes des tailles des nouveau-nés suivant le sexe ?

Exercice 2

Le tableau suivant présente les taux de calcium chez des malades atteints d'une insuffisance rénale chronique

Malade	Echantillon 1	Echantillon 2
1	50	80
2	25	130
3	120	70
4	26	120
5	48	70
6	113	100
7	150	/

-Peut-on admettre au seuil de signification de 5% (risque d'érreur), qu'il existe une défférance significative entre les moyennes de deux échantillons ?

Exercice 3

On veut tester la signigication des différences des valeurs de la teneur en protéines totales du grain de blé lorsqu'une des protéines : $SG-HPM1$ ou $SG-HPM2$ est présente.

Les résultats de l'expérience sont résumés dans le tableau suivant :

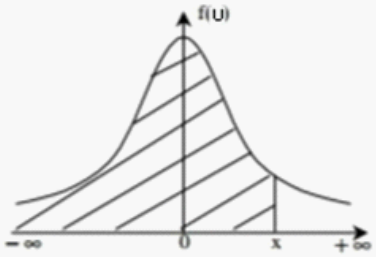
Ech1($SG-HPM1$)	15.63	13.86	16.85	15.88	17.46	15.40	13.40	17.08	17.47
Ech2($SG-HPM2$)	14.91	15.57	13.71	14.57	16.37	11.70	16.23	15.87	13.14

-Peut-on admettre au seuil de signification de 5% (risque d'érreur). Tester l'égalité des moyennes et des varainces ?

4.8 Tables statistiques

Loi Normale centrée réduite

Probabilité de trouver une valeur inférieure à u



u	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	0,9496	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
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Table pour les grandes valeurs de u :

u	3	3,2	3,4	3,6	3,8	4	4,2	4,4	4,6	4,8
F	0,99865003	0,99931280	0,99966302	0,99984085	0,99992763	0,99996831	0,99998665	0,99999458	0,99999789	0,99999920

LOI DE STUDENT AVEC k DEGRÉS DE LIBERTÉ
QUANTILES D'ORDRE $1 - \gamma$

	γ										
k	0.25	0.20	0.15	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

LOI DU KHI-DEUX AVEC k DEGRÉS DE LIBERTÉ
QUANTILES D'ORDRE $1-\gamma$

	γ										
k	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.94	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.81	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

Si k est entre 30 et 100 mais n'est pas un multiple de 10, on utilise la table ci-haut et on fait une interpolation linéaire. Si $k > 100$ on peut, grâce au théorème limite central, approximer la loi $\chi^2(k)$ par la loi $N(k, 2k)$.

QUANTILES D'ORDRE 0.95 DE LA LOI DE FISHER

Degrés de liberté du numérateur sur la première ligne
Degrés de liberté du dénominateur sur la colonne de gauche

	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993
70	3.978	3.128	2.736	2.503	2.346	2.231	2.143	2.074	2.017	1.969
80	3.960	3.111	2.719	2.486	2.329	2.214	2.126	2.056	1.999	1.951
90	3.947	3.098	2.706	2.473	2.316	2.201	2.113	2.043	1.986	1.938
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927
150	3.904	3.056	2.665	2.432	2.274	2.160	2.071	2.001	1.943	1.894
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878
400	3.865	3.018	2.627	2.394	2.237	2.121	2.032	1.962	1.903	1.854

QUANTILES D'ORDRE 0.95 DE LA LOI DE FISHER

Degrés de liberté du numérateur sur la première ligne
Degrés de liberté du dénominateur sur la colonne de gauche

	11	12	13	14	15	16	17	18	19	20
1	243.0	243.9	244.7	245.4	245.9	246.5	246.9	247.3	247.7	248.0
2	19.40	19.41	19.42	19.42	19.43	19.43	19.44	19.44	19.44	19.45
3	8.763	8.745	8.729	8.715	8.703	8.692	8.683	8.675	8.667	8.660
4	5.936	5.912	5.891	5.873	5.858	5.844	5.832	5.821	5.811	5.803
5	4.704	4.678	4.655	4.636	4.619	4.604	4.590	4.579	4.568	4.558
6	4.027	4.000	3.976	3.956	3.938	3.922	3.908	3.896	3.884	3.874
7	3.603	3.575	3.550	3.529	3.511	3.494	3.480	3.467	3.455	3.445
8	3.313	3.284	3.259	3.237	3.218	3.202	3.187	3.173	3.161	3.150
9	3.102	3.073	3.048	3.025	3.006	2.989	2.974	2.960	2.948	2.936
10	2.943	2.913	2.887	2.865	2.845	2.828	2.812	2.798	2.785	2.774
11	2.818	2.788	2.761	2.739	2.719	2.701	2.685	2.671	2.658	2.646
12	2.717	2.687	2.660	2.637	2.617	2.599	2.583	2.568	2.555	2.544
13	2.635	2.604	2.577	2.554	2.533	2.515	2.499	2.484	2.471	2.459
14	2.565	2.534	2.507	2.484	2.463	2.445	2.428	2.413	2.400	2.388
15	2.507	2.475	2.448	2.424	2.403	2.385	2.368	2.353	2.340	2.328
16	2.456	2.425	2.397	2.373	2.352	2.333	2.317	2.302	2.288	2.276
17	2.413	2.381	2.353	2.329	2.308	2.289	2.272	2.257	2.243	2.230
18	2.374	2.342	2.314	2.290	2.269	2.250	2.233	2.217	2.203	2.191
19	2.340	2.308	2.280	2.256	2.234	2.215	2.198	2.182	2.168	2.155
20	2.310	2.278	2.250	2.225	2.203	2.184	2.167	2.151	2.137	2.124
21	2.283	2.250	2.222	2.197	2.176	2.156	2.139	2.123	2.109	2.096
22	2.259	2.226	2.198	2.173	2.151	2.131	2.114	2.098	2.084	2.071
23	2.236	2.204	2.175	2.150	2.128	2.109	2.091	2.075	2.061	2.048
24	2.216	2.183	2.155	2.130	2.108	2.088	2.070	2.054	2.040	2.027
25	2.198	2.165	2.136	2.111	2.089	2.069	2.051	2.035	2.021	2.007
26	2.181	2.148	2.119	2.094	2.072	2.052	2.034	2.018	2.003	1.990
27	2.166	2.132	2.103	2.078	2.056	2.036	2.018	2.002	1.987	1.974
28	2.151	2.118	2.089	2.064	2.041	2.021	2.003	1.987	1.972	1.959
29	2.138	2.104	2.075	2.050	2.027	2.007	1.989	1.973	1.958	1.945
30	2.126	2.092	2.063	2.037	2.015	1.995	1.976	1.960	1.945	1.932
40	2.038	2.003	1.974	1.948	1.924	1.904	1.885	1.868	1.853	1.839
50	1.986	1.952	1.921	1.895	1.871	1.850	1.831	1.814	1.798	1.784
60	1.952	1.917	1.887	1.860	1.836	1.815	1.796	1.778	1.763	1.748
70	1.928	1.893	1.863	1.836	1.812	1.790	1.771	1.753	1.737	1.722
80	1.910	1.875	1.845	1.817	1.793	1.772	1.752	1.734	1.718	1.703
90	1.897	1.861	1.830	1.803	1.779	1.757	1.737	1.720	1.703	1.688
100	1.886	1.850	1.819	1.792	1.768	1.746	1.726	1.708	1.691	1.676
150	1.853	1.817	1.786	1.758	1.734	1.711	1.691	1.673	1.656	1.641
200	1.837	1.801	1.769	1.742	1.717	1.694	1.674	1.656	1.639	1.623
400	1.813	1.776	1.745	1.717	1.691	1.669	1.648	1.630	1.613	1.597

QUANTILES D'ORDRE 0.95 DE LA LOI DE FISHER

Degrés de liberté du numérateur sur la première ligne
Degrés de liberté du dénominateur sur la colonne de gauche

	21	22	23	24	25	26	27	28	29	30
1	248.3	248.6	248.8	249.1	249.3	249.5	249.6	249.8	250.0	250.1
2	19.45	19.45	19.45	19.45	19.46	19.46	19.46	19.46	19.46	19.46
3	8.654	8.648	8.643	8.639	8.634	8.630	8.626	8.623	8.620	8.617
4	5.795	5.787	5.781	5.774	5.769	5.763	5.759	5.754	5.750	5.746
5	4.549	4.541	4.534	4.527	4.521	4.515	4.510	4.505	4.500	4.496
6	3.865	3.856	3.849	3.841	3.835	3.829	3.823	3.818	3.813	3.808
7	3.435	3.426	3.418	3.410	3.404	3.397	3.391	3.386	3.381	3.376
8	3.140	3.131	3.123	3.115	3.108	3.102	3.095	3.090	3.084	3.079
9	2.926	2.917	2.908	2.900	2.893	2.886	2.880	2.874	2.869	2.864
10	2.764	2.754	2.745	2.737	2.730	2.723	2.716	2.710	2.705	2.700
11	2.636	2.626	2.617	2.609	2.601	2.594	2.588	2.582	2.576	2.570
12	2.533	2.523	2.514	2.505	2.498	2.491	2.484	2.478	2.472	2.466
13	2.448	2.438	2.429	2.420	2.412	2.405	2.398	2.392	2.386	2.380
14	2.377	2.367	2.357	2.349	2.341	2.333	2.326	2.320	2.314	2.308
15	2.316	2.306	2.297	2.288	2.280	2.272	2.265	2.259	2.253	2.247
16	2.264	2.254	2.244	2.235	2.227	2.220	2.212	2.206	2.200	2.194
17	2.219	2.208	2.199	2.190	2.181	2.174	2.167	2.160	2.154	2.148
18	2.179	2.168	2.159	2.150	2.141	2.134	2.126	2.119	2.113	2.107
19	2.144	2.133	2.123	2.114	2.106	2.098	2.090	2.084	2.077	2.071
20	2.112	2.102	2.092	2.082	2.074	2.066	2.059	2.052	2.045	2.039
21	2.084	2.073	2.063	2.054	2.045	2.037	2.030	2.023	2.016	2.010
22	2.059	2.048	2.038	2.028	2.020	2.012	2.004	1.997	1.990	1.984
23	2.036	2.025	2.014	2.005	1.996	1.988	1.981	1.973	1.967	1.961
24	2.015	2.003	1.993	1.984	1.975	1.967	1.959	1.952	1.945	1.939
25	1.995	1.984	1.974	1.964	1.955	1.947	1.939	1.932	1.926	1.919
26	1.978	1.966	1.956	1.946	1.938	1.929	1.921	1.914	1.907	1.901
27	1.961	1.950	1.940	1.930	1.921	1.913	1.905	1.898	1.891	1.884
28	1.946	1.935	1.924	1.915	1.906	1.897	1.889	1.882	1.875	1.869
29	1.932	1.921	1.910	1.901	1.891	1.883	1.875	1.868	1.861	1.854
30	1.919	1.908	1.897	1.887	1.878	1.870	1.862	1.854	1.847	1.841
40	1.826	1.814	1.803	1.793	1.783	1.775	1.766	1.759	1.751	1.744
50	1.771	1.759	1.748	1.737	1.727	1.718	1.710	1.702	1.694	1.687
60	1.735	1.722	1.711	1.700	1.690	1.681	1.672	1.664	1.656	1.649
70	1.709	1.696	1.685	1.674	1.664	1.654	1.646	1.637	1.629	1.622
80	1.689	1.677	1.665	1.654	1.644	1.634	1.626	1.617	1.609	1.602
90	1.675	1.662	1.650	1.639	1.629	1.619	1.610	1.601	1.593	1.586
100	1.663	1.650	1.638	1.627	1.616	1.607	1.598	1.589	1.581	1.573
150	1.627	1.614	1.602	1.590	1.580	1.570	1.560	1.552	1.543	1.535
200	1.609	1.596	1.583	1.572	1.561	1.551	1.542	1.533	1.524	1.516
400	1.582	1.569	1.556	1.545	1.534	1.523	1.514	1.505	1.496	1.488

QUANTILES D'ORDRE 0.95 DE LA LOI DE FISHER

Degrés de liberté du numérateur sur la première ligne
Degrés de liberté du dénominateur sur la colonne de gauche

	40	50	60	70	80	90	100	150	200	400
1	251.1	251.8	252.2	252.5	252.7	252.9	253.0	253.5	253.7	253.8
2	19.47	19.48	19.48	19.48	19.48	19.48	19.49	19.49	19.49	19.49
3	8.594	8.581	8.572	8.566	8.561	8.557	8.554	8.545	8.540	8.537
4	5.717	5.699	5.688	5.679	5.673	5.668	5.664	5.652	5.646	5.643
5	4.464	4.444	4.431	4.422	4.415	4.409	4.405	4.392	4.385	4.381
6	3.774	3.754	3.740	3.730	3.722	3.716	3.712	3.698	3.690	3.686
7	3.340	3.319	3.304	3.294	3.286	3.280	3.275	3.260	3.252	3.248
8	3.043	3.020	3.005	2.994	2.986	2.980	2.975	2.959	2.951	2.947
9	2.826	2.803	2.787	2.776	2.768	2.761	2.756	2.739	2.731	2.726
10	2.661	2.637	2.621	2.610	2.601	2.594	2.588	2.572	2.563	2.558
11	2.531	2.507	2.490	2.478	2.469	2.462	2.457	2.439	2.431	2.426
12	2.426	2.401	2.384	2.372	2.363	2.356	2.350	2.332	2.323	2.318
13	2.339	2.314	2.297	2.284	2.275	2.267	2.261	2.243	2.234	2.229
14	2.266	2.241	2.223	2.210	2.201	2.193	2.187	2.169	2.159	2.154
15	2.204	2.178	2.160	2.147	2.137	2.130	2.123	2.105	2.095	2.089
16	2.151	2.124	2.106	2.093	2.083	2.075	2.068	2.049	2.039	2.034
17	2.104	2.077	2.058	2.045	2.035	2.027	2.020	2.001	1.991	1.985
18	2.063	2.035	2.017	2.003	1.993	1.985	1.978	1.958	1.948	1.942
19	2.026	1.999	1.980	1.966	1.955	1.947	1.940	1.920	1.910	1.903
20	1.994	1.966	1.946	1.932	1.922	1.913	1.907	1.886	1.875	1.869
21	1.965	1.936	1.916	1.902	1.891	1.883	1.876	1.855	1.845	1.838
22	1.938	1.909	1.889	1.875	1.864	1.856	1.849	1.827	1.817	1.810
23	1.914	1.885	1.865	1.850	1.839	1.830	1.823	1.802	1.791	1.784
24	1.892	1.863	1.842	1.828	1.816	1.808	1.800	1.779	1.768	1.761
25	1.872	1.842	1.822	1.807	1.796	1.787	1.779	1.757	1.746	1.739
26	1.853	1.823	1.803	1.788	1.776	1.767	1.760	1.738	1.726	1.719
27	1.836	1.806	1.785	1.770	1.758	1.749	1.742	1.719	1.708	1.701
28	1.820	1.790	1.769	1.754	1.742	1.733	1.725	1.702	1.691	1.683
29	1.806	1.775	1.754	1.738	1.726	1.717	1.710	1.686	1.675	1.667
30	1.792	1.761	1.740	1.724	1.712	1.703	1.695	1.672	1.660	1.652
40	1.693	1.660	1.637	1.621	1.608	1.597	1.589	1.564	1.551	1.542
50	1.634	1.599	1.576	1.558	1.544	1.534	1.525	1.498	1.484	1.475
60	1.594	1.559	1.534	1.516	1.502	1.491	1.481	1.453	1.438	1.428
70	1.566	1.530	1.505	1.486	1.471	1.459	1.450	1.420	1.404	1.394
80	1.545	1.508	1.482	1.463	1.448	1.436	1.426	1.395	1.379	1.368
90	1.528	1.491	1.465	1.445	1.429	1.417	1.407	1.375	1.358	1.348
100	1.515	1.477	1.450	1.430	1.415	1.402	1.392	1.359	1.342	1.331
150	1.475	1.436	1.407	1.386	1.369	1.356	1.345	1.309	1.290	1.278
200	1.455	1.415	1.386	1.364	1.346	1.332	1.321	1.283	1.263	1.249
400	1.425	1.383	1.352	1.329	1.311	1.296	1.283	1.242	1.219	1.204

Bibliographie

[1] Bernard.Ycart ; Méthodes Statistiques pour la Biologie ; *Université Joseph Fourier, Grenoble I.*

[2] Carrat.F, Mallet. A, Morice .V ; Biostatistique ; *Université Pierre et Marie Curie.* 2013 – 2014.

[3] Dagnelie. P ; Statistique théorique et appliquée ; *Tome* 1 *et* 2. *Ed, Université Larcier et De-Boeck, Belgique.*2009.

[4] Gaetan Morin ; Biostatistique ; *Tome* 1 *et* 2. $2^{\text{ème}}$ *Ed. Scherrer, Canada.B.* 2009

[5] Gilbert Demengel ; Probabilités statistique inférentielle fiabilité ; *Premier cycle, IUP, Prépa, BTS, IUT.* 2007.

[6] Harvey.J ; Biostate, une approche intuitive. *Ed. Univ. De Boeck et Larcier ; Motulsky. .Belgique.*1995.

[7] Huguier.M ; Biostatistique au quotidien ; *Ed. Elsevier.A.* 2003.

[8] Jean-Christophe Breton ; Statistiques ; *Université de La Rochelle. Octobre-Novembre* 2008.

[9] Jean-Jacques Ruch ; Statistique : Estimation ; *Préparation à l'Agrégation Bordeaux 1. Année* 2012 – 2013

[10] Khalidi Khaled ; Méthodes statistiques ; *Rappels et cours ; Office des publication universitaires 1, Place centrale de Ben-Aknoun (Alger).*1998.

[11] Nakache.J.P ; Statistique explicative appliquée ; *Ed. Technip, France.J.* 2003.