KHALDI KHALED

METHODES STATISTIQUES

Rappels de cours
Exercices corrigés

Filières
Des Sciences Exactes

Office Des Publications Universitaires

KHALDI KHALED

METHODES STATISTIQUES

Rappels de cours - Exercices corrigés

Réimpression 2005

OFFICE DES PUBLICATIONS UNIVERSITAIRES
1, Place centrale de Ben-Aknoun (Alger)

EDITION: 1.01.4043
I.S.B.N : 9961.0.0918.5
Dépôt légal : 3168/2005

Avant propos

L'outil statistique trouve une place de plus en plus importante dans de nombreux domaines tels que : Sciences de l'ingénieur ; sciences physiques , économiques , biologiques , sociales ; agronomie ; médecine ;...

Notre souci , dans ce travail , a été de donner au lecteur les moyens d'aborder et de résoudre différents problèmes statistiques ; mais d'éviter de donner des recettes .

Nous nous sommes efforcés dans cet ouvrage de présenter les concepts et principales méthodes de la statistique et d'incorporer un nombre relativement important d'exercices . Par souci pédagogique , ceux-ci sont entièrement résolus avec parfois beaucoup de détails .

Ce recueil d'exercices est destiné à être utilisé comme complément à un cours de base de statistique mathématique . Outre la présentation des méthodes statistiques fondamentales, objectif principal de l'ouvrage , nous avons consacré une partie aux fondements théoriques des principales lois de probabiltés utilisées dans différents domaines statistiques (estimation ; théorie de la décision) ainsi qu'une autre à la statistique descriptive .

La plupart des exercices de ce recueil ont déjà été proposés aux étudiants dans le cadre de séances de travaux dirigés . Certains d'entre eux sont inspirés de manuels cités en référence . Ils correspondent dans la plupart des cas à des données réelles .

Nous exprimons notre reconnaissance aux collègues enseignants M^{me} KHALDI.Y , M^{lle} TAHI.R et M^r DOUMAZ.M. pour avoir bien voulu se charger de la correction et pour leurs précieux conseils qu'ils nous adressés . Nos remerciements vont également à M^r GOUIGAH.M pour la mise en forme de l'ouvrage . Nous espérons que le lecteur nous fera part de ses suggestions et remarques .

TABLE DES MATIERES

CHAPITRE 1

STATISTIQUE DESCRIPTIVE

I-SÉRIES STATISTIQUES À UNE DIMENSION

1-Définitions -Tableaux statistiques

Un ensemble fini Ω est dit population . Les éléments de Ω sont appelés individus . Une application de Ω dans R est dite 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 $\sum_i n_i = n$ et $f_i = \frac{n_i}{n}$ La fréquence correspondante .

Un tableau statistique est présenté sous la forme :

Tableau 1.1

x_i	n_i	f_i
x_1	n_1	f_1
x_2	n_2	f_2
.	.	.
.	.	.
.	.	.
.	.	.
x_k	n_k	f_k
Total	n	1

b) Représentation graphique

Diagramme en bâtons.

On porte $f_i(ou \quad n_i)$ en ordonnée en fonction de x_i

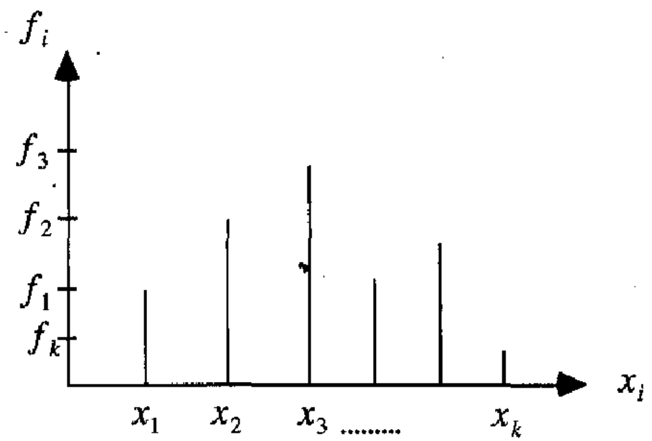


Fig. 1.1

Variables continues - Données groupées

a) Tableau

Les valeurs de la variable aléatoire X sont regroupées en classes $[c_i, c_{i+1}[$, $(i = 1, 2 \ldots, k-1)$. Un centre de classe x_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 $F_i = \sum_{j=1}^{i} f_j$ les fréquences cumulées

On représente ces données sous la forme du tableau

Tableau 1.2

Classes	x_i	n_i	f_i	F_i
c_1..........	....	....	....	....
	x_1	n_1	f_1	
c_2..........	....	....	....	... F_1.....
	x_2	n_2	f_2	
c_3..........	....	....	....	..... F_2....
.				
.				
.				
c_{k-1}......	....	....	....	... F_{k-2}...
.	x_{k-1}	n_{k-1}	f_{k-1}	
c_k......	....	....	....	... F_{k-1}...

L'amplitude de la classe i est $a_i = c_{i+1} - c_i$.
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

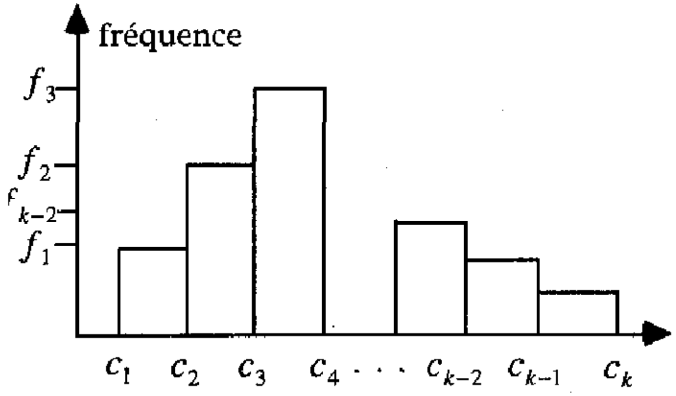


Fig. 1.2

2-Caractéristiques de position

Mode :

La valeur que la variable prend le plus fréquemment est dite mode .Pour une distribution discrète , c'est la valeur la plus fréquente . Dans le cas de données groupées en classe , on parle d'une classe modale.

Médiane

La médiane (notée M) est la valeur de la variable telle que

$F(M)=\frac{1}{2}$, $\left(F_i=\sum_{j=1}^{i} f_j \;,\; f_j=\frac{n_j}{n}\right)$ Elle déterminée par interpolation linéaire pour les données groupées . C'est une caractéristique qui n'est pas affectée par les valeurs extrêmes ; mais les observations doivent être ordonnées pour sa détermination .

Moyenne :

La moyenne (notée $\bar{x}$; $\bar{x}=\frac{\sum_i x_i}{n}$) est une caractéristique qui prend compte de toutes les données observées . Elle est utilisée dans beaucoup d'applications (estimations , tests statistiques) et est sensible aux valeurs extrêmes .

Pour une distribution en classe on a : $\bar{x}=\sum_i f_i x_i=\frac{\sum_i n_i x_i}{n}$ où

x_i est le centre de la classe i

On simplifie parfois le calcul en faisant un changement de variable . On introduit une variable :

$x_i'=\frac{x_i-x_o}{a}$ avec x_o et a convenablement choisis.

La moyenne $\bar{x}$ des x_i est alors calculée à partir de celle $\bar{x}'$ des x_i' puis on fait $\bar{x}=a\bar{x}'+x_0$

3-Caractéristique de dispersion

Étendue : $e=x_{max}-x_{min}$

Quartiles : Les nombres Q_1,Q_2 et Q_3 tels que $F(Q_1)=0{,}25$; $F(Q_2)=0{,}5$; $F(Q_3)=0{,}75$ sont dits quartiles .
Q_3-Q_1 est dit écart interquartile.

Déciles: Les nombres D_1,D_2,.........,D_9 tels que $F(D_1)=0{,}1$;
$F(D_2)=0{,}2$;........; $F(D_9)=0{,}9$ sont dits déciles.

Les centiles sont définis de la mêmes manière .

Variance et écart-type : On appelle variance le nombre réel positif ou nul

$$V = \sum_i f_i\left(x_i - \bar{x}\right)^2 = \sum_i f_i x_i^2 - \left(\sum_i f_i x_i\right)^2$$
$$= \sum_i f_i x_i^2 - \bar{x}^2$$

On appelle écart-type ou écart quadratique le nombre réel $\sigma = \sqrt{V}$

Moment d'ordre n $(n \in N)$: $m_n = \sum_i f_i x_i^n$

Moment centré d'ordre n

$$(n \in N) : \ \mu_n = \sum_i f_i\left(x_i - \bar{x}\right)^n$$

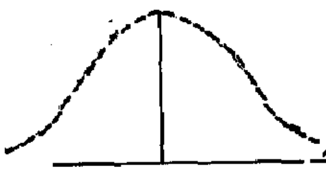
4-Caractéristiques de forme .

Cœfficient d'asymétrie : $\gamma = \mu_3 / \sigma^3$

Cœfficient d'aplatissement : $\delta = \mu_4 / \sigma^4$

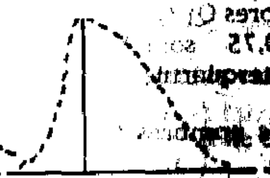
Lorsque la statistique est symétrique on a $\gamma = 0$

Pour une distribution normale réduite , $\delta - 3 = 0$



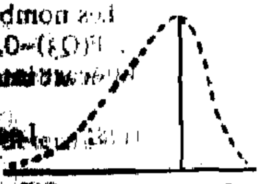
Courbe symétrique $\gamma = 0$

Fig. 1.3



$\gamma > 0$

Fig. 1.4



$\gamma < 0$

Fig. 1.5

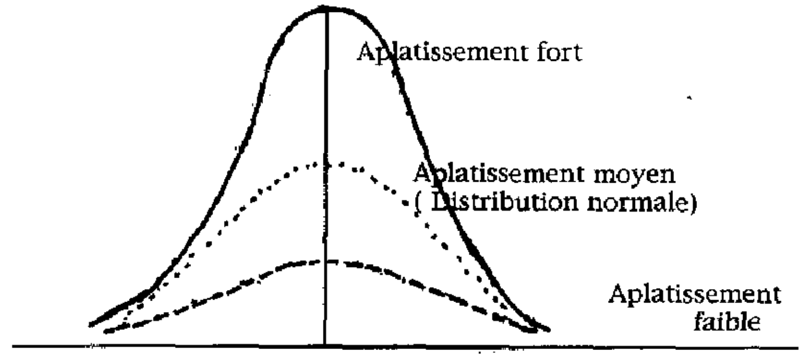


Fig. 1.6

II. SÉRIES STATISTIQUES À DEUX DIMENSIONS

1- Tableaux statistiques .

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

Tableau 1.3

X	x_1	x_2	....	x_i	....	x_m
Y	y_1	y_2	....	y_i	....	y_m

$X = x_1, x_2, \ldots\ldots\ldots, x_m$

$Y = y_1, y_2, \ldots\ldots\ldots, y_m$

Le couple (x_i, y_i) ; $i = 1, \ldots, m$; représente la valeur prise par (X,Y) dans la i[eme] observation .

On représente la distribution sous forme d'un nuage de points dans R^2 .

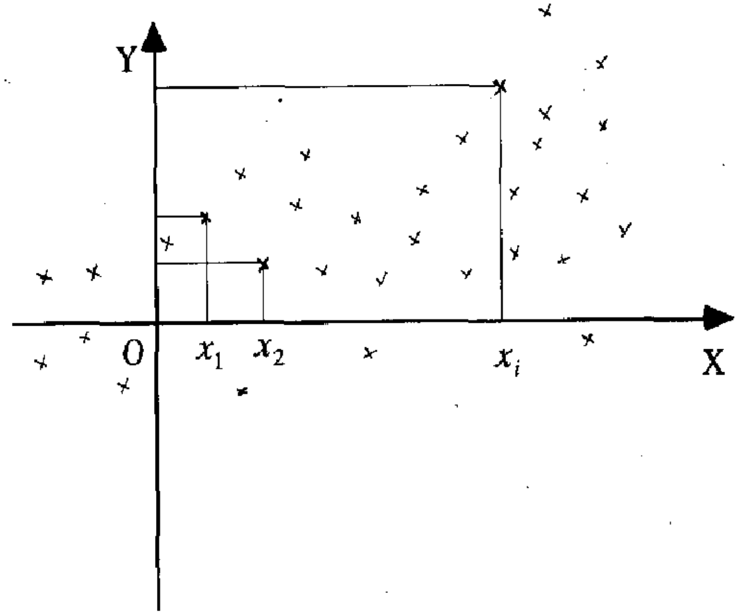


Fig. 1.7

b) Soit n_{ij} l'effectif du caractère

$$\left(x_i, y_j\right)\ ,\ i=1,\ldots,p; j=1,\ldots,m$$

X prend les valeurs $x_1, x_2, \ldots, x_p$

Y prend les valeurs $y_1, y_2, \ldots, y_m$

On utilise la représentation (tableau de contingence)

Tableau 1.4

X \ Y	y_1	y_2	......	y_j	....	y_m	Total partiel
x_1	n_{11}	n_{12}	......	n_{1j}	....	n_{1m}	$n_{1.}=\sum_{j=1}^{m} n_{1j}$
x_2	n_{21}	n_{22}	......	n_{2j}	....	n_{2m}	$n_{2.}=\sum_{j=1}^{m} n_{2j}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	......	n_{ij}	....	n_{im}	$n_{i.}=\sum_{j=1}^{m} n_{ij}$
$\vdots$	$\vdots$	$\vdots$				$\vdots$	$\vdots$
x_p	n_{p1}	n_{p2}	......	n_{pj}	....	n_{pm}	$n_{p.}=\sum_{j=1}^{m} n_{pj}$
Total partiel	$n_{.1} = \sum_{i=1}^{p} n_{i1}$	$n_{.2} = \sum_{i}^{p} n_{i2}$	......	$n_{.j} = \sum_{i}^{p} n_{ij}$	....	$n_{.m} = \sum_{i}^{p} n_{im}$	$n=n_{..}=\sum_{i,j} n_{ij}$

$n_{i.} = \sum_{j=1}^{m} n_{ij}$ est la distribution marginale de X

$n_{.j} = \sum_{i=1}^{p} n_{ij}$ est la distribution marginale de Y

On peut aussi utiliser la représentation en fréquence f_{ij} avec

$f_{ij} = \frac{n_{ij}}{n}$; $f_{i.} = \frac{n_{i.}}{n}$ et $f_{.j} = \frac{n_{.j}}{n}$

2-Cœfficient de corrélation linéaire

a) Lorsqu'il n'y a qu'une observation pour un couple (x_i, y_i) ; $i = 1, \ldots, m$; on définit le cœfficient de corrélation par le réel ρ :

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

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

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

b) Soit n_{ij} l'effectif du caractère (x_i, y_j); $i = 1, \ldots, p$; $j = 1, \ldots, m$

Le cœfficient de corrélation linéaire est défini par :

$$\rho = \frac{\frac{1}{n}\sum_i \sum_j n_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sigma_X \sigma_Y}$$

où $\sigma_X^2 = \frac{1}{n}\sum_i n_{i.}(x_i - \bar{x})^2$

et $\sigma_Y^2 = \frac{1}{n}\sum_j n_{.j}(y_j - \bar{y})^2$

EXERCICES

I.1 Soit le tableau statistique donnant le nombre d'enfants dans 116 familles

Tableau 1.5

Nombre d'enfants x_i	0	1	2	3	4	5 et plus
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 .

I.2 Le tableau suivant donne les quantités hebdomadaires d'essence reçues (en milliers de litres) par une station durant 36 semaines .

Tableau 1.6

12,5	15	18,5	15,5	16	11	11,5	12	16
11	15	16	15	12	12	16	20	23
20	15	18	18	22	15	27	18	15
16	25	18	18	17,1	20	23,5	16,8	12,5

1) Ranger ces quantités par valeur croissante

2) Construire un tableau des intervalles de classe , centre de classe , fréquences et fréquences cumulées

3) Construire l'histogramme des fréquences et la courbe des fréquences cumulées .

4) Trouver la moyenne , la médiane , le mode , la variance et l'écart-type

a) pour les quantités données par le tableau ci-dessus .

b) pour le tableau construit à la 2eme question .

Expliquer les résultats .

I.3 Soit le tableau donnant le poids de 133 étudiants :

Tableau 1.7

Poids (Kilogramme)	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 caractéristiques : mode , médiane , étendue , quartiles , moyenne et écart-type.

I.4 La répartition de 638 fonctionnaires par tranches de salaire net mensuel dans un établissement public en 1993 était :

Salaire mensuel (10^3 DA)	Nombre de fonctionnaires
Moins de 6	69
de 6 à moins de 7	86
— 7 ________ 8	220
— 8 ________ 9	92
— 9 ________ 10	27
— 10 ________ 11	11
— 11 ________ 12	20
— 12 ________ 13	19
— 13 ________ 14	19
— 14 ________ 15	42
— 15 ________ 16	26
— 16 ________ 17	07
Total	638

1) Représenter l'histogramme de cette répartition .

2) Quel est le salaire mensuel moyen . Trouver l'écart-type de la distribution des salaires . .

3) Déterminer le mode et la médiane de cette répartition .

SOLUTIONS DES EXERCICES

I.1 a) On cherche les fréquences f_i puis les fréquences cumulées F_i

Tableau 1.9

x_i	n_i	f_i	F_i
0	6	0,052	0,052
1	18	0,155	0,207
2	25	0,216	0,423
3	33	0,284	0,707
4	21	0,181	0,888
5 et plus	13	0,112	1
Total	116	1	

La courbe des fréquences cumulées est :

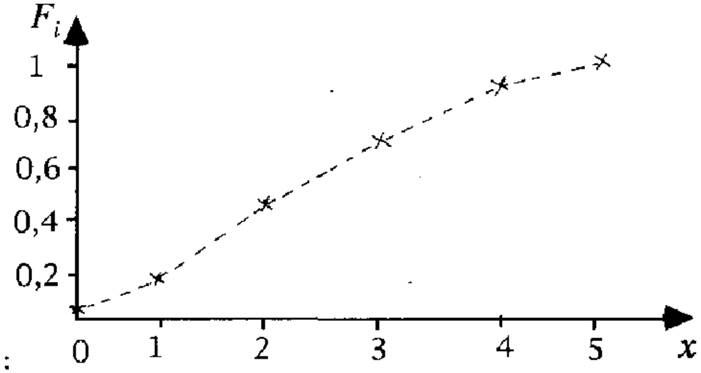


Fig. 1.8

b) Le mode est : 3 enfants

La médiane est : $M_e = 3$

Les quartiles sont : $Q_1 = 2$ et $Q_2 = 2$

c) En prenant une moyenne de 6 pour la classe " 5 et plus" , on trouve : $\bar{x} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{18+50+99+84+78}{116}$

$\bar{x} = 2{,}84$ enfants

et $\sigma^2 = \frac{\sum_i n_i x_i^2}{\sum_i n_i} - \bar{x}^2 = \frac{18+100+297+336+468}{116} - (2{,}84)^2$

$\sigma^2 = 2{,}46$

On a donc $\sigma = 1{,}57$

1.2. 1) **En classant les quantités par valeur croissante, on** obtient :

Tableau 1.10

11	11	11,5	12	12	12	12,5	12,5	15
15	15	15	15	15	15,5	16	16	16
16	16	16,8	17,1	18	18	18	18	18
18,5	20	20	20	22	23	23,5	25	27

2) On dresse le tableau

Tableau 1.11

Quantités 10^3 l	centres de classe x_i	Effectif n_i	Fréquences f_i	Fréquences cumulées F_i
De 10 à moins de 13	11,5	8	0,222	0,222
_13 ______16	14,5	7	0,194	0,416
_16 ______19	17,5	13	0,361	0,777
_19 ______21	20	3	0,083	0,860
_21 ______24	22,5	3	0,083	0,943
_24 ______27	25,5	2	0,057	1
Totaux		36	1	—

3) Histogramme des fréquences

L'amplitude élémentaire est $a = 3.10^3$. La hauteur du rectangle correspondant à la classe " De 19 à moins de 21" est

$$\frac{0,083}{2}.3 = 0,125$$

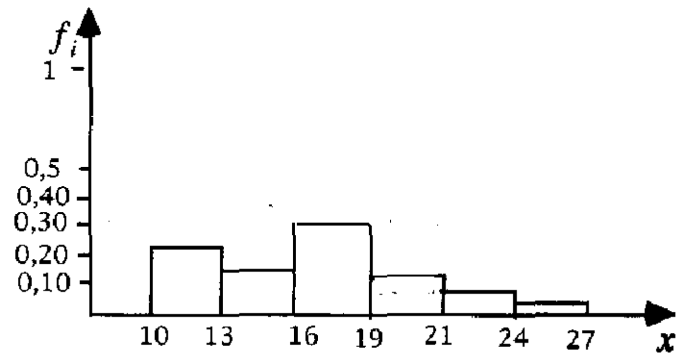


Fig.1.9
Courbe des fréquences cumulées

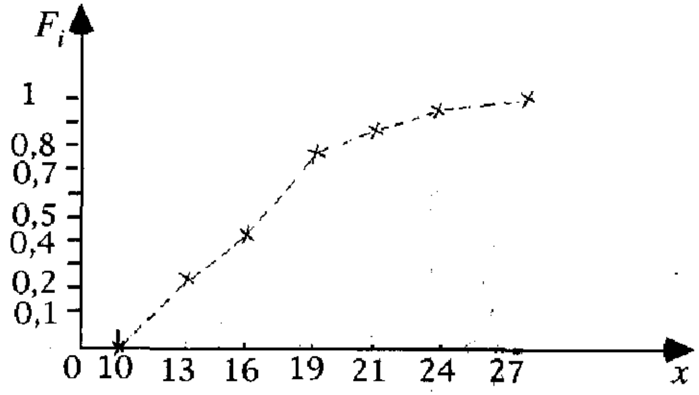


Fig.1.10

4)a) La moyenne $\bar{x}$ est :

$$\bar{x} = \frac{12,5+15+........+16,8+12,5}{36} = 16,75$$

La **médiane M_e correspondant** à la valeur du $\left(\frac{36+1}{2}\right)^{ieme} = 18,5$ élément , c'est à dire la moyenne du 18^{e} et du 19^{e} élément . Donc la médiane M_e vaut 16

Le mode est 15

la variance σ^2 vaut :

$$\sigma^2 = \sum x^2 - \bar{x}^2 = \frac{(12,5)^2+(15)^2+......+(12,5)^2}{36} - (16,75)^2$$

$\sigma^2 = 15,136$ et $\sigma = 3,89$

b) La moyenne pour les données groupées est $\bar{x} = \sum_i f_i x_i$

Tableau 1.12

x_i	f_i	$f_i x_i$	$f_i x_i^2$
11,5	0,222	2,553	29,360
14,5	0,194	2,813	40,789
17,5	0,361	6,318	110,556
20	0,083	1,66	33,2
22,5	0,083	1,868	42,019
25,5	0,057	1,454	37,064
Totaux	1	16,666	292,988

$\bar{x} = 16,67$

La médiane M_e est

$$M_e = 14,5 + 3\frac{0,5-0,416}{0,777-0,416} = 15,2$$

La classe modale est celle qui a la plus forte fréquence par unité d'amplitude , c'est à dire la classe " De 16 à moins de 19"

La variance σ^2 vaut 292,988 - (16,67)^2

$$\sigma^2 = \sum_i f_i x_i^2 - (\bar{x})^2 = 292,988 - (16,67)^2$$

$\sigma^2 = 15,099$ et $\sigma = 3,89$

Les calculs sont faits comme si tous les individus d'**une classe** avaient pour valeur le centre de la classe . Ceci **entraîne** évidemment une approximation . Les résultats obtenus sont différents de ceux obtenus à partir des données brutes . On perd parfois beaucoup en précision en regroupant les classes.

I.3

1) On construit le tableau

Tableau 1.13

Poids (Kg)	Centre de classe x_i	Fréquence f_i	Fréquence cumulée F_i	$f_i x_i$	$f_i x_i^2$
56			0		
	57	0,038		2,166	123,462
58			0,038		
	59	0,09		5,31	313,29
60			0,12		
	61	0,135		8,235	502,335
62			0,263		
	63	0,293		18,459	1162,917
64			0,556		
	65	0,271		17,615	1144,975
66			0,827		
	67,5	0,113		7,594	514,856
69			0,94		
	70,5	0,06		4,23	298,215
72			1		
Totaux	—	1	—	63,609	4060,05

Histogramme des fréquences

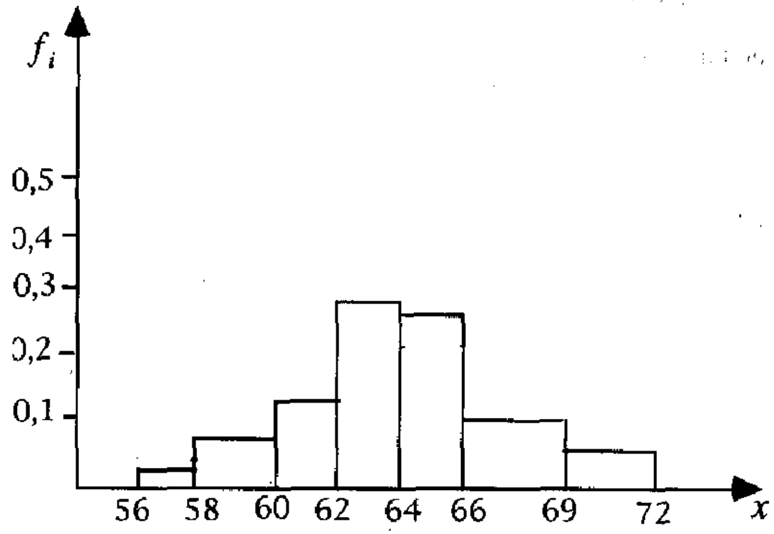


Fig. 1.11

Courbe des fréquences cumulées

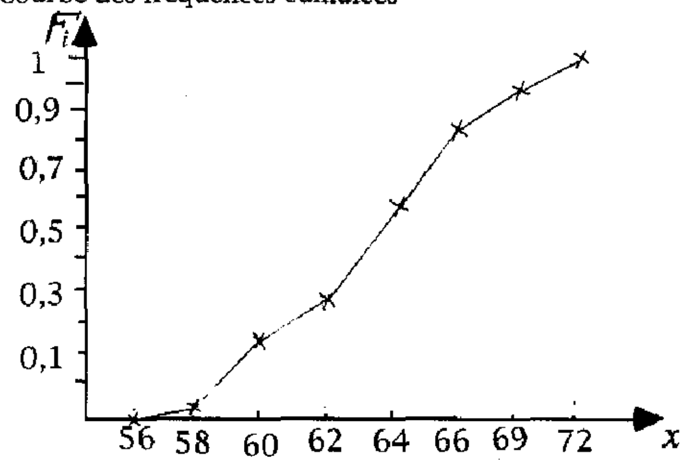


Fig. 1.12

2) La classe modale est : " De 62 à moins de 64 "

La médiane est déterminée par interpolation linéaire :

$$M_e = 62 + 2\frac{0,5-0,263}{0,556-0,263} = 63,62$$

L'étendue e est : e=72-56=16 Kilogrammes

Les deux quartiles Q_1 et Q_2 sont déterminés par interpolation linéaire . Le deuxième quartile correspond à la médiane .

$$Q_1 = 60 + 2\frac{0,25-0,12}{0,263-0,12} = 61,82$$

$$Q_3 = 64 + 2\frac{0,75-0,556}{0,827-0,556} = 65,43$$

La moyenne $\bar{x}$ est :

$$\bar{x} = \sum_i x_i f_i \quad , \quad \bar{x} = 63,61$$

La variance σ^2 est :

$$\sigma^2 = \sum_i f_i x_i^2 - \left(\bar{x}\right)^2$$

$$= 4060,05 - (63,61)^2 = 13,95$$

L'écart-type σ est : σ = 3,73

I.4

1) On construit le tableau (en supposant qu'aucun salaire n'est inférieur à 5.10^3 DA)

Tableau 1.14

Salaires 10^3DA	x_i	n_i	f_i	F_i	$f_i x_i$	$f_i x_i^2$
5.....				0		
	5,5	69	0,108		0,594	3,267
6.....				0,108		
	6,5	86	0,135		0,878	5,704
7.....				0,243		
	7,5	220	0,345		2,588	19,406
8.....				0,588		
	8,5	92	0,144		1,224	10,404
9.....				0,732		
	9,5	27	0,042		0,399	3,791
10.....				0,744		
	10,5	11	0,017		0,179	1,874
11.....				0,791		
	11,5	20	0,031		0,357	4,1
12.....				0,822		
	12,5	19	0,030		0,375	4,688
13.....				0,852		
	13,5	19	0,030		0,405	5,468
14.....				0,882		
	14,5	42	0,066		0,957	13,877
15.....				0,948		
	15,5	26	0,041		0,636	9,85
16.....				0,989		
	16,5	7	0,011		0,182	2,995
17.....				1		
Totaux	----	638	1	-----	8,774	85,424

On représente l'histogramme des fréquences de la répartition

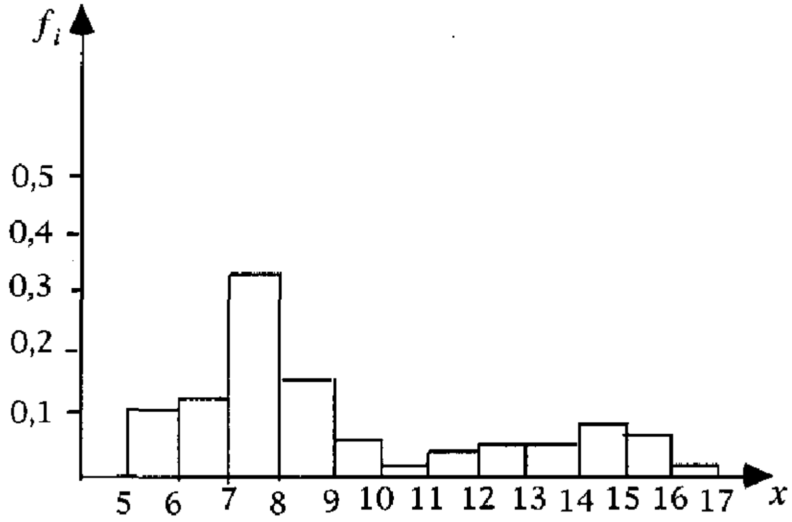


Fig. 1.13

2) Le salaire mensuel moyen $\bar{x}$ est $\bar{x} = \sum_i f_i x_i$

$$\bar{x} = 8,77.10^3\text{DA}$$

On a : $\sigma^2 = \sum_i f_i x_i^2 - \left(\bar{x}\right)^2 = 85,424 - (8,77)^2$

$$\sigma^2 = 8,511$$

d'où l'écart-type $\sigma = 2,92.10^3\text{DA}$

3) La classe modale est celle qui a la plus forte fréquence par unité d'amplitude : " De 7 à moins de 8 "

La médiane M_e est déterminée par interpolation linéaire . On a :

$$M_e = 7 + 1\,\frac{0,5-0,243}{0,588-0,243} = 7,745$$

$$M_e = 7,75.10^3$$

CHAPITRE 2

NOTIONS DE CALCUL DES PROBABILITÉS-VARIABLES ALÉATOIRES

I. PROBABILITÉ

1-Définitions L'ensemble Ω , dit ensemble fondamental , désigne l'ensemble des résultats élémentaires d'une expérience aléatoire . Une épreuve est un élément w de Ω.

Un sous-ensemble $\mathcal{A}$ de $\mathcal{P}(\Omega)$ est dit $\sigma-$algèbre de parties de Ω si :

a) $\Omega \in \mathcal{A}$

b) si $A \in \mathcal{A}$, $\overline{A} \in \mathcal{A}$

c) si $A_1, A_2, \ldots, \in \mathcal{A}$, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Le triplet $(\Omega, \mathcal{A}, P)$ est dit espace de probabilité où $\mathcal{A}$ est la $\sigma-$**algèbre** des événements et P une application définie sur $\mathcal{A}$ et vérifiant :

a) $P(A) \geq 0$, $\forall A \in \mathcal{A}$

b) $P(\Omega) = 1$

c) Pour $A_1, A_2, \ldots, \in \mathcal{A}$ et $A_i \cap A_j = \phi$ (pour $i \neq j$),

On a $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

On en déduit :

a) $P(\phi) = 0$

b) $\forall A \in \mathcal{A}$, $\forall B \in \mathcal{A}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

c) $\forall A \in \mathcal{A}$, $P(\overline{A}) = 1 - P(A)$

2-Indépendance Soient $(\Omega, \mathcal{A}, P)$ un espace de probabilité et deux événements A et B avec $P(B) > 0$. La probabilité conditionnelle de A à B (ou probabilité de A sachant B) $P(A/B)$ est définie par :

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

A est indépendant de B si $P(A/B) = P(A)$ et donc $P(AB) = P(A)P(B)$

3-Formule de Bayes . Soient un événement A et n événement $B_1, B_2, \ldots, B_n$; la famille $(B_i)_i$ constituant un système complet d'événements de Ω (c'est à dire : $\bigcup_{i=1}^{n} B_i = \Omega$ et $B_i \cap B_j = \phi$ pour $i \neq j$) . On a :

$$P(A) = \sum_{i=1}^{n} P(A/B_i)P(B_i)$$

Soit A un événement tel que $P(A) > 0$ et soient n événements $B_1, B_2, \ldots, B_n$ constituant un système complet d'événements de Ω avec : $P(B_i) > 0$, $\forall i = 1, \ldots, n$

On a :

$$P(B_i/A) = \frac{P(B_i)P(A/B_i)}{\sum_{j=1}^{n} P(B_j)P(A/B_j)}$$

II-VARIABLES ALÉATOIRES DISCRETES

1-Fonction de répartition . Une variable aléatoire réelle X est une application de Ω dans R . L'ensemble des réalisations de X est noté X(Ω) . X est discrète lorsque X(Ω) comporte un nombre fini ou une infinité dénombrable d'éléments .

Si X prend n valeurs distinctes $x_1, x_2, \ldots, x_n$ avec des probabilités respectives $p_1, p_2, \ldots, p_n$; ces probabilités $p_i = P(X = x_i), (i = 1, 2, \ldots, n)$ vérifient la condition :

$$\sum_{i=1}^{n} p_i = 1$$

Si X prend une infinité dénombrable de valeurs $x_1, x_2, \ldots, \ldots x_n, \ldots$ (On suppose ces valeurs ordonnées de façon croissante) avec des probabilités respectives $p_1, p_2, \ldots, p_n, \ldots$; ces probabilités p_i vérifient la condition :

$$\sum_{i=1}^{\infty} p_i = 1$$

On appelle fonction de répartition de la variable aléatoire X définie sur (Ω A,P) la fonction $F(x)$ définie sur R par :

$$F(x) = p(X \le x)$$

La fonction de répartition F possède les propriétés suivantes:

a) F est croissante

b) F est constante sur tout intervalle $[x_k, x_{k+1}[$

c) F est continue à droite en tout point x de R

d) $\lim_{x \to -\infty} F(x) = 0$ et $\lim_{x \to \infty} F(x) = 1$

On a $F(x) = \sum_{x_i \le x} P(X = x_i)$ si $x_k \le x < x_{k+1}$

$$= p_1 + p_2 + \ldots\ldots p_k$$

Pour tous a et b réels , $P(a < X \le b) = F(b) - F(a)$

2- Caractéristiques de variables aléatoires

Si la variable aléatoire X prend un nombre fini d'éléments , on définit les caractéristiques :

a) Espérance mathématique de X : $E(X) = \sum_{i=1}^{n} p_i x_i$

b) Moment d'ordre k de X : $m_k = E(X^k) = \sum_{i=1}^{n} p_i x_i^k$

c) Moment centré d'ordre k de X :

$$\mu_k = E\left[(X - m_1)^k\right] = \sum_{i=1}^{n} p_i (x_i - m_1)^k$$

d) Variance de X (c'est le moment centré d'ordre 2)

$$V(X) = E\left[(X - E(X))^2\right] = E\left[(X - m_1)^2\right]$$

$$= \sum_{i=1}^{n} p_i (x_i - m_1)^2 = E(X)^2 - [E(X)]^2 = E(X^2) - m_1^2$$

La racine carré de la variance $\sqrt{V(X)}$ est appelée écart-type et est notée $\sigma(X)$

Si la variable aléatoire X prend une infinité dénombrable d'éléments , la somme de la série $\sum_{i=1}^{\infty} p_i x_i$ peut ne pas exister .
On dira alors que X n'a pas d'espérance mathématique . Cette condition de convergence est valable pour les moments d'ordre k .

III-VARIABLES ALÉATOIRES CONTINUES

Soit (Ω A,,P) un espace de probabilité et F la fonction de répartition de X définie sur (Ω,A,,P) . X est dite continue si X(Ω) est un intervalle ou réunion d'intervalles de R .
X est dite absolument continue , si de plus , il existe une fonction $f(x)$ définie sur R telle que :

a) $f(x) \geq 0 \qquad \forall x \in R$

b) $f(x)$ est continue sur R sauf , peut -être en un nombre fini de points.

c) $\int_{-\infty}^{\infty} f(x)dx$ converge et $\int_{-\infty}^{\infty} f(x)dx = 1$

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

f est dite densité de la variable X .
Pour tous réels a et b tels que a<b , on a :

$$P(a < X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$$

$$P(X = a) = 0$$

On définit les caractéristiques suivantes

a) Espérance mathématique de x .

L'intégrale , si elle existe , $E(X) = \int_{-\infty}^{\infty} f(x)dx$ **est dite** espérance mathématique de X

b) Moment d'ordre k de X .

L'intégrale , si elle existe , $m_k = E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx$ **est** dite moment d'ordre k de X

c) Moment centré d'ordre k de X

L'intégrale , si elle existe , $\mu_k = E\left[(X - m_1)^k\right]$

$$= \int_{-\infty}^{\infty} (x - m_1)^k f(x)dx$$

est dite moment centré d'ordre k de X .

Si k=2 , on obtient la variance $V(X)$ notée $\sigma^2(X)$

$$\sigma^2(X) = V(X) = \int_{-\infty}^{\infty} (x - m_1)^2 f(x)dx$$
$$= \int_{-\infty}^{\infty} x^2 f(x)dx - m_1^2$$

EXERCICES

2.1. Un atelier de fabrication dispose de 3 machines . Les probabilités que les machines 1,2 ou 3 tombent en panne en une journée sont 0,02 ; 0,08 et 0,12 respectivement .
Calculer les probabilités qu'en une journée :

1) On n'ait aucune panne .
2) Une machine et une seule tombe en panne .
3) Deux machines seulement tombent en panne .
4) Au moins une machine tombe en panne .

2.2. Une urne contient 4 boules blanches et 6 noires .

1) On tire sans remise 2 boules de cette urne . Trouver la suite de répartition de la variable aléatoire X correspondant au nombre de boules noires extraites . Trouver la fonction de répartition $F(x)$ de X . Calculer $E(X)$ et $\sigma(X)$

2) Mêmes questions dans le cas où le tirage se fait avec remise .

2.3. Parmi les articles produits par une machine 3% sont défectueux . Calculer la probabilité que parmi 112 articles produits en une journée il y ait :

1) Deux défectueux
3) Dix défectueux

2.4. Un magasin reçoit pour la vente des produits en provenance de 3 usines dont les parts relatives sont :
Usine 1 : 50% ; Usine 2 : 30% ; Usine 3 : 20%
Le pourcentage de produits défectueux dans la production de ces usines est respectivement : 2% , 3% et 5%.
Quelle est la probabilité qu'un produit acheté au hasard soit de bonne qualité ?

2.5. Deux tireurs , indépendamment l'un de l'autre , tirent un coup chacun sur une cible mobile . La probabilité d'atteindre la cible est égale à 0,2 pour le premier et 0,1 pour le second . La cible est atteinte par un seul coup . Quelle est la probabilité pour que la cible soit atteinte par le premier tireur ?

2.6. La densité d'une variable aléatoire X est :

$$f(x) = \alpha e^{-\alpha x} \quad ; \quad x \geq 0 \quad , \quad \alpha \in R$$

1) Trouver la fonction de répartition $F(x)$ de X
2) Trouver les densités des variables aléatoires :

a) $Y = \sqrt{X}$; b) $Z = X^2$; c) $T = \dfrac{1}{\alpha} L_n X$

SOLUTIONS DES EXERCICES

2.1. On désigne par $A_i \quad (i=1,2,3)$ l'événement :
A_i : " La machine i tombe en panne " et on suppose que les événements A_1, A_2 et A_3 sont indépendants . Les événements A_1', A_2' et A_3' (où $A_i' = A_i$ ou $\overline{A}_i$) le seront aussi .

1) L'événement A : " Aucune machine ne tombe en panne " est $\overline{A}_1\overline{A}_2\overline{A}_3$ et

$$P(A) = P(\overline{A}_1\overline{A}_2\overline{A}_3) = P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)$$
$$= 0{,}98 \,.\, 0{,}92 \,.\, 0{,}88 \,. = 0{,}79$$

2) L'événement B : " Une machine et seule tombe en panne " s'écrit :

$$B = A_1\overline{A}_2\overline{A}_3 + \overline{A}_1A_2\overline{A}_3 + \overline{A}_1\overline{A}_2A_3 \quad \text{et}$$
$$P(B) = P(A_1\overline{A}_2\overline{A}_3) + P(\overline{A}_1A_2\overline{A}_3) + P(\overline{A}_1\overline{A}_2A_3)$$
$$= P(A_1)P(\overline{A}_2)P(\overline{A}_3) + P(\overline{A}_1)P(A_2)P(\overline{A}_3) + P(\overline{A}_1)P(\overline{A}_2)P(A_3)$$
=0,02.0,92.0,88+0,98.0,08.0,88+0,98.0,92.0,12.
= 0,193

3) L'événement C : "Deux machines seulement tombent en panne " est

$$C = A_1A_2\overline{A}_3 + A_1\overline{A}_2A_3 + \overline{A}_1A_2A_3$$
P(C)=0,02.0,08.0,88+0,02.0,92.0,12+0,98.0,08.0,12
= 0,012

4) L'événement D : " Au moins une machine tombe en panne" est D=B+C+E où E est l'événement : " Trois machines tombent en panne "

On a $P(E) = P(A_1A_2A_3)$
$$= P(A_1)P(A_2)P(A_3)$$
= 0,02.0,08.0,12=0,0002

et P(D)=P(B)+P(C)+P(E)
=0,193+0,012+0,0002=0,2052

On peut retrouver ce résultat en remarquant que
$D = \Omega - A = \overline{A}$ et
P(D)=1-P(A)
=1-0,79=0,21

2.2. On note $N_i (i=1,2,3)$ l'événement :
" La boule extraite au i^{eme} tirage est noire "
Soient $x_1 = 0, x_2 = 1, x_3 = 2$ les différents valeurs que peut prendre X.

1) On a $p_1 = P(X=0) = P(\overline{N}_1\overline{N}_2) = \frac{4}{10}\frac{3}{9} = 0{,}133$

$$p_2 = P(X=1) = P(N_1\overline{N}_2 + \overline{N}_1N_2)$$
$$= P(N_1\overline{N}_2) + P(\overline{N}_1N_2)$$
$$= \frac{6}{10}\frac{4}{9} + \frac{4}{10}\frac{6}{9} = 0{,}533$$
$$p_3 = P(X=2) = P(N_1N_2) = \frac{6}{10}\frac{5}{9} = 0{,}333$$

La suite de répartition est

Tableau 2.1

x_i	0	1	2
P_i	0,133	0,533	0,333

et $F(x)=\begin{cases}0 & si \quad x<0\\ 0,133 & si \quad 0\le x<1\\ 0,666 & si \quad 1\le x<2\\ 1 & si \quad 2\le x\end{cases}$

On a : $E(X)=\sum_{i=1}^{3}p_ix_i=1,199$

$$V(X)=E\left(X^2\right)-E(X)^2=\sum_{i=1}^{3}p_ix_i^2-(1,2)^2=0,43$$

et $\sigma(X)=0,65$

2) Si le tirage se fait avec remise on a $P(N_i)=\dfrac{6}{10}$

et $P\left(\overline{N}_i\right)=\dfrac{4}{10}$ $(i=1,2)$; donc

$$p_1=P(X=0)=P\left(\overline{N}_1\,\overline{N}_2\right)=\frac{4}{10}\frac{4}{10}=0,16$$

$$p_2=P(X=1)=P\left(N_1\overline{N}_2+\overline{N}_1N_2\right)$$

$$=\frac{6}{10}\frac{4}{10}+\frac{4}{10}\frac{6}{10}=0,48$$

$$p_3=P(X=2)=P\left(N_1N_2\right)=\frac{6}{10}\frac{6}{10}=0,36$$

La suite de répartition est :

Tableau 2.2

x_i	0	1	2
p_i	0,16	0,48	0,36

et $F(x)=\begin{cases}0 & si \quad x<0\\ 0,16 & si \quad 0\le x<1\\ 0,64 & si \quad 1\le x<2\\ 1 & si \quad 2\le x\end{cases}$

On a : $E(X)=\sum_{i=1}^{3}p_ix_i=1,2$

$$V(X)=E\left(X^2\right)-E(X)^2=0,48$$

et $\sigma(X)=0,69$

2.3. Soit X la variable aléatoire correspondant au nombre d'articles défectueux parmi les 112 produits . Si la probabilité qu'un article soit défectueux est 0,03 , la probabilité qu'un article est non défectueux est 0,97 .

1) On a : $P(X=2)=C_{112}^{2}(0,03)^2(0,97)^{110}=0,1958$

C_{112}^{2} représente le nombre de manières différentes de choisir 2 articles (défectueux) parmi les 112 .

2) $P(X=10)=C_{112}^{10}(0,03)^{10}(0,97)^{102}=0,0015$

Nous verrons plus loin qu'on peut résoudre ce type de problèmes de façon beaucoup plus simple (du point de vue calcul) en utilisant une approximation .

2.4. Soient les événements :

H_i : " Le produit acheté provient de l'usine i" $i = 1,2,3$
et A : " Le produit acheté est de bonne qualité "

On a : $P(H_1) = 0,5$; $P(H_2) = 0,3$; $P(H_3) = 0,2$

$$P(A/H_1) = 0,98 \ ; \ P(A/H_2) = 0,97 \ ; \ P(A/H_3) = 0,95$$

Les H_i $(i = 1,2,3)$ forment un système complet d'événements. L'application de la formule des probabilités totales donne :

$$P(A) = \sum_{i=1}^{3} P(H_i)P(A/H_i) = 0,97$$

2.5. Soient les événements :

A: " La cible est atteinte par le premier tireur "
B: " La cible est atteinte par le second tireur "
C: " La cible est atteinte par un seul coup "

$H_1 = AB$: " La cible est atteinte par les deux tireurs "
$H_2 = A\overline{B}$: " La cible est atteinte par le premier tireur seulement "
$H_3 = \overline{A}B$: " La cible est atteinte par le second tireur seulement"
$H_4 = \overline{A}\,\overline{B}$: " La cible n'est atteinte par aucun tireur "

Les événements (A,B) sont indépendants . Il en est de même des événements $(A,\overline{B})$, $(\overline{A},B)$ et $(\overline{A},\overline{B})$.

Les événements H_i $(i = 1,2,3,4)$ forment un système complet d'événements .

On a:

$$P(H_1) = P(A)P(B) = 0,2.0,1 = 0,02$$
$$P(H_2) = P(A\overline{B}) = P(A)P(\overline{B}) = 0,2.0,99 = 0,198$$
$$P(H_3) = P(\overline{A}B) = P(\overline{A})P(B) = 0,98.0,1 = 0,098$$
$$P(H_4) = P(\overline{A}\,\overline{B}) = P(\overline{A})P(\overline{B}) = 0,98.0,99 = 0,97$$

La probabilité cherchée est donnée par la formule de Bayes:

$$P(H_2/C) = \frac{P(H_2)P(C/H_2)}{\sum_{i=1}^{4} P(H_i)P(C/H_i)} = 0,67$$

avec $P(C/H_1) = P(C/H_4) = 0$

et $P(C/H_2) = P(C/H_3) = 1$

2.6. 1) On a $F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x \alpha e^{-\alpha t}dt = 1 - e^{-\alpha x} ; x \geq 0$

2) On cherche d'abord les fonctions de répartition Y,Z et T ; puis on calcule les dérivées correspondantes

a) $F_Y(y) = P(Y \leq y) = P(\sqrt{X} \leq y) = P(X \leq y^2)$

$$= 1 - e^{-\alpha y^2}$$

et $f_Y(y) = 2\alpha y e^{-\alpha y^2}$; $y \geq 0$

b) $F_Z'(z) = P(Z \leq z) = P(X^2 \leq z) = P(X \leq \sqrt{z})$

$$= 1 - e^{-\alpha\sqrt{z}}$$

et $f_Z(z)=\dfrac{\alpha}{2\sqrt{z}}e^{-\alpha\sqrt{z}}\quad z>0$

c) $F_T(t)=P(T\le t)=P\left(\dfrac{1}{\alpha}\ln X\le t\right)=P\left(X\le e^{\alpha t}\right)$

$=1-\exp\left(-\alpha e^{\alpha t}\right)$

et $f_T(t)=\alpha^2\exp\left(\alpha\left(t-e^{\alpha t}\right)\right)$

CHAPITRE 3

PRINCIPALES LOIS DE PROBABILITÉ. -CONVERGENCES

I-LOIS DE PROBABILITÉ

Les représentations graphiques des séries statistiques des phénomènes observées ont des formes particulières dont certaines se rencontrent si souvent qu'elles ont fait l'objet d'études particulières . Ainsi il est préférable parfois de substituer à une série observée une loi de probabilité qui a l'avantage d'être tabulée et de donner quelques caractéristiques simples telles que la moyenne ou l'écart-type.
Nous présentons dans cette partie les principales lois utilisées.

1-Loi de BERNOULLI.

La variable aléatoire X définie sur un espace de probabilité $(\Omega,\mathcal{A},P)$, où Ω est l'ensemble fondamental des événements, $\mathcal{A}$ la σ-algèbre des événements et P la probabilités définie sur l'espace probabilisable $(\Omega,\mathcal{A})$, est dite suivre une loi de BERNOULLI de paramètre p si $X(\Omega)=\{0,1\}$ Ces deux valeurs sont dites aussi échec et succès .

On a : $P(X=0)=q$, $P(X=1)=p$ avec $p+q=1$

$$E(X)=p\ ;\ V(X)=pq$$

2- Loi binomiale B(n,p)

Soit la répétition de n épreuves indépendantes de BERNOULLI de paramètre p . La loi de la variable aléatoire X égale au nombre de succès dans la répétition de ces n épreuves est dite binomiale ;

On a $P(X = k) = C_n^k p^k (1-p)^{n-k}$

$$E(X) = np \;;\; V(X) = npq \quad \text{avec } p + q = 1$$

Loi de la somme de deux variables binomiales

Soient deux variables aléatoires indépendantes X_1 et X_2 suivant une loi binomiale $B(n_1, p)$ et $B(n_2, p)$ respectivement.

La loi de de $Z = X_1 + X_2$ est une loi binomiale $B(n_1 + n_2, p)$.

3- Loi hypergéométrique

Parmi les $N = n + b$ individus d'une population , n possèdent un certain caractère A . On procède à k tirages sans remise . Soit X la variable aléatoire représentant le nombre d'individus tirés ne possédant pas le caractère A . La probabilité que l'on ait tiré x individus ne possédant pas le caractère étudié est :

$$P(X = x) = \frac{C_b^x C_n^{k-x}}{C_{b+n}^k}$$

On a

$$E(X) = \frac{kb}{b+n} \;;\; V(X) = \frac{kbn(b+n-k)}{(b+n)^2(b+n-1)}$$

Si N tend vers l'infini , la loi hypergéométrique tend vers la loi binomiale.

4-Loi géométrique

Soit une suite d'épreuves de BERNOULLI de paramètre p . La probabilité pour que le premier succès apparaisse à la k^{ieme} épreuve est :

$$P(X = k) = pq^{k-1} \quad \text{avec} \quad q = 1 - p$$

On a : $E(X) = \frac{1}{p}$; $V(X) = \frac{1-p}{p^2}$.

5-Loi de POISSONP (λ)

Une variable aléatoire X est dite suivre une loi de Poisson de paramètre λ (λ constante positive) si :

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

On a : $E(X) = \lambda$; $V(X) = \lambda$

Cette loi s'applique pour la détermination d'une probabilité de succès relativement à un événement réparti dans le temps (par exemple loi du nombre d'appels téléphoniques pendant un intervalle de temps) . λ est le nombre moyen de succès correspondant à l'intervalle de temps .

Loi de la somme de deux variables indépendantes

Soient 2 variables aléatoires X_1 et X_2 indépendantes suivant une loi P (λ_1) et P (λ_2) respectivement . La loi $Z = X_1 + X_2$ est une loi de Poisson P $(\lambda_1 + \lambda_2)$.

6- Loi uniforme

Une variable aléatoire X absolument continue est dite distribuée uniformément sur l'intervalle $[a,b]$ si sa densité de probabilité est :

$$f(x)=\begin{cases}\dfrac{1}{b-a} & ; \quad x\in[a,b]\\ 0 & ; \quad Ailleurs\end{cases}$$

On a : $$E(X)=\frac{a+b}{2} \quad ; \quad V(X)=\frac{(a-b)^2}{12}$$

7- Loi exponentielle

Une variable aléatoire X absolument continue est dite variable aléatoire exponentielle de paramètre α (α constante positive) si sa densité de probabilité est :

$$f(x)=\begin{cases}\alpha e^{-\alpha x} & ; \quad x\geq 0\\ 0 & ; \quad x<0\end{cases}$$

On a : $$E(X)=\frac{1}{\alpha} \quad ; \quad V(X)=\frac{1}{\alpha^2}$$

8-Loi de CAUCHY

Une variable aléatoire X absolument continue suit une loi de Cauchy si sa densité de probabilité est :

$$f(x)=\frac{a}{\pi\left(x^2+a^2\right)} \quad ; \quad a>0 \quad , \quad x\in R$$

$E\left(X^k\right)$ n'existe pas ; $k\geq 1$

9-Loi normale (ou de LAPLACE-GAUSS) $N(m,\sigma)$

Une variable aléatoire X absolument continue est dite normale (ou suit une loi normale , ou gaussienne) si elle admet pour densité de probabilité la fonction :

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

où m et σ sont respectivement la moyenne et l'écart-type . On dit aussi que X suit une loi normale de paramètres (m,σ): $N(m,\sigma)$.

Si $X^*=\dfrac{X-m}{\sigma}$ est la variable aléatoire réduite correspondante , alors $E\left(X^*\right)=0$ et $V\left(X^*\right)=1$ et la densité de X^* s'exprime par :

$$g(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)$$

Si X suit une loi normale $N(m,\sigma)$, alors $E(X)=m$ et $V(X)=\sigma^2$

On peut ramener tout calcul sur la fonction de répartition d'une variable aléatoire normale $N(m,\sigma)$ à un calcul sur la fonction de répartition d'une variable aléatoire normale réduite $N(0,1)$. On désigne par Φ cette fonction :

$$P(X \le a) = P\left(\frac{X-m}{\sigma} \le \frac{a-m}{\sigma}\right) = P\left(X^* \le \frac{a-m}{\sigma}\right) = \Phi\left(\frac{a-m}{\sigma}\right)$$

Les valeurs de Φ sont données par les tables . Elles permettent de faire les calculs sur une loi normale de paramètres (m, σ) quelconques .
Cette loi est très importante tant du point de vue théorique que pratique .

Loi de la somme de variables aléatoires normales

Soient k variables aléatoires $X_1, X_2, \dots, X_k$ indépendantes distribuées suivant une loi normale $N(m_i, \sigma_i); i = 1, \dots, k;$
La variable aléatoire

$$Z = \sum_{i=1}^{k} X_i$$

suit une loi normale $N\left(\sum_{i=1}^{k} m_i, \sqrt{\sum_{i=1}^{k} \sigma_i^2}\right)$

10-Loi Gamma γ_r

Une variable aléatoire positive X suit une loi gamma de paramètre $r(\gamma_r)$ si sa densité est :

$$f(x) = \frac{1}{\Gamma(r)} e^{-x} x^{r-1}$$

Où $\Gamma(r) = \int_0^{\infty} e^{-t} t^{r-1} dt \quad ; \quad r > 0$

On a : $E(X) = r \quad ; \quad V(X) = r$

La loi de gamma est utilisée surtout dans les calculs d'autres lois . Elle sert aussi en modélisation .

11-Loi Bêta

Une variable aléatoire absolument continue suit une loi bêta si sa densité de probabilité est :

$$f(x) = \begin{cases} \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & ; \quad 0 < x < 1 \\ 0 & ; \quad Ailleurs \end{cases}$$

où α et β sont des constantes et $B(\alpha, \beta)$ est la fonction bêta.

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

On a : $E(X) = \dfrac{\alpha}{\alpha+\beta} \quad ; \quad V(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Comme la loi gamma , la loi bêta sert dans les calculs d'autres lois . On l'utilise aussi pour les répartitions concentrées sur l'intervalle $[0,1]$

12-Loi du Khi-carré

Soient k variables aléatoires $X_1, X_2, \dots, X_k$ indépendantes suivant une loi normale $N(m_i, \sigma_i)$, $i = 1, \dots, k.$

La variable aléatoire

$$Z = \sum_{i=1}^{k} \left(\frac{X_i - m_i}{\sigma_i} \right)^2$$

suit une loi de Khi-carré $\left(\chi_k^2\right)$ à k degrés de liberté .

La densité de probabilité de la variable aléatoire Z est :

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)} e^{-\frac{x}{2}} x^{\frac{k}{2}-1} & ; \quad x \geq 0 \\ 0 & ; \quad x < 0 \end{cases}$$

On a : $E(Z) = k \quad ; \quad V(Z) = 2k$

Cette loi est surtout utile dans les tests statistiques

13-Loi de FISHER-SNEDECOR

Soient deux variables aléatoires indépendantes X et Y distribuées suivant une loi du χ_p^2 et χ_q^2 respectivement .

La variable $F(p,q) = \dfrac{X/p}{Y/q}$ suit une loi de Fisher à p et q degrés de liberté .

La densité de probabilité de la variable F est :

$$g(x) = \frac{1}{B\left(\frac{p}{2}, \frac{q}{2}\right)} \frac{\left(\frac{p}{q}\right)^{\frac{p}{2}} x^{\frac{p}{2}-1}}{\left(1 + \frac{p}{q} x\right)^{\frac{p+q}{2}}} \quad ; \quad x > 0$$

On a : $E(F) = \dfrac{q}{q-2} \quad ; \quad V(F) = 2\dfrac{q^2}{p} \dfrac{p+q-2}{(q-2)^2(q-4)}$

La loi F est utilisé pour le traitement des données (analyse de la variance , tests statistiques).

14-Loi de STUDENT

Soient deux variables aléatoires indépendantes X et Y distribuées suivant une loi normale $N(0,1)$ et celle du χ_k^2 respectivement .

La variable $T = \dfrac{X}{\sqrt{\frac{Y}{k}}}$ suit une loi de Student à k degrés de liberté

On a : $E(T) = 0 \;,\; k > 1 \quad ; \quad V(T) = \dfrac{k}{k-2} \;,\; k > 2$

Si on a k=1 , alors on obtient la loi de Cauchy dont la densité est $g(x) = \dfrac{1}{\pi\left(1+x^2\right)}$

D'une façon générale la densité T s'écrit :

$$g(x) = \frac{1}{\sqrt{k} B\left(\frac{1}{2}, \frac{k}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{k}\right)^{\frac{k+1}{2}}}$$

La loi T de Student est utilisée surtout dans les tests statistiques

15-Loi de WEIBULL

Une variable aléatoire X absolument continue est dite suivre une loi de Weibull si la densité de probabilité est :

$$f(x)=\begin{cases}\beta\alpha x^{\alpha-1}e^{-\beta x^{\alpha}} & , \quad x\geq 0\\ 0 & , \quad x<0\end{cases}$$

où β et α sont des constantes positives .

On a : $$E(X)=\beta^{-\frac{1}{2}}\Gamma\left(1+\frac{1}{\alpha}\right)$$

$$V(X)=\beta^{-\frac{2}{\alpha}}\left[\Gamma\left(1+\frac{2}{\alpha}\right)-\Gamma^2\left(1+\frac{1}{\alpha}\right)\right]$$

La loi de Weibull est utilisée particulièrement en théorie de la fiabilité . Elle a l'avantage de s'ajuster à différents résultats expérimentaux .

II-CONVERGENCES

1-Inégalité de BIENAYMÉ-TCHEBICHEV

Soit $\varepsilon\in R^+$ et une variable aléatoire X telle que : $E(X)=m$ et $V(X)$ existent . On a :

$$P\left(|X-E(X)|\geq\varepsilon\right)\leq\frac{V(X)}{\varepsilon^2}$$

Si on pose $\varepsilon=t\sigma$ (où σ est l'écart-type et $t\in R^+$) l'inégalité s'écrit:

$$P\left(|X-m|\geq t\sigma\right)\leq\frac{1}{t^2}$$

2-Loi des grands nombres

Soit une suite de variables aléatoires indépendantes $X_1,X_2,\ldots,X_n,\ldots,$ de même loi de distribution , telle que $E(X_i)=m \quad (i=1,\ldots,n,\ldots)$ et soit $S_n=\dfrac{X_1+X_2+\ldots+X_n}{n}$

On a : $$\lim_{n\to\infty}P\left(|S_n-m|<\varepsilon\right)=1$$

3-Théorème de la limite centrale

Soit $X_1,X_2,\ldots,X_n,\ldots,$ une suite de variables aléatoires indépendantes de même loi de distribution telle que : $E(X_i)=m$ et $V(X_i)=\sigma^2 \quad (i=1,\ldots,n,\ldots)$

et soit $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$. La variable aléatoire :

$$\frac{\overline{X}-m}{\sigma/\sqrt{n}}$$

converge en loi vers une variable aléatoire normale $N(0,1)$.

III-APPROXIMATIONS DES LOIS BINOMIALE.DE POISSON ET NORMALE

Tirage avec remise
loi binomiale
$P(X=x)=C_n^x p^x (1-p)^{n-x}$

Si $p<0,1; n \geq 50$ approximation de la loi binomiale par la loi de poisson avec
$\lambda = np$

Pour $n>18$,p non petit .approximation de la loi binomiale par une loi normale $N(m,\sigma)$ avec
$m=E(X)=np$
et
$\sigma=\sqrt{np(1-p)}$

Loi de Poisson
$P(X=x)=e^{-\lambda}\dfrac{\lambda^x}{x!}$

Pour $\lambda \geq 18$ approximation de la loi de poisson par une loi normale $N(m,\sigma)$ avec
$m=E(X)=\lambda \quad et \quad \sigma=\sqrt{m}$

$Y=\dfrac{X-E(X)}{\sigma}$ suit une loi N(0,1)

EXERCICES

3.1. Démontrer que pour une distribution normale réduite le cœfficient d'asymétrie γ est nul et le cœfficient d'aplatissement δ vaut 3 .

3.2. Montrer que si X_1 suit une loi gamma de paramètre $r(\gamma_r)$ on a : $E(X_1)=r$ et $V(X_1)=r$.
Soit une autre variable aléatoire X_2 suivant une loi gamma de paramètre $S(\gamma_S)$. Montrer que la variable aléatoire $Z=X_1+X_2$ suit une loi gamma de paramètre $r+s(\gamma_{r+s})$

3.3. Soient $X_1,X_2,\ldots,X_n$, n variables indépendantes suivant une loi normale $N(m_i,\sigma_i)$, $i=1,2,\ldots n$. Montrer que $X=X_1+X_2+\ldots+X_n$ suit une loi normale $N\left(\sum_{i=1}^{n} m_i \ , \ \sqrt{\sum_{i=1}^{n}\sigma_i^2}\right)$

3.4. Soient $X_1,X_2,\ldots,X_n$, n variables aléatoires indépendantes suivant une loi $N(0,1)$. Montrer que $Z=X_1^2+X_2^2+\ldots+X_n^2$ suit une loi du Khi-carré χ_n^2 ; n étant le nombre de degrés de liberté .

3.5. Soient deux variables aléatoires indépendantes X et Y suivant respectivement des lois normales $N(0,1)$ et du χ_k^2 .

La variable aléatoire $T = \dfrac{X}{\sqrt{\dfrac{Y}{k}}}$ est dite suivre une loi de Student à k degrés de liberté . Trouver la densité de T .

3.6. Soient deux variables aléatoires indépendantes X et Y suivant une loi de χ_p^2 et χ_q^2 respectivement (p et q étant les nombres de degrés de liberté de X et Y) . La variable aléatoire $F(p,q) = \dfrac{X/p}{Y/q}$ est dite suivre une loi de Fisher . Trouver la densité de F .

3.7. On répète une expérience de Bernoulli n fois . Soit X la variable aléatoire exprimant le nombre de succès ; p est la probabilité de succès et q=1-p celle d'échec .
Montrer que pour tout ε positif on a :

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) = 0$$

3.8. Combien faut-il lancer une pièce de monnaie pour que ; avec une probabilité d'au moins 0,95 , la fréquence de pile diffère de 0,5 de moins de 0,1 ?
1) Utiliser l'approximation de la loi normale à la lo binomiale .
2) Utiliser l'inégalité de Bienaymé-Tchebichev

3.9. Résoudre l'exercice 2.3 en utilisant l'approximation de la loi de Poisson à la loi binomiale .

SOLUTIONS DES EXERCICES

3.1. On a par définition $\gamma = \mu_3 / \sigma^3$ et $\delta = \mu_4 / \sigma^4$
Le moment centré d'ordre k est défini par μ_k : : $\mu_k = E\left[(X-m)^k\right]$. Pour une distribution continue $\mu_k = \int_{-\infty}^{\infty} (x-m)^k f(x)dx$ et dans le cas d'une distribution normale N(0,1) on a : $m = 0$ et $\sigma = 1$

En effet $m = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ mais

$$\int_{-\infty}^{0} xe^{-\frac{x^2}{2}} dx = -\int_{0}^{\infty} xe^{-\frac{x^2}{2}} dx \quad \text{donc}$$

$$m = \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{0} x \; e^{-\frac{x^2}{2}} dx + \int_{0}^{\infty} x \; e^{-\frac{x^2}{2}} dx\right] = 0$$

et $$\mu_2 = E\left(X^2\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} x^2 e^{-\frac{x^2}{2}} dx$$

$$= \frac{2}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{2+1}{2}\right)}{2\left(\frac{1}{2}\right)^{\frac{2+1}{2}}} = \frac{2}{\sqrt{2\pi}} \frac{1}{2} \frac{\Gamma\left(\frac{1}{2}\right)}{2\left(\frac{1}{2}\right)^{\frac{3}{2}}} = \frac{2}{\sqrt{2\pi}} \frac{1}{2} \frac{\sqrt{\pi}}{2} 2\sqrt{2} = 1$$

Par conséquent $\sigma = 1$. On a utilisé la relation :

$$\int_0^{\infty} x^{\alpha} e^{-\frac{x^2}{2}} dx = \frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{2\left(\frac{1}{2}\right)^{\frac{\alpha+1}{2}}} \quad \text{où } \alpha > -1$$

et où $\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt$

Si k=3 on a : $\mu_3 = \int_{-\infty}^{\infty} x^3 f(x) dx$. Par symétrie

$$\int_{-\infty}^{0} x^3 e^{-\frac{x^2}{2}} dx = -\int_0^{\infty} x^3 e^{-\frac{x^2}{2}} dx \text{ , de sorte que } \mu_3 = 0$$

Si k =4 on a : $\mu_4 = \int_{-\infty}^{\infty} x^4 f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-\frac{x^2}{2}} dx$

$$\mu_4 = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^4 e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \frac{\Gamma\left(\frac{5}{2}\right)}{2\left(\frac{1}{2}\right)^{\frac{5}{2}}} = \frac{2}{\sqrt{2\pi}} \frac{3}{2} \frac{1}{2} \frac{\Gamma\left(\frac{1}{2}\right)}{2\left(\frac{1}{2}\right)^{\frac{5}{2}}}$$

$$= \frac{2}{\sqrt{2\pi}} \frac{3}{2} \frac{1}{2} \frac{\sqrt{\pi}}{2} 2^{\frac{5}{2}} = 3$$

On a utilisé les propriétés suivantes de la fonction gamma :

$$\Gamma(n+1) = n\Gamma(n); \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}; \Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right)$$

En conclusion : $\gamma = \mu_3 / \sigma^3 = 0 / 1^3 = 0$

$$\delta = \mu_4 / \sigma^4 = 3 / 1^4 = 3$$

3.2. On a :

$$E(X_1) = \frac{1}{\Gamma(r)} \int_0^{\infty} x\, e^{-x} x^{r-1} dx = \frac{1}{\Gamma(r)} \int_0^{\infty} e^{-x} x^r dx = \frac{\Gamma(r+1)}{\Gamma(r)}$$

Comme $\Gamma(r+1) = r\Gamma(r)$, il vient $E(X_1) = r$

$$V(X_1) = E(X_1^2) - (E(X_1))^2 = \frac{1}{\Gamma(r)} \int_0^{\infty} x\, e^{-x} x^{r+1} dx - r^2$$

Mais $$\frac{\int_0^{\infty} e^{-x} x^{r+1} dx}{\Gamma(r)} = \frac{\Gamma(r+2)}{\Gamma(r)} = \frac{(r+1)\Gamma(r+1)}{\Gamma(r)},$$

$$= \frac{(r+1)r\Gamma(r)}{\Gamma(r)} = (r+1)r$$

Donc $V(X_1) = (r+1)r - r^2 = r$

Pour montrer que $Z = X_1 + X_2$ suit une loi gamma γ_{r+s}, on utilise la formule du produit de convolution pour le calcul de la densité de la variable aléatoire $Z = X_1 + X_2$.
Cette densité s'écrit :

$$f(z)=\int_0^z \frac{1}{\Gamma(r)}e^{-x}x^{r-1}\frac{1}{\Gamma(s)}e^{-(z-x)}(z-x)^{s-1}dx$$

$$f(z)=\frac{e^{-z}}{\Gamma(r)\Gamma(s)}\int_0^z x^{r-1}(z-x)^{s-1}dx$$ En posant $x=zu$

on obtient $f(z)=\dfrac{e^{-z}}{\Gamma(r)\Gamma(s)}\displaystyle\int_0^1 (zu)^{r-1}(z-zu)^{s-1}z\ du$

$$=\frac{e^{-z}}{\Gamma(r)\Gamma(s)}z^{r+s-1}\int_0^1 u^{r-1}(1-u)^{s-1}du$$

$$=\frac{e^{-z}}{\Gamma(r)\Gamma(s)}z^{r+s-1}B(r,s)$$

$$=\frac{e^{-z}}{\Gamma(r)\Gamma(s)}z^{r+s-1}\frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

$$=\frac{1}{\Gamma(r+s)}e^{-z}z^{r+s-1}$$

On a utilisé : $\displaystyle\int_0^1 u^{r-1}(1-u)^{s-1}du=B(r,s)=\frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$

3.3. On fait une démonstration par récurrence.
Soient X_1 et X_2 deux variables indépendantes suivant une loi normale $N(m_1,\sigma_1)$ et $N(m_2,\sigma_2)$ respectivement.
On cherche la loi de $X=X_1+X_2$. On démontre pour deux variables aléatoires centrées réduites, c'est à dire pour

$X_1^*=\dfrac{X_1-m_1}{\sigma_1}$ et $X_2^*=\dfrac{X_2-m_2}{\sigma_2}$. X_1^*et X_2^* suivent une loi normale N(0,1).
La densité de $X=X_1^*+X_2^*$ (formule du produit de convolution)est :

$$f(z)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-x)^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{1}{2\pi}e^{-\frac{z^2}{2}}\int_{-\infty}^{\infty}e^{zx-x^2}dx$$

$$=\frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{\infty}e^{-\left(x-\frac{z}{2}\right)^2}dx$$

En posant $u=x-\dfrac{z}{2}$, on obtient :

$$f(z)=\frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{\infty}e^{-u^2}du=\frac{1}{2\pi}e^{-\frac{z^2}{4}}\sqrt{\pi}$$

donc $f(z)=\dfrac{1}{\sqrt{2}\sqrt{2\pi}}e^{-\frac{1}{2}\frac{z^2}{2}}$ qui est la densité d'une variable aléatoire suivant une loi normale $N(0,\sqrt{2})$.
On suppose que la somme $X=X_1+X_2+\ldots+X_{n-1}$ de n-1 variables aléatoires suit une loi normale $N(0,\sqrt{n-1})$ et on montre que ceci est vrai pour l'ordre n , c'est à dire que $X=X_1+X_2+\ldots+X_{n-1}+X_n$ suit une loi normale $N(0,\sqrt{n})$.

La densité de $X_1+X_2+\ldots+X_{n-1}$ est $f(x)=\dfrac{1}{\sqrt{n-1}\sqrt{2\pi}}e^{-\frac{1}{n-1}\frac{x^2}{2}}$

celle de X_n est : $f(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

La densité de $X = X_1+X_2+\ldots+X_{n-1}+X_n$ s'écrit :

$$f(z)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{n-1}\sqrt{2\pi}}e^{-\frac{1}{n-1}\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-x)^2}{2}}dx$$

$$=\frac{1}{2\pi\sqrt{n-1}}\int_{-\infty}^{\infty}e^{-\frac{x^2}{2(n-1)}-\frac{(z-x)^2}{2}}dx$$

$$=\frac{1}{2\pi\sqrt{n-1}}\int_{-\infty}^{\infty}e^{\left(-\frac{x^2}{2}\frac{n}{n-1}+xz-\frac{z^2}{2}\right)}dx$$

$$=\frac{1}{2\pi\sqrt{n-1}}\int_{-\infty}^{\infty}e^{-\left[x\sqrt{\frac{n}{2(n-1)}}-\frac{1}{2}z\sqrt{\frac{2(n-1)}{n}}\right]^2-\frac{z^2}{2n}}dx$$

$$=\frac{1}{2\pi\sqrt{n-1}}e^{-\frac{z^2}{2n}}\int_{-\infty}^{\infty}e^{-\left[x\sqrt{\frac{n}{2(n-1)}}-\frac{1}{2}z\sqrt{\frac{2(n-1)}{n}}\right]^2}dx$$

En posant $x\sqrt{\dfrac{n}{2(n-1)}}-\dfrac{1}{2}z\sqrt{\dfrac{2(n-1)}{n}}=\dfrac{t}{\sqrt{2}}$, il vient

$$f(z)=\frac{e^{-\frac{z^2}{2n}}}{2\pi\sqrt{n-1}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}\frac{1}{\sqrt{2}}\sqrt{\frac{2(n-1)}{n}}dt$$

$$=\frac{e^{-\frac{z^2}{2n}}}{2\pi\sqrt{n-1}}\sqrt{\frac{2(n-1)}{n}}\frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt$$

Comme $\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$

On obtient $f(z)=\dfrac{1}{\sqrt{n}\sqrt{2\pi}}e^{-\frac{z^2}{2n}}$

qui est la densité d'une variable aléatoire suivant une loi normale $N\left(0,\sqrt{n}\right)$

3.4. On désigne par F_i et f_i $(i=1,\ldots,n)$ la fonction de répartition et la densité de probabilité de la variable aléatoire X_i .
La densité de probabilité d'une variable aléatoire suivant une loi du χ_k^2 s'écrit :

$$g(t)=\frac{1}{2\Gamma\left(\frac{k}{2}\right)}\left(\frac{t}{2}\right)^{\frac{k}{2}-1}e^{-\frac{t}{2}}$$

Pour montrer que $Z=X_1^2+X_2^2+\ldots+X_n^2$ suit une loi du χ_n^2 , on utilise un raisonnement par récurrence .

a) Soit X_1 une variable aléatoire normale suivant **une loi** normale N(0,1) . On détermine la loi de probabilité de $Z = X_1^2$.
On a : $G(z)$ fonction de répartition de Z

$$G(z)=P(Z<z)=P\left(X_1^2<z\right)=P\left(-\sqrt{z}<X_1<\sqrt{z}\right)$$

$$= F_1(\sqrt{z}) - F_1(-\sqrt{z}) \qquad ; \quad z > 0$$

La densité de probabilité se Z est :

$$g_1(z) = G'(z) = \frac{1}{2\sqrt{z}} f_1(\sqrt{z}) + \frac{1}{2\sqrt{z}} f_1(-\sqrt{z})$$

$$= \frac{1}{2\sqrt{z}}\left(f_1(\sqrt{z}) + f_1(-\sqrt{z})\right) = \frac{1}{2\sqrt{z}}\left[\frac{1}{\sqrt{2\pi}} e^{-\frac{z}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{z}{2}}\right]$$

$$= \frac{2}{2\sqrt{z}\sqrt{2\pi}} e^{-\frac{z}{2}} = \frac{1}{\sqrt{2\pi}} z^{-\frac{1}{2}} e^{-\frac{z}{2}}$$

On a bien $g_1(z) = \dfrac{1}{2\Gamma\left(\frac{1}{2}\right)}\left(\dfrac{z}{2}\right)^{\frac{1}{2}-1} e^{-\frac{z}{2}}$ avec $\Gamma\left(\dfrac{1}{2}\right) = \sqrt{\pi}$

b) Soit X_2 une variable aléatoire suivant une loi normale N(0,1) . On détermine la loi de $Z = X_1^2 + X_2^2$.
On sait que la densité de X_1^2 est :

$$g_1(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$$

La densité de X_2^2 s'écrit de la même façon :

$$g_1(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

La densité de $Z = X_1^2 + X_2^2$ s'écrira alors :

$$g_2(z) = \int_0^z g_1(y) g_1(z-y) dy$$

$$= \int_0^z \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \frac{1}{\sqrt{2\pi}} (z-y)^{-\frac{1}{2}} e^{-\frac{(z-y)}{2}} dy$$

$$= \frac{e^{-\frac{z}{2}}}{2\pi} \int_0^z y^{-\frac{1}{2}} (z-y)^{-\frac{1}{2}} dy$$

En posant $y = tz$, on obtient :

$$g_2(z) = \frac{1}{2\pi} e^{-\frac{z}{2}} \int_0^1 (tz)^{-\frac{1}{2}} (1-t)^{-\frac{1}{2}} z^{-\frac{1}{2}} z \; dt$$

$$g_2(z) = \frac{1}{2\pi} e^{-\frac{z}{2}} \int_0^1 t^{-\frac{1}{2}} (1-t)^{-\frac{1}{2}} \; dt = \frac{e^{-\frac{z}{2}}}{2\pi} \int_0^1 \frac{dt}{\sqrt{t(1-t)}}$$

En posant $t = Cos^2\theta$, il vient :

$$g_2(z) = \frac{e^{-\frac{z}{2}}}{2\pi} \int_0^{\frac{\pi}{2}} \frac{2 Cos\theta \sin\theta \; d\theta}{\sqrt{Cos^2\theta\left(1 - Cos^2\theta\right)}} = \frac{e^{-\frac{z}{2}}}{2\pi} \int_0^{\pi/2} 2 d\theta = \frac{1}{2} e^{-\frac{z}{2}}$$

On a bien $g_2(z) = \dfrac{1}{2\Gamma\left(\frac{2}{2}\right)}\left(\dfrac{z}{2}\right)^{\frac{2}{2}-1} e^{-\frac{z}{2}}$ avec $\Gamma\left(\dfrac{1}{2}\right) = 1$

c) Soit X_3 une variable aléatoire suivant une loi normale N(0,1).

On détermine la loi de $Z = X_1^2 + X_2^2 + X_3^2$

La densité de X_3^2 est :

$$g_1(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$$

celle de $X_1^2 + X_2^2$ **est :**

$$g_2(y) = \frac{1}{2} e^{-\frac{y}{2}}$$

La densité de Z s'écrit **alors :**

$$g_3(z) = \int_0^z g_2(y) g_1(z-y) dy$$

$$= \int_0^z \frac{1}{2} e^{-\frac{y}{2}} \frac{1}{\sqrt{2\pi}} (z-y)^{-\frac{1}{2}} \; e^{-\frac{(z-y)}{2}} dy$$

$$= \frac{1}{2\sqrt{2\pi}} e^{-\frac{z}{2}} \int_0^z (z-y)^{-\frac{1}{2}} \; dy$$

En posant $y = tz$, on obtient :

$$g_3(z) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{z}{2}} \int_0^1 z^{-\frac{1}{2}} (1-t)^{-\frac{1}{2}} \; z \; dt$$

$$= \frac{1}{2\sqrt{2\pi}} e^{-\frac{z}{2}} z^{\frac{1}{2}} \int_0^1 (1-t)^{-\frac{1}{2}} \; dt = \frac{1}{\sqrt{2\pi}} e^{-\frac{z}{2}} z^{\frac{1}{2}}$$

On a bien :

$$g_3(z) = \frac{1}{2\Gamma\left(\frac{3}{2}\right)} \left(\frac{z}{2}\right)^{\frac{3}{2}-1} e^{-\frac{z}{2}}$$

avec $\Gamma\left(\frac{3}{2}\right) = \Gamma\left(\frac{1}{2}+1\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$

d) On suppose que la variable aléatoire $U = X_1^2 + X_2^2 + \ldots + X_{n-1}^2$ suit une loi du χ_{n-1}^2 et on montre que la variable aléatoire $Z = U + X_n^2$ suit une loi du χ_n^2.
On détermine la loi de Z.

La densité de χ_n^2 est :

$$g_1(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}} \; ;$$

celle de U est :

$$g_{n-1}(y) = \frac{1}{2\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{y}{2}\right)^{\frac{n-1}{2}-1} e^{-\frac{y}{2}}$$

On a alors : $g_n(z) = \int_0^z g_{n-1}(y) g_1(z-y) dy$

$$g_n(z) = \int_0^z \frac{1}{2\Gamma\left(\frac{n-1}{2}\right)} \left(\frac{y}{2}\right)^{\frac{n-1}{2}-1} e^{-\frac{y}{2}} \frac{1}{\sqrt{2\pi}} (z-y)^{-\frac{1}{2}} e^{-\frac{(z-y)}{2}} dy$$

$$= \frac{e^{-\frac{z}{2}}}{2\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}} \int_0^z \left(\frac{y}{2}\right)^{\frac{n-1}{2}-1} (z-y)^{-\frac{1}{2}} dy$$

En posant $y = tz$, on obtient :

$$g_n(z) = \frac{e^{-\frac{z}{2}}}{2\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}} \int_0^1 \frac{t^{\frac{n-1}{2}-1} z^{\frac{n-1}{2}-1}}{2^{\frac{n-1}{2}-1}} z^{-\frac{1}{2}} (1-t)^{-\frac{1}{2}} \; z \; dt$$

$$= \frac{e^{-\frac{z}{2}} z^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}\, 2^{\frac{n-1}{2}-1}} \int_0^1 t^{\frac{n-1}{2}-1} (1-t)^{-\frac{1}{2}} dt$$

$$= \frac{e^{-\frac{z}{2}} z^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}\, 2^{\frac{n-1}{2}-1}} B\left(\frac{n-1}{2}, \frac{1}{2}\right)$$

$$= \frac{e^{-\frac{z}{2}} z^{\frac{n}{2}-1}}{2\Gamma\left(\frac{n-1}{2}\right)\sqrt{2\pi}\, 2^{\frac{n-1}{2}-1}} \frac{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

car $\quad B(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$

Sachant que $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, il vient

$$g_n(z) = \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{z}{2}\right)^{\frac{n}{2}-1} e^{-\frac{z}{2}}$$

3.5. On cherche la densité de la variable $\quad Z = \sqrt{\dfrac{Y}{k}}$

La fonction de répartition $\; F(z)$ de Z est :

$$F(z) = P(Z < z) = P\left(\sqrt{\frac{Y}{k}} < z\right) = P\left(Y < kz^2\right)$$

$$= \int_0^{kz^2} \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} e^{-\frac{t}{2}} t^{\frac{k}{2}-1} dt$$

En posant $u = \sqrt{\dfrac{t}{k}}$ on obtient

$$F(z) = \int_0^z \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} e^{-\frac{1}{2}ku^2} \left(ku^2\right)^{\frac{k}{2}-1} k\,2u\,du$$

La densité $\; f(z)$ de $Z = \sqrt{\dfrac{Y}{k}}$ s'écrit alors :

$$f(z) = \frac{2kz}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} e^{-\frac{kz^2}{2}} \left(kz^2\right)^{\frac{k}{2}-1}$$

On détermine ensuite la densité $\; g(t)$ de $T = \dfrac{X}{Z}$

La fonction de répartition de T est :

$$G(t) = P(T < t) = \int_0^\infty f(z)dz \int_{-\infty}^{tz} h(x)dx$$

où $h(x)$ est la densité de probabilité de X et $f(z)$ est la densité de probabilité de $Z = \sqrt{\frac{Y}{k}}$

On a : $g(t) = \int_0^\infty f(z)h(tz)z \ dz$

$$= \int_0^\infty \frac{2kz \ e^{-\frac{kz^2}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}\left(kz^2\right)^{\frac{k}{2}-1}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2z^2}{2}} z \ dz$$

$$g(t) = \frac{k \ k^{\frac{k}{2}-1}}{2^{\frac{k}{2}-1}\Gamma\left(\frac{k}{2}\right)}\frac{1}{\sqrt{2\pi}}\int_0^\infty \left(z^2\right)^{\frac{k-1}{2}} e^{-\frac{z^2}{2}\left(t^2+k\right)} z \ dz$$

En posant $\frac{z^2}{2}\left(t^2+k\right) = v$, on obtient

$$g(t) = \frac{k^{\frac{k}{2}}}{2^{\frac{k}{2}-1}\Gamma\left(\frac{k}{2}\right)\sqrt{\pi}}\int_0^\infty \left(\frac{2v}{t^2+k}\right)^{\frac{k-1}{2}} e^{-v} \frac{dv}{\left(t^2+k\right)}$$

$$= \frac{k^{\frac{k}{2}}}{2^{\frac{k-1}{2}}\Gamma\left(\frac{k}{2}\right)\sqrt{\pi}}\frac{2^{\frac{k-1}{2}}}{\left(t^2+k\right)^{\frac{k+1}{2}}}\int_0^\infty v^{\frac{k-1}{2}} e^{-v} \ dv$$

Mais $\int_0^\infty v^{\frac{k-1}{2}} e^{-v} \ dv = \int_0^\infty v^{\frac{k+1}{2}-1} e^{-v} \ dv = \Gamma\left(\frac{k+1}{2}\right)$

et $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ d'où

$$g(t) = \frac{k^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right)}\Gamma\left(\frac{k+1}{2}\right)\frac{1}{\left(t^2+k\right)^{\frac{k+1}{2}}} = \frac{k^{\frac{k}{2}}}{B\left(\frac{k}{2},\frac{1}{2}\right)\left(t^2+k\right)^{\frac{k+1}{2}}}$$

$$g(t) = \frac{1}{\sqrt{k} \ B\left(\frac{k}{2},\frac{1}{2}\right)\left(1+\frac{t^2}{k}\right)^{\frac{k+1}{2}}}$$

3.6. On cherche les densité $f(x)$ et $g(y)$ des variables aléatoires $X_1 = \frac{X}{p}$ et $Y_1 = \frac{Y}{q}$

Soit $F(x)$ la fonction de répartition de X_1 . On a :

$$F(x) = P(X_1 < x) = P\left(\frac{X}{p} < x\right) = P(X < px)$$

$$= \int_0^{px} \frac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} e^{-\frac{t}{2}} t^{\frac{p}{2}-1} dt$$

En posant $\frac{t}{p} = u$, on obtient

$$F(x) = \int_0^x \frac{1}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} e^{-\frac{pu}{2}} (pu)^{\frac{p}{2}-1} \, pdu$$

La densité $f(x)$ s'écrit :

$$f(x) = \frac{p}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} e^{-\frac{px}{2}} (px)^{\frac{p}{2}-1}$$

On obtient de la même façon la densité de Y_1

$$g(y) = \frac{q}{2^{\frac{q}{2}}\Gamma\left(\frac{q}{2}\right)} e^{-\frac{qy}{2}} (qy)^{\frac{q}{2}-1}$$

On détermine ensuite la densité $g(z)$ de $F = \frac{X_1}{Y_1}$

La fonction de répartition de F est :

$$G(z) = P(F < z) = \int_0^\infty g(y)dy \int_{-\infty}^{zy} f(x)dx$$

$g(z)$ s'écrit : $g(z) = \int_0^\infty g(y) f(zy) y \, dy$

$$g(z) = \int_0^\infty \frac{q}{2^{\frac{q}{2}}\Gamma\left(\frac{q}{2}\right)} e^{-\frac{qy}{2}} (qy)^{\frac{q}{2}-1} \frac{p}{2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right)} e^{-\frac{pzy}{2}} (pzy)^{\frac{p}{2}-1} ydy$$

$$= \frac{q^{\frac{q}{2}} p^{\frac{p}{2}}}{2^{\frac{q}{2}} 2^{\frac{p}{2}} \Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} z^{\frac{p}{2}-1} \int_0^\infty e^{-\frac{y}{2}(q+pz)} y^{\frac{p+q}{2}-1} \, dy$$

En posant $\frac{y}{2}(q+pz) = t$, on obtient

$$g(z) = \frac{q^{\frac{q}{2}} p^{\frac{p}{2}}}{2^{\frac{q}{2}} 2^{\frac{p}{2}} \Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} z^{\frac{p}{2}-1} \int_0^\infty e^{-t} \left(\frac{2t}{q+pz}\right)^{\frac{p+q}{2}-1} \frac{2dt}{q+pz}$$

$$= \frac{q^{\frac{q}{2}} p^{\frac{p}{2}}}{2^{\frac{q}{2}} 2^{\frac{p}{2}} \Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} z^{\frac{p}{2}-1} \frac{2^{\frac{p+q}{2}}}{(q+pz)^{\frac{p+q}{2}}} \int_0^\infty e^{-t} \, t^{\frac{p+q}{2}-1} \, dt$$

$$= \frac{q^{\frac{q}{2}} p^{\frac{p}{2}}}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)} \frac{z^{\frac{p}{2}-1}}{(q+pz)^{\frac{p+q}{2}}} \Gamma\left(\frac{p+q}{2}\right)$$

$$= \frac{q^{\frac{q}{2}} p^{\frac{p}{2}}}{B\left(\frac{p}{2}, \frac{q}{2}\right)} \frac{z^{\frac{p}{2}-1}}{(q+pz)^{\frac{p+q}{2}}}$$

3.7. X suit une loi binomiale , on a $E(X)=np$ et $V(X)=npq$
L'utilisation de l'inégalité de Bienaymé-Tchebichev donne :

$$P\left(|X-np| \geq t\sqrt{npq}\right) \leq \frac{1}{t^2}$$

ou encore $$P\left(\left|\frac{X}{n}-p\right| \geq t\sqrt{\frac{pq}{n}}\right) \leq \frac{1}{t^2}$$

En posant $\varepsilon = t\sqrt{\frac{pq}{n}}$ on obtient :

$$P\left(\left|\frac{X}{n}-p\right| \geq \varepsilon\right) \leq \frac{pq}{n\varepsilon^2} \quad \text{et}$$

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n}-p\right| \geq \varepsilon\right) = 0$$

3.8. Soit la variable aléatoire $X_i \quad (i=1,2,\dots)$

$$X_i = \begin{cases} 1 & \text{si pile est apparu au cours de l' expérience } i \\ 0 & \text{si face est apparu au cours de l' expérience } i \end{cases}$$

On a :$P(X_i=1)=P(X_i=0)=\frac{1}{2}$

$$E(X_i)=p=\frac{1}{2}$$

$$V(X_i)=p(1-p)=\frac{1}{4}$$

Par $f=\frac{\sum_{i=1}^{n} X_i}{n}$ on désigne la variable aléatoire ,**fréquence** d'apparitions de pile au cours de n expériences ..
On a :

$$\lim_{n\to\infty} P\left(|f-p| < t\sqrt{\frac{p(1-p)}{n}}\right) = \frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx = 0,95$$

car pour une variable binomiale $\sigma = \sqrt{\frac{p(1-p)}{n}}$ on choisit t

tel que $\frac{2}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx = 0,95$ ou $\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{x^2}{2}}\,dx = 0,475$

La valeur de t est obtenue à partir de la table de la loi normale N(0,1):t=1,96

Si n est suffisamment grand on a approximativement

$$P\left(\left|f-\frac{1}{2}\right| < t\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = P\left(\left|f-\frac{1}{2}\right| < \frac{0,98}{\sqrt{n}}\right) = 0,95$$

On doit avoir $\frac{0,98}{\sqrt{n}} = 0,1$ d'où n=97

Il faut lancer au moins 97 fois la pièce de monnaie pour avoir 95% de chances d'obtenir $p=\frac{1}{2}$ à 0,1 près .

3.9. L'application de la loi binomiale donne (voir exercice 2.3) :

$$P(X=2)=0,196 \quad \text{et} \quad P(X=10)=0,001$$

Nous sommes dans les conditions d'approximation par une loi de Poisson puisque p=0,03 < 0,1 et n=112 > 50 . D'où

$$\lambda = np = 112.0,03 = 3,36$$

et $$P(X=2)=e^{-3,36}\frac{(3,36)^2}{2!}=0,196$$

et $$P(X=10)=e^{-3,36}\frac{(3,36)^{10}}{10!}=0,0017$$

CHAPITRE 4

ECHANTILLONNAGE-ESTIMATION

I-INTRODUCTION

Pour avoir des informations sur une population on peut examiner chaque individu de cette population . On peut aussi examiner un échantillon représentatif de cette population . Cette deuxième méthode a l'avantage de pouvoir s'appliquer car la taille de l'échantillon est relativement réduite . Elle est le plus souvent utilisée car la première est coûteuse ou parfois impossible à appliquer .

La théorie de l'échantillonnage est l'étude des liens existants entre les caractéristiques (en général la moyenne et l'écart-type) de la population et ceux des échantillons de cette population . Les échantillons de la population doivent être représentatifs .

La théorie de l'estimation est l'étude des liens qui existent entre les caractéristiques connues d'échantillons et ceux correspondants de la population . La notion de précision est très importante dans ce cas : on détermine un intervalle dans lequel se trouve le paramètre à estimer avec une probabilité fixée .

II-DISTRIBUTION D'ÉCHANTILLONNAGE DE LA MOYENNE

Soit une population de taille N . On désigne par m et σ la moyenne et l'écart-type de cette population respectivement . On extrait de la population une série d'échantillons de taille n. Chacun de ces échantillons a une moyenne $\bar{x}$. Les différentes moyennes obtenues constituent une distribution d'échantillonnage de moyenne : $\overline{X}$ (les différentes moyennes sont considérées comme des réalisations d'une variable aléatoire $\overline{X}$) . On désigne par $\mu_{\overline{X}}$ et $\sigma_{\overline{X}}$ la moyenne et l'écart-type de la distribution d'échantillonnage de la moyenne.

On a : $\mu_{\overline{X}} = m$

Si le tirage est exhaustif (sans remise) $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

Dans le cas où la population est infinie ou si le tirage est non exhaustif (avec remise) on a : $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Si n est petit devant N , la distinction entre exhaustivité et non exhaustivité est sans objet car $\frac{N-n}{N-1} \approx 1$

Si la taille n des échantillons est assez grande (en pratique $n \geq 30$) , la distribution d'échantillonnage de la moyenne approche la distribution normale quelle que soit la distribution de la population.

Si la population est normalement distribuée , la distribution d'échantillonnage de la moyenne est une loi normale quelle que soit la valeur n de la taille des échantillons.

III DISTRIBUTION D'ÉCHANTILLONNAGE DES FRÉQUENCES

La probabilité de réalisation d'un événement A est égale à p . On considère les échantillons aléatoires de tailles n extraits d'une population de taille N . Pour chaque échantillon on détermine la proportion f de réalisation de l'événement A . La population et les échantillons suivent des lois binomiales $B(N,p)$ et $B(N,f)$ respectivement.

La moyenne μ_f et l'écart-type σ_f de la distribution d'échantillonnage des fréquences valent :

$\mu_f = p$ et

$\sigma_f = \sqrt{\frac{p(1-p)}{n}}$ si le tirage est non exhaustif ou si la population est infinie ou

$\sigma_f = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$ si le tirage est exhaustif.

Si $n \geq 30$ la distribution d'échantillonnage des **fréquences** approche la distribution normale.

IV-DISTRIBUTION D'ÉCHANTILLONNAGE DE DIFFÉRENCES DES MOYENNES

Soient deux populations P_1 et P_2 (moyennes m_1 et m_2 et écarts-type σ_1 et σ_2 respectivement) . On extrait de chacune des deux populations un échantillon de taille n_1 (de P_1) et n_2 (de P_2)

La moyenne et l'écart-type de la distribution de la différence des moyennes des deux populations sont :

$\mu_{\overline{X}_1-\overline{X}_2} = m_1 - m_2$ et

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{(tirage non exhaustif)}$$

V-DISTIBUTION DE LA VARIANCE

Chaque échantillon de taille n de la population a une variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$. Ces variances sont des valeurs observées d'une même variable aléatoire.

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

On a : $E(S^2) = \frac{n-1}{n}\sigma^2$

$$V(S^2)=\frac{n-1}{n^3}\left[(n-1)\mu_4-(n-3)\sigma^4\right]$$

et $Cov(\overline{X},S^2)=\frac{n-1}{n^2}\mu_3$

VI-ESTIMATEURS

A partir de l'examen d'un échantillon , on essaye d'obtenir une information quantitative sur des paramètres de la population à estimer (en général m et σ)
L'estimation est dite ponctuelle si on estime un paramètre de la population avec un seul nombre .
L'estimation est dite par intervalle si on estime un paramètre inconnu θ par une construction d'un intervalle $]a,b[$ qui contient ce paramètre avec une probabilité α fixée : $P(a<\theta<b)=\alpha$, a et b sont dits limites de confiance et α seuil (ou niveau) de confiance .
T est un estimateur de θ si T converge en moyenne quadratique vers θ :

$$E(T)\to\theta$$

$$V(T)\to 0 \qquad \text{quand } n\to\infty$$

Un estimateur T de θ est dit sans biais si : $E(T)=\theta$

$\overline{X}$ est , par exemple , un estimateur ponctuel non biaisé de m car $E(\overline{X})=\mu_{\overline{X}}=m$

S^2 est un estimateur biaisé de σ^2 car : $E(S^2)=\frac{n-1}{n}\sigma^2$

Si on utilise la variance $S'^2=\frac{n}{n-1}S^2$ on a : $E(S'^2)=\sigma^2$, et S'^2 est un estimateur sans biais de σ^2

VII-ESTIMATEURS PONCTUELS -MÉTHODE DU MAXIMUM DE VRAISEMBLANCE

Soit un paramètre θ (en général m ou σ) d'une population à estimer . Il faut trouver un estimateur T , à partir d'un échantillon , qui reflète le mieux le paramètre considéré .

a) $E(T)=\theta$

Par exemple $\overline{X}$ et S'^2 sont des estimateurs sans biais de m et σ^2 respectivement .

b) $P(|\theta-T|\le\varepsilon)\to 1$ quand $n\to\infty$

$\overline{X}$ converge en probabilité vers m
On a : $V(\overline{X})\to 0$ quand $n\to\infty$

c) La taille n de l'échantillon étant fixé , parmi les estimateurs T de θ on choisit celui dont la variance est la plus petite .
La méthode du maximum de vraisemblance consiste à choisir comme estimateur de θ la valeur particulière de θ qui maximise la fonction de vraisemblance :

$$L(x_1,x_2,\ldots,x_n,\theta)$$

L'estimateur $L(x_1,x_2,\ldots,x_n)$ est la fonction des observations qui maximise $L(x_1,x_2,\ldots,x_n,\theta)$ par rapport à tous les θ possibles .
Cet estimateur T est solution de l'équation :

$$\frac{\partial L(x_1,x_2,\ldots x_n,\theta)}{\partial\theta}=0 \quad \text{ou} \quad \frac{\partial LnL(x_1,x_2,\ldots x_n,\theta)}{\partial\theta}=0$$

VIII-INTERVALLE DE CONFIANCE AVEC LA DISTRIBUTION NORMALE

1- Intervalle de confiance de la moyenne .

On veut estimer la moyenne m d'une population normale à l'aide d'un échantillon aléatoire . (Si la taille de l'échantillon n est telle que $n \geq 30$ la distribution d'échantillonnage de la moyenne est normale quelle que soit la distribution de la population)

$\overline{X}$ suit une loi normale $N\left(m, \sigma_{\overline{X}}\right)$ avec

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \text{ (tirage non exhaustif) ou}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \text{ (tirage exhaustif) N: taille de la population}$$

On a : $\dfrac{\overline{X}-m}{\sigma_{\overline{X}}}$ suit une loi normale N(0,1)

On cherche un intervalle centré sur m avec une probabilité égale à α :

$$P\left(\frac{\left|\overline{X}-m\right|}{\sigma_{\overline{X}}} < u_\alpha\right) = \alpha$$

donc $\quad m \in \left]\overline{X} - u_\alpha \sigma_{\overline{X}}, \overline{X} + u_\alpha \sigma_{\overline{X}}\right[$

avec une probabilité égale à α .

Dans le cas où σ (écart-type de la population) est inconnu on utilise l'estimation S' (ou S) de σ

(L'échantillon étant par hypothèse de taille $n > 30$)

Si $\alpha = 0,95$ on a : $u_\alpha = 1,96$

2- Intervalle de confiance d'une fréquence

Dans une population on estime la proportion p d'un certain caractère à partir de la proportion f trouvée dans un échantillon de taille n

On a : $p \in \left]f - u_\alpha \sigma_f \ , f + u_\alpha \sigma_f\right[$

avec $\sigma_f = \sqrt{\dfrac{p(1-p)}{n}}$ (tirage non exhaustif) ou

$$\sigma_f = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}} \text{ (tirage exhaustif)}$$

N: taille de la population

Si n est grand ($n > 30$) , on estime p par f pour le calcul des limites de confiance .

Ces résultats sont dus au fait que si Z est le nombre d'éléments de l'échantillon qui présentent le caractère étudié on a :

$$P(Z = k) = C_n^k p^k (1-p)^{n-k}$$

Si la taille de l'échantillon est assez grande , Z suit une loi normale ainsi que $f = \dfrac{Z}{n}$ suit une loi normale $N\left(\mu_f, \sigma_f\right)$avec :

$$\mu_f = p \text{ et } \sigma_f = \sqrt{\frac{p(1-p)}{n}} \text{ (tirage non exhaustif)}$$

d'où $\dfrac{f-\mu_f}{\sigma_f} = \dfrac{f-p}{\sigma_f}$ suit une loi normale N(0,1) et $p \in \left]f - u_\alpha \sigma_f \, , f + u_\alpha \sigma_f\right[$

f étant un estimateur sans biais de p,

$$\left(E(f) = \mu_f = E\left(\frac{Z}{n}\right) = \frac{1}{n} E(Z) = \frac{1}{n} np = p \right) ,$$

on estime p par f dans le calcul des limites de confiance d'où:

$$p \in \left] f - u_\alpha \sqrt{\frac{f(1-f)}{n}} \ , \ f + u_\alpha \sqrt{\frac{f(1-f)}{n}} \right[$$

IX-Intervalle de confiance de différence des moyennes

On extrait un échantillon de taille n_i de la population P_i (i=1,2) de moyenne m_i et d'écart-type σ_i .
Si le tirage est non exhaustif on a :

$$m_1 - m_2 = \mu_{\overline{X}_1} - \mu_{\overline{X}_2}$$
$$= \mu_{\overline{X}_1 - \overline{X}_2} \in \left] \overline{X}_1 - \overline{X}_2 - u_\alpha \sigma_{\overline{X}_1 - \overline{X}_2} \, , \overline{X}_1 - \overline{X}_2 + u_\alpha \sigma_{\overline{X}_1 - \overline{X}_2} \right[$$

avec $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma^2_{\overline{X}_1} + \sigma^2_{\overline{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

X-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 .

distribution t de Student pour déterminer l'intervalle de confiance de la moyenne .

Il existe pour chaque valeur de n une distribution t particulière , alors qu'il n'existe qu'une seule distribution normale réduite .

$t = \dfrac{\overline{X} - m}{S} \sqrt{n-1} = \dfrac{\overline{X} - m}{S' / \sqrt{n}}$ suit une loi de Student à n-1degrés de liberté
Le nombre de degrés de liberté ν est défini par $\nu = n - 1$.
A partir de l'intervalle $\left]-u_\alpha , u_\alpha\right[$ pour lequel :

$$p\left(\left| \frac{\overline{X} - m}{S' / \sqrt{n}} \right| < u_\alpha \right) = \alpha$$

on déduit : $\overline{X} - u_\alpha \dfrac{S'}{\sqrt{n}} < m < \overline{X} + u_\alpha \dfrac{S'}{\sqrt{n}}$

si $\alpha = 0,95$; $\nu = 9$ on a : $u_\alpha = 2,262$
Dès que n>30 , la loi de Student tend à se confondre avec la loi normale .

EXERCICES

4.1. Montrer que la moyenne et l'écart-type de la distribution d'échantillonnage de la moyenne sont:

$$\mu_{\overline{X}} = m \quad \text{et} \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \quad \text{(tirage non exhaustif)}$$

$$\text{ou} \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \quad \text{(tirage exhaustif)}$$

où N, m, σ sont la taille , la moyenne et l'écart-type de la population et n la taille de l'échantillon .

4.2. Calculer la covariance de la moyenne et de la variance des échantillons extraits d'une population ayant une distribution :

a) quelconque
b) de Poisson

On suppose que les échantillons sont de taille n .

4.3. On choisit au hasard sans remise six nombres parmi les nombres entiers de 1 à 9 ; chacun de ces nombres a la même probabilité d'être choisi . Calculer la moyenne et l'écart-type de la distribution d'échantillonnage des moyennes .

4.4. On choisit au hasard avec remise six nombres parmi les nombres entiers de 1 à 9 ; chacun de ces nombres a la même probabilité d'être choisi . Calculer la moyenne et l'écart-type de la distribution d'échantillonnage des moyennes .

4.5. La moyenne des notes d'une épreuve de mathématiques de 300 étudiants est égale à 9,8 . L'écart-type vaut 3,68 .

Trouver la probabilité qu'un échantillon aléatoire de notes de 40 étudiants extrait de l'ensemble ait une moyenne :

a) comprise entre 10 et 13
b) inférieure à 10

dans les deux cas :

1) l'échantillon est exhaustif
2) l'échantillon est non exhaustif

4.6. Soit une variable aléatoire X suivant une loi de Poisson . Estimer le paramètre λ de la loi .

4.7. Soit une variable aléatoire X suivant une loi normale $N(m, \sigma^2)$. Estimer les paramètres m et σ^2 de la loi .

4.8. Soit une population suivant une loi binomiale . On extrait un échantillon de taille n . Construire un estimateur pour la proportion p en utilisant la méthode du maximum de vraisemblance .

4.9. Trouver les valeurs u_α correspondantes au seuil de confiance égal à α=95% en utilisant :

a) la loi normale
b) la distribution t avec i) $\nu = 4$
ii) $\nu = 30$

4.10. Soit une variable aléatoire X ayant pour moyenne m et pour variance σ^2 . Montrer que la variance d'un échantillon de taille n , $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ est un estimateur asymptotiquement sans biais de σ^2 . En déduire un estimateur sans biais de σ^2 .

4.11. a) Dans un échantillon de taille n , montrer que les bornes de l'intervalle de confiance de la proportion de succès au seuil de confiance α sont :

$$p = \frac{\left(2f + \frac{u_\alpha}{n}\right) \pm \sqrt{\frac{u_\alpha^4}{n^2} + 4f\frac{u_\alpha^2}{n} - 4f^2\frac{u_\alpha^2}{n}}}{2\left(1 + \frac{u_\alpha^2}{n}\right)}$$

où f et p sont les proportions de succès dans un échantillon et dans la population respectivement et u_α la valeur correspondante au seuil de confiance α.

b) Si n est grand , montrer que la formule précédente s'écrit :

$$p = f \pm u_\alpha\sqrt{\frac{f(1-f)}{n}}$$

4.12. La taille moyenne d'un échantillon aléatoire de 40 personnes extrait d'une population de 780 individus est de 1,70m . L'écart-type pour toute la population vaut 24cm . Trouver l'intervalle de confiance pour la taille moyenne de la population à 95% .

4.13. 500 étudiants se présentent à un examen de mathématiques . Un échantillon aléatoire de 38 notes donne une moyenne égale à 8,65 et un écart-type égal à 2,82 . Trouver l'intervalle de confiance pour la moyenne des notes de la population :

a) à 90% ; b) à 95% ; c) à 99%

4.14. Les résultats d'un institut de sondage , sur la base d'un échantillon de 58 personnes , indiquent que le candidat A remportera dans sa circonscription 52% des voix au suffrage des électeurs .

a) Trouver l'intervalle de confiance de la proportion des électeurs votant pour le candidat A à un seuil de confiance de 95%

b) Quelle doit être la taille minimale de l'échantillon si on veut que le candidat A remportera plus de 50% des voix à un seuil de confiance de 95%?

4.15. Les notes de mathématiques d'une promotion nombreuse d'étudiants sont normalement distribuées . De cette promotion on en extrait un échantillon de 9 notes . La note moyenne est 9,55 avec écart-type 3,65 . Trouver l'intervalle de confiance de la moyenne de l'ensemble des notes à : a) 95%
b) 99%

Quels seront les résultats si on applique la méthode des grands échantillons ?

4.16. Un échantillon de 180 composants électroniques produits par une usine A ont une durée de vie moyenne égale à 3400 heures avec un écart-type de 200 heures . Un autre échantillon de 130 composants produits par une usine B ont une durée de vie moyenne égale à 3000 heures avec un écart-type de 230 heures . Trouver les limites de confiance à 95% de la différence des durées de vie des composants produits par les deux usines .

4.17. Soit un échantillon aléatoire de taille n extrait d'une population qui suit une loi normale $N(m, \sigma)$. Montrer que la variable aléatoire $t = \frac{\overline{X} - m}{S}\sqrt{n-1}\left(= \frac{\overline{X} - m}{S'}\sqrt{n}\right)$ suit une loi de Student à n-1 degrés de liberté .

SOLUTIONS DES EXERCICES

4.1. Soit un échantillon de taille n . Il est décrit par les valeurs $x_i (i = 1, 2, \dots, n)$ prises par n variables aléatoires X_i . Ces variables aléatoires X_i ont même distribution que la population . La moyenne de la population étant égale à m , on a :

$$E(X_i) = m \text{ et } \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad ; \text{ alors}$$

$$\mu_{\overline{X}} = E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n} nm$$

Si le tirage est non exhaustif (**avec remise**) ou si la population est de taille infinie , les **variables** aléatoires X_i sont indépendantes et ont pour **variance** σ^2 , variance de la population .

On a : $\sigma^2_{\overline{X}} = V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} n\sigma^2$$

d'où $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Si le tirage est **exhaustif (sans remise**) , la population étant de taille N , il y a C_N^n **échantillons de** taille n et les variables aléatoires X_i ne sont **pas indépendantes**

On a : $\sigma^2_{\overline{X}} = V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^{n} X_i\right)$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{n} V(X_i) + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} Cov(X_i, X_j)\right]$$

avec $\sum_{i=1}^{n} V(X_i) = n\sigma^2$

et $Cov(X_i, X_j) = E[(X_i - m)(X_j - m)]$

$$= \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l - m)(x_k - m)P(X_i = x_l, X_j = x_k)$$

$$= \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l - m)(x_k - m)P(X_i = x_l)P(X_j = x_k / X_i = x_l)$$

$$= \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l - m)(x_k - m)\frac{1}{N}P(X_j = x_k / X_i = x_l)$$

$$Cov(X_i, X_j) = \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l - m)(x_k - m)\frac{1}{N}\frac{1}{N-1} \quad ; \text{ si } k \neq l$$

$$= 0 \quad ; \text{ si } k = l$$

Donc $Cov(X_i, X_j) = \dfrac{1}{N}\dfrac{1}{N-1}\sum_{\substack{l,k=1 \\ l \neq k}}^{N}(x_l - m)(x_k - m)$

Comme

$$\left[\sum_{i=1}^{N}(x_i-m)\right]^2=\sum_{i=1}^{N}(x_i-m)^2+\sum_{\substack{l,k=1\\l\neq k}}^{N}(x_l-m)(x_k-m)$$

et

$$\left[\sum_{i=1}^{N}(x_i-m)\right]^2=[Nm-Nm]^2=0$$

et

$$\sum_{i=1}^{N}(x_i-m)^2=N\sigma^2$$

On obtient

$$\sum_{\substack{l,k=1\\l\neq k}}^{N}(x_l-m)(x_k-m)=-N\sigma^2$$

et $$Cov\left(X_i,X_j\right)=\frac{1}{N}\frac{1}{N-1}\left(-N\sigma^2\right)=\frac{-\sigma^2}{N-1}$$

On obtient finalement :

$$\sigma_{\overline{X}}^2=\frac{1}{n^2}\left[n\sigma^2+\sum_{\substack{i,j=1\\i\neq j}}^{n}\left(\frac{-\sigma^2}{N-1}\right)\right]=\frac{1}{n^2}\left[n\sigma^2-\frac{\sigma^2}{N-1}n(n-1)\right]$$

$$=\frac{\sigma^2}{n^2}n\left[1-\frac{1}{N-1}(n-1)\right]$$

d'où $\sigma_{\overline{X}}=\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

4.2. a) On a :

$$Cov\left(\overline{X},S^2\right)=E\left[\left(\overline{X}-m\right)\left(S^2-\frac{n-1}{n}\sigma^2\right)\right]$$

car

$$E\left(\overline{X}\right)=m$$

et

$$E\left(S^2\right)=E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right]=E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2-\left(\overline{X}-m\right)^2\right]$$

$$=\frac{1}{n}\sum_{i=1}^{n}E(X_i-m)^2-E\left(\overline{X}-m\right)^2$$

$$=\sigma^2-\frac{\sigma^2}{n}=\frac{n-1}{n}\sigma^2$$

d'où

$$Cov\left(\overline{X},S^2\right)=E\left[\left(\overline{X}-m\right)S^2\right]-E\left[\left(\overline{X}-m\right)\frac{n-1}{n}\sigma^2\right]$$

$$=E\left[\left(\overline{X}-m\right)S^2\right]$$

car

$$=E\left[\left(\overline{X}-m\right)\frac{n-1}{n}\sigma^2\right]=\frac{n-1}{n}\sigma^2E\left(\overline{X}-m\right)=0$$

$$Cov\left(\overline{X},S^2\right)=E\left[\left(\overline{X}-m\right)S^2\right]$$

$$=E\left[\left(\overline{X}-m\right)\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2-\left(\overline{X}-m\right)^2\right)\right]$$

$$= E\left[\left(\overline{X}-m\right)\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2\right] - E\left[\left(\overline{X}-m\right)\left(\overline{X}-m\right)^2\right]$$

$$= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{nm}{n}\right)\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^2\right] - E\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \frac{nm}{n}\right)^3\right]$$

$$= \frac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i-m)\sum_{i=1}^{n}(X_i-m)^2\right] - \frac{1}{n^3}E\left[\sum_{i=1}^{n}(X_i-m)\right]^3$$

En tenant compte du fait que

$E\left[(X_i-m)(X_j-m)\right]=0$ pour $i \neq j$ à cause de l'indépendance des variables aléatoires , on en déduit :

$$\frac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i-m)\sum_{i=1}^{n}(X_i-m)^2\right] = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^3\right) = \frac{1}{n}\mu_3$$

et

$$\frac{1}{n^3}E\left[\sum_{i=1}^{n}(X_i-m)\right]^3 = \frac{1}{n^2}E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-m)^3\right] = \frac{1}{n^2}\mu_3$$

Par conséquent

$$Cov\left(\overline{X},S^2\right) = \frac{n-1}{n^2}\mu_3$$

b) On a :

$$Cov\left(\overline{X},S^2\right) = \frac{n-1}{n^2}\mu_3$$

On cherche μ_3 pour une distribution de Poisson .

$$\mu_3 = \sum_{k=0}^{\infty}(k-E(X))^3 P(X=k)$$

avec

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{et} \quad E(X) = \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} =$$

$$e^{-\lambda}\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

Donc

$$\mu_3 = \sum_{k=0}^{\infty}(k-\lambda)^3 e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=0}^{\infty}\left(k^3-3k^2\lambda+3\lambda^2k-\lambda^3\right)e^{-\lambda}\frac{\lambda^k}{k!}$$

$$\mu_3 = e^{-\lambda}\left[\sum_{k=0}^{\infty}k^3\frac{\lambda^k}{k!} - 3\lambda\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!} + 3\lambda^2\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!} - \lambda^3\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\right]$$

On a :

$$\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{\lambda}$$

$$\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{\lambda}$$

$$\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!} = \sum_{k=0}^{\infty}\left[k(k-1)+k\right]\frac{\lambda^k}{k!}$$

$$= \sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)} + \sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}$$

$$= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda^2 e^{\lambda} + \lambda e^{\lambda}$$

$$\sum_{k=0}^{\infty} k^3 \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \left[k(k-1)(k-2)+3k(k-1)+k\right] \frac{\lambda^k}{k!}$$

$$= \sum_{k=3}^{\infty} \frac{\lambda^k}{(k-3)!} + \sum_{k=2}^{\infty} 3 \frac{\lambda^k}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}$$

$$= \lambda^3 \sum_{k=3}^{\infty} \frac{\lambda^{k-3}}{(k-3)!} + 3\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda^3 e^{\lambda} + 3\lambda^2 e^{\lambda} + \lambda e^{\lambda}$$

d'où

$$\mu_3 = e^{-\lambda} e^{\lambda} \left[\lambda^3 + 3\lambda^2 + \lambda - 3\lambda\left(\lambda^2 + \lambda\right) + 3\lambda^2 . \lambda - \lambda^3 \right]$$

$$= \lambda$$

et

$$Cov\left(\overline{X}, S^2\right) = \frac{n-1}{n^2} \lambda$$

4.3. La moyenne de la population m vaut :

$$m = \frac{1+2+3+\ldots+8+9}{9} = 5$$

La variance de cette population est :

$$\sigma^2 = \frac{1}{9}\left[(1-5)^2 + (2-5)^2 + \ldots + (9-5)^2\right]$$

$$= \frac{1}{9}(16+9+4+1+0+1+4+9+16)$$
$$= 6,67$$

et l'écart-type est : $\sigma = 2,58$

Il y a $C_9^6 = \frac{9!}{6!3!} = 84$ façons de choisir six nombres parmi les nombres de 1 à 9 . Chacun de ces 84 échantillons a une moyenne $\overline{X} = \frac{1}{6}\sum_{i=1}^{6} x_i$ où les x_i représentent six des nombres de 1 à 9 .
Par exemple , si on choisit l'échantillon (3,8,7,2,5,1), la moyenne qui lui correspond est

$$\overline{X} = \frac{3+8+7+2+5+1}{6} = 4,33$$

On obtient ainsi 84 moyennes et la moyenne de la distribution d'échantillonnage des moyennes notée $\mu_{\overline{X}}$ vaut :

$$\mu_{\overline{X}} = m = 5$$

La variance de la distribution d'échantillonnage est :

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$; N étant la taille de la population et **n** la taille de l'échantillon .

$$\sigma_{\overline{X}}^2 = \frac{6,67}{6}\left(\frac{9-6}{9-1}\right) = 0,417$$

d'où $\sigma_{\overline{X}} = 0,645$

4.4. Il y a 9^6 façons de choisir six nombres parmi les nombres de 1 à 9 (le tirage étant avec remise) ; chacun de ces $9^6 = 531441$ échantillons a une moyenne $\overline{X} = \frac{1}{6}\sum_{i=1}^{6} x_i$ où les $x_i (i = 1, \dots, 6)$ représentent six des nombres de 1 à 9 .
Par exemple , si l'échantillon (4,3,4,5,7,8) est choisi , la moyenne est :

$$\overline{X} = \frac{4+3+4+5+7+8}{6} = 5,17$$

Nous obtenons ainsi 9^6 moyennes et la moyenne de la distribution d'échantillonnage des moyennes est :

$$\mu_{\overline{X}} = m = 5,17$$

La variance de la distribution d'échantillonnage est :

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} = \frac{6,67}{6} = 1,11$$

car

$$\sigma^2 = \frac{1}{9}\left[(1-5)^2 + (2-5)^2 + \dots + (9-5)^2\right] = 6,67$$

L'écart-type $\sigma_{\overline{X}}$ vaut : $\sigma_{\overline{X}} = 1,05$

4.5. 1) L'échantillon est exhaustif et la taille n=40 est assez grande (n>30) , la distribution d'échantillonnage de la moyenne suit une distribution normale de moyenne $\mu_{\overline{X}} = m$ et d'écart-type $\sigma_{\overline{X}} = \frac{\sigma}{n}\sqrt{\frac{N-n}{N-1}}$

m et N sont la moyenne et la taille de la population .

a) On a :

$$P\left(10 < \overline{X} < 13\right) = P\left(\frac{10-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{13-\mu_{\overline{X}}}{\sigma_{\overline{X}}}\right)$$

$$= P\left(u_1 < Z < u_2\right)$$

avec $Z = \frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}}$ suit une loi normale N(0,1)

$$u_1 = \frac{10-9,8}{\frac{3,68}{\sqrt{40}}\sqrt{\frac{300-40}{300-1}}} = 0,37$$

$$u_2 = \frac{13-9,8}{\frac{3,68}{\sqrt{40}}\sqrt{\frac{300-40}{300-1}}} = 5,89$$

Donc

$$P\left(10 < \overline{X} < 13\right) = P(0,37 < Z < 5,89) = \Phi(5,89) - \Phi(0,37)$$

$$= 0,9999 - 0,6443$$

$$= 0,3556$$

Φ est la fonction de répartition de la variable aléatoire Z qui suit une loi normale N(0,1)

b) On a :

$$P\left(\overline{X} < 10\right) = P\left(\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} < \frac{10-9,8}{0,543}\right)$$

$\Phi(0,37)=0,6443$

2) L'échantillon étant non exhaustif on a :

$$\mu_{\overline{X}} = m \text{ et } \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} \quad , \quad \text{d'où :}$$

a)

$$P\left(10 < \overline{X} < 13\right) = P\left(\frac{10-9,8}{\frac{3,68}{\sqrt{40}}} < \frac{\overline{X}-\mu_{\overline{X}}}{\frac{\sigma}{\sqrt{n}}} < \frac{13-9,8}{\frac{3,68}{\sqrt{40}}}\right)$$

$$= P(0,34 < Z < 5,50)$$

$$= \Phi(5,50) - \Phi(0,34)$$

$$= 0,9999 - 0,6331 = 0,3668$$

b)

$$P\left(\overline{X} < 10\right) = P\left(\frac{\overline{X}-\mu_{\overline{X}}}{\sigma_{\overline{X}}} \le \frac{10-\mu_{\overline{X}}}{\sigma_{\overline{X}}}\right)$$

$$= P\left(Z \le \frac{10-9,8}{\frac{3,68}{\sqrt{40}}}\right) = \Phi(0,34)$$

$$= 0,6331$$

4.6. Pour une variable aléatoire X suivant une loi de Poisson on a :

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad ; \quad x = 0,1,2,\ldots$$

où λ est le paramètre inconnu.

On a :

$$L(x_1, x_2, \ldots x_n, \lambda) = \frac{\lambda^{x_1}}{x_1\,!} e^{-\lambda} \frac{\lambda^{x_2}}{x_2\,!} e^{-\lambda} \ldots \frac{\lambda^{x_n}}{x_n\,!} e^{-\lambda}$$

$$= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i\,!} e^{-\lambda}$$

et

$$\ln L = \sum_{i=1}^{n} \left(x_i \ln \lambda - \ln(x_i\,!) - \lambda\right)$$

$$= \ln \lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i\,!) - n\lambda$$

On résoud l'équation :

$$\frac{\partial \ln L}{\partial \lambda} = 0 \Leftrightarrow \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0 \qquad \text{d'où}$$

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

4.7. La densité de probabilité de la loi est :

$$f(x, m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

On a :

$$L(x_1, x_2, x_3, \ldots, x_n, m, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - m)^2}$$

et

$$\ln L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - m)^2$$

On dérive $\ln L$ par rapport à m et à σ^2 et on annule les dérivées partielles . On obtient le système d'équations :

$$\begin{cases}\dfrac{\partial \ln L}{\partial m} = 0 \\ \dfrac{\partial \ln L}{\partial \sigma^2} = 0\end{cases} \quad \text{ou} \quad \begin{cases}\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i - m)^2 = 0 \\ -\dfrac{n}{2}\dfrac{1}{\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(x_i - m)^2 = 0\end{cases}$$

La résolution de ce système donne les estimations :

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

et

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

4.8. On a :

$$L_i(x_i, p) = \begin{cases}1-p & \text{si } x_i = 0 \\ p & \text{si } x_i = 1\end{cases}$$

La fonction

$$L(x,p) \text{ est } L(x,p) = L_1(x_1,p)L_2(x_2,p)\ldots L_n(x_n,p)$$

$$= p^k(1-p)^{n-k}$$

où $k = \sum_{i=1}^{n} x_i$

$$\frac{\partial L}{\partial p} = kp^{k-1}(1-p)^{n-k} - p^k(n-k)(1-p)^{n-k-1} = 0$$

$$\Leftrightarrow \quad p^k(1-p)^{n-k}\left(\frac{k}{p} - \frac{n-k}{1-p}\right) = 0$$

On en déduit

$$\left(\frac{k}{p} - \frac{n-k}{1-p}\right) = 0 \qquad \text{c'est à dire}$$

$$k(1-p) = p(n-k) \qquad \text{ou} \qquad p = \frac{k}{n}$$

4.9. On sait qu'à chaque valeur de n correspond une distribution t particulière . Quand n augmente , la distribution t tend vers la distribution normale réduite . Elles sont approximativement égales dès que n>30

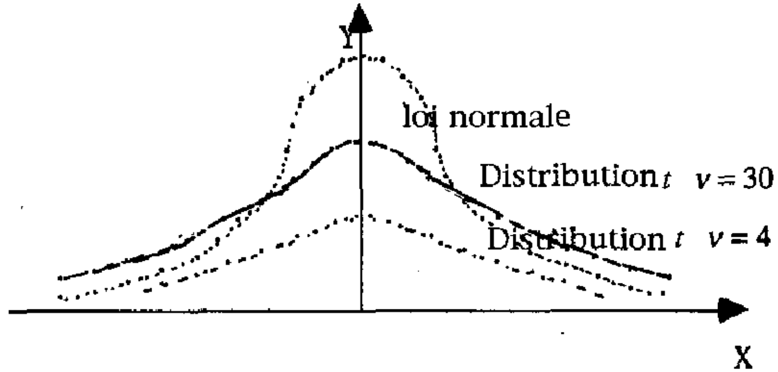


Fig. 4.1

a) Au seuil $\alpha = 0,95$ on cherche u_α telle que $P(|Z| < u_\alpha) = 0,95$ où Z suit une loi normale N(0,1).
La distribution de Z est symétrique par rapport à l'origine ;

$$P(|Z| < u_\alpha) = P(-u_\alpha < Z < u_\alpha)$$
$$= \Phi(u_\alpha) - \Phi(-u_\alpha)$$
$$= \Phi(u_\alpha) - [1 - \Phi(u_\alpha)]$$

Φ est la fonction de répartition de Z .

$$P(|Z| < u_\alpha) = 2\Phi(u_\alpha) - 1 = 0,95$$

On obtient

$$\Phi(u_\alpha) = \frac{1,95}{2} = 0,975$$

La table de la loi normale N(0,1) donne $u_\alpha = 1,96$

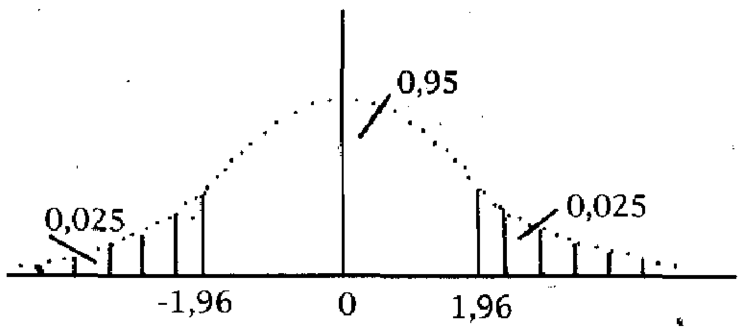


Fig. 4.2

L'aire hachurée vaut 0,025.2=0,5 ;celle non hachurée vaut 0,95

b) Au seuil de confiance $\alpha = 0,95$, on cherche u_α telle que $P(|t| < u_\alpha) = 0,95$ où t suit une distribution de Student avec $v = 4$(ou $v = 30$) degrés de liberté .

i) si $v = 4$, la distribution t étant symétrique par rapport à l'origine on a :

$$P(|t| < u_\alpha) = P(-u_\alpha < t < u_\alpha) = G(u_\alpha) - G(-u_\alpha)$$
$$= 2G(u_\alpha) - 1 = 0,95$$

Donc

$$G(u_\alpha) = \frac{1,95}{2} = 0,975$$

G est la fonction de répartition de t .
La lecture de la table pour la distribution t donne $u_\alpha = 2,776$

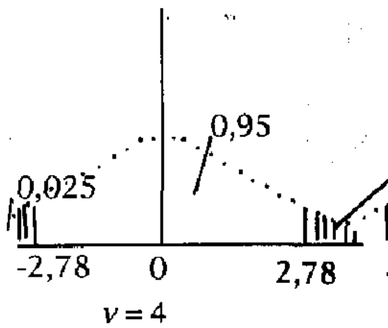


Fig. 4.3

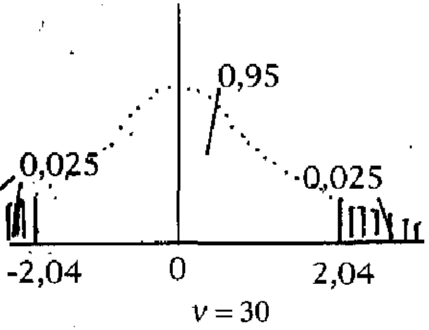


Fig. 4.4

ii) Si $v = 30$, le même raisonnement que précédemment donne $G(u_\alpha) = 0,975$ et on lit dans la table(distribution t de Student) $u_\alpha = 2,042$

4.10. La variance de l'échantillon aléatoire de taille n est :

$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$. S^2 n'est pas un estimateur sans biais de σ^2 ; $\sigma^2 = V(X) = E\left[(X-m)^2\right]$. On calcule $E\left(S^2\right)$

On a :

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n}\left[\left(X_i - m\right) - \left(\overline{X} - m\right)\right]^2$$

$$= \sum_{i=1}^{n}\left(X_i - m\right)^2 - 2\left(\overline{X} - m\right)\sum_{i=1}^{n}\left(X_i - m\right) + \sum_{i=1}^{n}\left(\overline{X} - m\right)^2$$

$$= \sum_{i=1}^{n}\left(X_i - m\right)^2 - 2\left(\overline{X} - m\right)n\left(\overline{X} - m\right) + n\left(\overline{X} - m\right)^2$$

$$= \sum_{i=1}^{n}\left(X_i - m\right)^2 - n\left(\overline{X} - m\right)^2$$

d'où

$$nE\left(S^2\right) = E\left[\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] = E\left[\sum_{i=1}^{n}\left(X_i - m\right)^2\right] - nE\left(\overline{X} - m\right)^2$$

$$= \sum_{i=1}^{n} E\left(X_i - m\right)^2 - nE\left(\overline{X} - m\right)^2$$

$$= n\sigma^2 - nV\left(\overline{X}\right) = n\sigma^2 - n\frac{\sigma^2}{n}$$

et

$$E\left(S^2\right) = \sigma^2 - \frac{\sigma^2}{n}$$

$$= \frac{n-1}{n}\sigma^2$$

Quand $n \to \infty$, on a : $E\left(S^2\right) = \frac{n-1}{n}\sigma^2 \to \sigma^2$

Pour obtenir un estimateur sans biais de σ^2 , il faut donc :

$$S'^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

4.11. a) La **fréquence** d'échantillonnage centrée et réduite est: $\frac{f-p}{\sigma_f} = \frac{f-p}{\sqrt{\frac{p(1-p)}{n}}}$. Les bornes de l'intervalle de confiance sont : $f = p \pm u_\alpha\sqrt{\frac{p(1-p)}{n}}$

Soit

$$(f-p)^2 = \frac{u_\alpha^2 p(1-p)}{n} \quad \text{ou} \quad f^2 + p^2\left(1 + \frac{u_\alpha^2}{n}\right) - 2fp - \frac{u_\alpha^2 p}{n} = 0$$

On résoud l'**équation en p** , étant donné une valeur de f observée :

$$p^2\left(1 + \frac{u_\alpha^2}{n}\right) - p\left(\frac{u_\alpha^2}{n} + 2f\right) + f^2 = 0$$

Le discriminant Δ vaut :

$$\Delta = \left(\frac{u_\alpha^2}{n} + 2f\right)^2 - 4\left(1 + \frac{u_\alpha^2}{n}\right)f^2 = \frac{u_\alpha^4}{n^2} + 4f\frac{u_\alpha^2}{n} - 4f^2\frac{u_\alpha^2}{n}$$

d'où

$$p=\frac{\left(2f+\frac{u_\alpha^2}{n}\right)\pm\sqrt{\frac{u_\alpha^2}{n^2}+4f\frac{u_\alpha^2}{n}-4f^2\frac{u_\alpha^2}{n}}}{2\left(1+\frac{u_\alpha^2}{n}\right)}$$

b) Si n est grand le terme $\dfrac{2f+\frac{u_\alpha^2}{n}}{2\left(1+\frac{u_\alpha^2}{n}\right)}\sim f$

On transforme le second terme :

$$\frac{\sqrt{\frac{u_\alpha^4}{n^2}+4f\frac{u_\alpha^2}{n}-4f^2\frac{u_\alpha^2}{n}}}{2\left(1+\frac{u_\alpha^2}{n}\right)}=\sqrt{\frac{u_\alpha^4+4fnu_\alpha^2-4f^2nu_\alpha^2}{4\left(n+u_\alpha^2\right)^2}}$$

$$=\sqrt{\frac{u_\alpha^4+4fnu_\alpha^2-4f^2nu_\alpha^2}{4n^2+8u_\alpha^2n+4u_\alpha^4}}$$

On a :

$$u_\alpha^4+4fnu_\alpha^2-4f^2nu_\alpha^2\sim 4fnu_\alpha^2-4f^2nu_\alpha^2$$

$$4n^2+8u_\alpha^2n+4u_\alpha^4\sim 4n^2$$

Donc le terme vaut :

$$\sqrt{\frac{4fnu_\alpha^2-4f^2nu_\alpha^2}{4n^2}}=u_\alpha\sqrt{\frac{f(1-f)}{n}}$$

Pour n grand , on obtient :

$$p=f\pm u_\alpha\sqrt{\frac{f(1-f)}{n}}$$

4.12. On a une distribution théorique d'échantillonnage normale car n>30 et $\sigma_{\overline{X}}=\frac{\sigma}{\sqrt{n}}=\frac{24}{\sqrt{40}}=3,79$

Donc $m\in\left]\overline{X}-1,96\sigma_{\overline{X}},\overline{X}+1,96\sigma_{\overline{X}}\right[$

$$162,56 < m < 177,44$$

m est compris entre 162,56 et 177,44 avec un seuil de 95% . On a supposé que la population est suffisamment grande pour que le tirage puisse être assimilé à un tirage non exhaustif .

Si on considère le tirage exhaustif (sans remise) , on a :

$$\sigma_{\overline{X}}=\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\overline{X}}=\frac{24}{\sqrt{40}}\sqrt{\frac{780-40}{780-1}}=3,70$$

et $m\in]162,75\ ;\ 177,25[$

4.13. On suppose que la taille de la population (N=500) est suffisamment grande par rapport à celle de l'échantillon (n=38) pour être dans les conditions d'un tirage non exhaustif.
Puisque n>30 , on utilise la distribution normale pour construire l'intervalle de confiance pour la moyenne des notes de la population et on considère que S=2,82 est une estimation de la valeur inconnue σ .

On aura alors

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{S}{\sqrt{n}} = \frac{2,82}{\sqrt{38}} = 0,46$$

a) A un seuil de confiance de 90%

$$m \in \left]\overline{X} - 1,64\sigma_{\overline{X}} \ , \ \overline{X} + 1,64\sigma_{\overline{X}}\right[$$

$$7,90 < m < 9,40$$

b) A un seuil de confiance de 95%

$$m \in \left]\overline{X} - 1,96\sigma_{\overline{X}} \ , \ \overline{X} + 1,96\sigma_{\overline{X}}\right[$$

$$7,75 < m < 9,55$$

c) A un seuil de confiance de 99%

$$m \in \left]\overline{X} - 2,58\sigma_{\overline{X}} \ , \ \overline{X} + 2,58\sigma_{\overline{X}}\right[$$

$$7,46 < m < 9,84$$

4.14. a) On suppose que la taille N de la population est suffisamment grande par rapport à celle de l'échantillon (n=58) pour être dans les conditions d'un tirage non exhaustif on a $\sigma_f = \sqrt{\dfrac{p(1-p)}{n}}$, Comme $n > 30$, $\dfrac{f-p}{\sigma_f}$ suit une loi normale N(0,1) et

$$p \in \left]f - 1,96\sigma_f \ , \ f + 1,96\sigma_f\right[$$

où on estime p (proportion de la population favorable au candidat) dans σ_f par f :

$$\sigma_f = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{f(1-f)}{n}} = \sqrt{\frac{0,52.0,48}{58}} = 0,066$$

et $\quad p \in]0,39 \ ; \ 0,65[$

b) Les bornes de l'intervalle de confiance sont :

$$p = f \pm 1,96\sqrt{\frac{p(1-p)}{n}}$$

Il vient $\quad p - f = \pm 1,96\sqrt{\dfrac{p(1-p)}{n}}$

$$(p-f)^2 = (1,96)^2\,\frac{p(1-p)}{n}$$

et $\quad n = \dfrac{(1,96)^2\,p(1-p)}{(p-f)^2} = \dfrac{(1,96)^2.0,5.0,5}{(0,5-0,52)^2} = 2401$

La taille de l'échantillon doit être au moins égale à n=2401.

4.15. a) Les bornes de l'intervalle de confiance sont :

$$m = \overline{X} \pm u_\alpha \frac{\sigma}{\sqrt{n}} \quad ; \quad m = 9,55 \pm u_\alpha \frac{3,65}{\sqrt{9}}$$

On considère que S=3,65 est une estimation de la valeur inconnue σ.

Pour $\alpha = 0,95$ et à $\nu = n - 1 = 9 - 1 = 8$degrés de liberté , on a : $u_\alpha = 2,306$

On obtient alors l'intervalle cherché :

$$m \in \left]9,55 - 2,306\frac{3,65}{3} \ ; \ 9,55 + 2,306\frac{3,65}{3}\right[$$

ou $\quad 6,74 < m < 12,36$

b) pour $\alpha = 0,99$ et à $\nu = 8$degrés de liberté on a :$u_\alpha = 3,355$ et on obtient l'intervalle :

$$m \in \left]9,55 - 3,355\frac{3,65}{3} \ ; \ 9,55 + 3,355\frac{3,65}{3}\right[$$

ou $\quad 5,47 < m < 13,63$

Si on applique la méthode des grands échantillons , la distribution sera normale et les bornes deviennent :

a) Pour $\alpha = 0,95$

$$m \in \left]9,55 - 1,96\frac{3,65}{3} \ ; \ 9,55 + 1,96\frac{3,65}{3}\right[$$

ou $7{,}17 < m < 11{,}93$

b) Pour $\alpha = 0,99$

$$m \in \left]9,55-2,58\frac{3,65}{3}\ ;\ 9,55+2,58\frac{3,65}{3}\right[$$

ou $6{,}41 < m < 12{,}69$

On remarque que la précision est moindre avec la méthode des petits échantillons qu'avec celles des plus grands.

4.16. On a :

$$P\left(\left|\frac{(\overline{X}_1-\overline{X}_2)-\mu_{\overline{X}_1-\overline{X}_2}}{\sigma_{\overline{X}_1-\overline{X}_2}}\right| < u_\alpha\right) = 0,95$$

avec

$$u_\alpha = 1,96\ ;\ \mu_{\overline{X}_1-\overline{X}_2} = m_1 - m_2$$

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\sigma^2_{\overline{X}_1}+\sigma^2_{\overline{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

où m_1 et σ_1 sont les moyennes et l'écart-type de la population 1 (composants produits par l'usine A) et m_2 et σ_2 sont les moyennes et l'écart-type de la population 2 (composants produits par l'usine B).
$S_1 = 200$ et $S_2 = 230$ sont des estimations des valeurs inconnues σ_1 et σ_2.

Les limites de confiance sont :

$$(\overline{X}_1-\overline{X}_2)-u_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\ ;\text{et}(\overline{X}_1-\overline{X}_2)+u_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$$= 355,72$$

$$(\overline{X}_1-\overline{X}_2)+u_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = (3400-3000)+1,96\sqrt{\frac{200^2}{180}+\frac{230^2}{130}}$$

$$= 444,28$$

et $355,72 < \mu_{\overline{X}_1-\overline{X}_2} < 444,28$

La différence des moyennes varie entre 355,72 et **444,28 avec** une probabilité égale à 0,95 .

4.17. L'échantillon aléatoire de taille n suit une loi normale $N\left(m,\frac{\sigma}{\sqrt{n}}\right)$ et donc la variable $\frac{\overline{X}-m}{\sigma/\sqrt{n}}$ suit une loi normale N(0,1).
Une variable t de Student à k degrés de liberté s'écrit : $t = \frac{X\sqrt{k}}{\sqrt{Y}}$ où X suit une loi normale N(0,1) et Y suit une loi du Khi-carré (χ_k^2) à k degrés de liberté .
La variable aléatoire

$$\frac{(n-1)S'^2}{\sigma^2}\left(\text{ où } S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right) \text{ sui}$$

loi du χ^2_{n-1} à $n-1$ degrés de liberté .
On conclut :

$$t = \frac{\frac{\overline{X}-m}{\sigma/\sqrt{n}}\sqrt{n-1}}{\sqrt{\frac{(n-1)S'^2}{\sigma^2}}} = \frac{(\overline{X}-m)\sqrt{n}}{S'}$$

$$t = \frac{(\overline{X}-m)\sqrt{n}}{S'} = \frac{(\overline{X}-m)\sqrt{n-1}}{S} \quad \text{car } S'^2 = \frac{n}{n-1}S^2$$

suit une loi de Student à n-1 degrés de liberté .

CHAPITRE 5

TESTS STATISTIQUES

I- Introduction :

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	H_1 fausse
ou bien	H_0 fausse	H_1 vraie

Il y a 4 solutions dont seulement les deux premières sont justes :

a) H_0 est vraie et on a choisi H_0
b) H_0 est fausse et on a rejeté H_0
c) H_0 est vraie et on a rejeté H_0
d) H_1 est vraie et on a choisi H_0

On distingue deux types d'erreurs :

a) Si H_0 est vraie et on l'a rejetée , on dit que l'on a une erreur de 1ère espèce . La probabilité de l'erreur de1ère espèce est notée α .

b) si H_1 est vraie et on a accepté H_0 , on dit que l'on a une erreur de 2e espèce . La probabilité de l'erreur de 2e espèce est notée β .
α est le seuil de signification du test et 1-α son seuil de confiance .

II- Catégories de tests :

1- Un test est dit d'hypothèse simple si on veut choisir entre deux valeurs d'un paramètre θ $(\theta_0$ et $\theta_1)$;
On a :

$$\begin{cases} H_0 & : \quad \theta = \theta_0 \\ H_1 & : \quad \theta = \theta_1 \end{cases}$$

2- Un test est dit d'hypothèse multiples si :

a)

$$\begin{cases} H_0 & : \quad \theta = \theta_0 \\ H_1 & : \quad \theta > \theta_1 \end{cases}$$

test unilatéral à droite ou test de supériorité

b)

$$\begin{cases} H_0 & : \quad \theta = \theta_0 \\ H_1 & : \quad \theta < \theta_1 \end{cases}$$

test unilatéral à gauche ou test d'infériorité

c)

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

test bilatéral

3- Un test est dit d'homogénéité si :

$$\begin{cases} H_0 & : \quad \theta = \theta_1 \\ H_1 & : \quad \theta \neq \theta_1 \end{cases}$$

où θ_0 et θ_1 sont deux valeurs d'un même paramètre dans deux populations différentes .

4- Un test est dit d'ajustement si

$$\begin{cases} H_0 : & F(x) = F_0(x) \\ H_1 : & F(x) \neq F_0(x) \end{cases}$$

où $F(x)$ est la fonction de répartition de la variable échantillonnée et $F_0(x)$ est la fonction de répartition d'une variable aléatoire connue .

5- Un test est dit d'indépendance si :

$$\begin{cases} H_0 : & X \text{ et } Y \text{ sont deux variables} \\ & \text{aléatoires indépendantes} \\ H_1 : & X \text{ et } Y \text{ ne sont pas des} \\ & \text{variables indépendantes} \end{cases}$$

III- Régions critique et d'acceptation d'une hypothèse.

La construction d'un test implique la détermination de la région critique W_0 de R^n . La valeur de α , erreur de 1ère espèce étant fixée (en général α=0,05 ; 0,01 ou 0,1) l'ensemble des valeurs de la variable de décision qui permettent d'écarter H_0 et de choisir H_1 est dit région critique ; le complémentaire $\overline{W_0}$ de cette région critique est dit région d'acceptation .

On a :

$$P(W_0 / H_0) = \alpha \quad ; \quad P(\overline{W_0} / H_0) = 1-\alpha$$

$$P(W_0 / H_1) = 1-\beta \quad ; \quad P(\overline{W_0} / H_1) = \beta$$

On extrait un échantillon aléatoire de la population et on accepte H_0 si la valeur de la variable de décision appartient à la région d'acceptation . Sinon on la rejette et on accepte H_1. Pour une valeur α fixée , on maximise la quantité $1-\beta$ dite puissance du test .

IV- Test entre deux hypothèses simples (Méthode de Neyman et Pearson)

On teste

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

On fixe le risque de 1ère espèce α .

$L(x,\theta) = L(x_1, x_2, \ldots, x_n, \theta)$ est la fonction de vraisemblance avec $x = (x_1, x_2, \ldots, x_n)$

W_0 , région critique , est définie par :

$$\alpha = P(W_0 / H_0) = \int_{W_0} L(x, \theta_0) dx$$

On maximise la quantité

$$1-\beta = \int_{W_0} L(x, \theta_1) dx = p(W_0 / H_1)$$

$$1-\beta = \int_{W_0} \frac{L(x,\theta_1)}{L(x,\theta_0)} L(x, \theta_0) dx$$

Pour maximiser $1-\beta$, on cherche l'ensemble des points de R^n tels que :

$$A = \frac{L(x,\theta_1)}{L(x,\theta_0)} \geq K_\alpha$$

La constante K_α est déterminée par :

$$\int_{A \geq K_\alpha} L(x,\theta_0)\,dx = \alpha$$

V- Test d'hypothèse multiples

On teste

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \quad \text{ou} \quad \theta > \theta_0 \quad \text{ou} \quad \theta \neq \theta_0 \end{cases}$$

On fixe **le risque** α et on **maximise** la puissance du test $1-\beta$. Le risque 1ère **espèce** α **étant fixé**, on définit pour chaque valeur $\theta_i(\theta_i \neq \theta_0)$ **une région critique** W_{0i}. Si ces régions sont identiques, **le test est dit UPP** (uniformément le plus puissant). Il n'existe pas de test UPP pour :

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

car sinon, il est UPP pour les deux tests

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \quad \text{et} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

Ces deux tests sont UPP et différents l'un de l'autre

VI-Test 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 entre :

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

θ_0 *et* θ_1 sont les deux valeurs d'un même paramètre des deux populations p_1 *et* p_2.

1- Test d'homogénéité de deux moyennes

Dans le cas où les tailles des échantillons sont élevées $(n_1, n_2 \geq 30)$, **les variables** $\overline{X}_1$ et $\overline{X}_2$ (correspondants aux populations p_1 et p_2) suivent les lois normales respectives :

$$N\left(m_1, \frac{\sigma_1}{\sqrt{n_1}}\right) \quad \text{et} \quad N\left(m_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$$

où m_i et σ_i sont la moyenne et l'écart-type de la population $p_i(i=1,2)$

La variable aléatoire $(\overline{X}_1 - \overline{X}_2)$ suit aussi une loi normale

$$N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

On choisit entre les deux hypothèses .

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

ou

$$\begin{cases} H_0 : m_1 - m_2 = 0 \\ H_1 : m_1 - m_2 \neq 0 \end{cases}$$

Si H_0 est vraie , $m_1 - m_2 = 0$ et $\overline{X}_1 - \overline{X}_2$ suit une loi normale

$$N\left(0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

On a alors :

$$p\left(|Z| < u_\alpha\right) = 1 - \alpha$$

où

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(m_1 - m_2\right)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

suit une loi normale N(0,1).

On accepte H_0 si la valeur $z = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ de Z telle que :

$$-u_\alpha < z < u_\alpha$$

où la valeur u_α est obtenue par la lecture de la table de la loi normale N(0,1)

On rejette H_0 si $|Z| \geq u_\alpha$. On dira que la différence est significative entre $\overline{X}_1$ et $\overline{X}_2$.

Remarque

i) Si σ_1^2 et σ_2^2 sont inconnues , on les remplace par les estimateurs :

$$S'^2_1 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(X_i - \overline{X}_1\right)^2$$

et

$$S'^2_1 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_2} \left(X_j - \overline{X}_2\right)^2 \text{ respectivement}$$

Comme les échantillons sont de tailles élevées on considère que $Z = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ suit une loi normale N(0,1)

(Si H_0 est vraie c'est à dire $m_1 - m_2 = 0$)

ii) Si les échantillons sont de tailles respectives $n_1 < 30$ et $n_2 < 30$, le test n'est plus valable car le théorème central limite ne s'applique plus . Mais pour deux populations p_1 et p_2 suivant des lois normales $N(m_1, \sigma_1)$ et $N(m_2, \sigma_2)$ respectivement ayant des écarts-type σ_1 et σ_2 égaux et inconnus , c'est à dire $\sigma_1 = \sigma_2 = \sigma$ on a :

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

suit une loi de Student à $n_1 + n_2 - 2$ degrés de liberté ; S^2 **étant** l'estimation ponctuelle de σ^2 .

On accepte alors H_0 si $|t| < u_\alpha$ où u_α est une valeur obtenue par lecture de la table de la loi t de Student-Fisher (nombre de degrés de liberté : $n_1 + n_2 - 2$; seuil de signification : α)

2- Test d'homogénéité de deux proportions

Chaque individu des deux populations P_1 et P_2 peut posséder ou non un certain caractère . On dit ce caractère est présent en proportions p_1 et p_2 dans les populations P_1 et P_2 respectivement :

On test au seuil de signification α

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

ou

$$\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0 \end{cases}$$

De la population P_i , on extrait un échantillon de taille n_i . Il lui correspond une proportion $f_i (i = 1, 2)$

Si les échantillons sont de tailles élevées $(n_1 \geq 30$ et $n_2 \geq 30)$, le théorème central limite permet d'affirmer que f_i suit une loi normale $N\left(p_i, \sqrt{\frac{p_i q_i}{n_i}}\right)$ avec $p_i + q_i = 1$. La variable aléatoire $f_1 - f_2$ suit alors une loi normale $N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$.

Si H_0 est vraie , $p_1 - p_2 = 0$ et $f_1 - f_2$ suit une loi normale $N\left(0, \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$ car $p_1 = p_2 = p$

Mais puisque p est inconnu , on fait une approximation . On estime p par $f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ et donc $f_1 - f_2$ suit une loi normale $N\left(0, \sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$ et

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

suit une loi normale N(0,1)

Soit u_α la constante telle que $P(|Z| \geq u_\alpha) = \alpha$. On a alors :

$$P(|Z| < u_\alpha) = 1 - \alpha$$

On accepte H_0 si la valeur $z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ de Z est telle que :

$$-u_\alpha < z < u_\alpha$$

On rejette H_0 si $|Z| \geq u_\alpha$

u_α est obtenue par lecture de la table de la loi normale N(0,1)

3- Test d'homogénéité de deux variance

Deux populations P_1 et P_2 supposées suivre des distributions normales ont pour variances inconnues σ_1 et σ_2 respectivement.

On teste au seuil de signification α

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

On extrait de chaque population un échantillon de taille $n_1(\text{pour } P_1)$ ou $n_2(\text{pour } P_2)$

Soient

$$S'^2_1 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(X_i - \overline{X}_1\right)^2 \quad \text{l'estimation de } \sigma_1^2$$

et

$$S'^2_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}\left(X_j - \overline{X}_2\right)^2 \quad \text{l'estimation de } \sigma_2^2$$

$X_i(i = 1, 2, \ldots, n_1)$ et $X_j(j = 1, 2, \ldots, n_2)$ sont les valeurs prises par les individus des échantillons des deux populations et $\overline{X}_1$ et $\overline{X}_2$ les moyennes respectives.

Soit

$$F = \frac{S'^2_1}{S'^2_2}.$$

Le test devient :

$$\begin{cases} H_0 : F = 1 \\ H_1 : F \neq 1 \end{cases}$$

Si H_0 est vraie $\left(\sigma_1^2 = \sigma_2^2\right)$, la variable

$$F_{n_1-1,n_2-1} = \frac{S'^2_1 / \sigma_1^2}{S'^2_2 / \sigma_2^2} = \frac{S'^2_1}{S'^2_2}$$

suit une loi de Snédecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté.

On a donc :

$$P\left(\frac{S'^2_1}{S'^2_2} < F^{\alpha}_{n_1-1,n_2-1}\right) = 1 - \alpha$$

α étant fixé, la valeur $F^{\alpha}_{n_1-1,n_2-1}$ est obtenue par lecture de la table de Fischer -Schedecor au seuil α et à $n_1 - 1$ et $n_2 - 1$ degrés de liberté.

Si la valeur $f = \frac{S'^2_1}{S'^2_2}$ de F calculée est telle que $f < F^{\alpha}_{n_1-1,n_2-1}$ on accepte H_0. On admet alors l'égalité des variances inconnues.

Si $f > F^{\alpha}_{n_1-1,n_2-1}$ on rejette H_0

En pratique on met toujours au numérateur la plus grande des deux quantités S'^2_1 et S'^2_2.

VII.- Test d'ajustement

1- Test du χ^2

Soit un échantillon aléatoire de taille n extrait d'une population et divisé en k classes d'effectifs respectifs $n_1, n_2, \ldots, n_k$ et de probabilités respectives $p_1, p_2, \ldots, p_k$ théoriques.

Il s'agit de tester

$$\begin{cases} H_0 \ : \ F(x) = F_0(x) \\ H_1 \ : \ F(x) \neq F_0(x) \end{cases}$$

Où $F(x)$ est la fonction de répartition de la variable échantillonnée et $F_0(x)$ est la fonction de répartition d'une variable aléatoire connue .

La quantité $\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ suit une loi du χ^2_{k-1} à $k-1$ degrés de liberté si np_i est supérieur à **5** .

Si la variable aléatoire **est continue** , on associe une probabilité p_j à la $j^{ème}$ classe en utilisant la densité de probabilité .

Le seuil de signification α étant fixé , on lit sur la table du χ^2 la valeur $\chi^2_{k-1}(\alpha)$ en fonction du nombre de degrés de liberté $k-1$ qui est le nombre de variables indépendantes . En effet comme $\sum_{i=1}^{k} n_i = n$, il suffit de connaître en fait $k-1$ variables

On a :

$$P\left(\chi^2_{k-1} < \chi^2_{k-1}(\alpha)\right) = 1 - \alpha$$

Si la valeur calculée $\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ est telle que :

$$\chi^2_{k-1} < \chi^2_{k-1}(\alpha)$$

on accepte l'hypothèse d'ajustement c'est à dire H_0 ; **sinon on** la rejette , la loi théorique proposée ne peut **représenter** valablement la distribution expérimentale .

On appelle la quantité np_i effectif théorique.

2- Test de Kolmogorov

Soit $F(x)$ la fonction de répartition pour la **population** et soit $F_n(x)$ la fonction de répartition **empirique** (fréquence cumulée relative) pour un échantillon de **taille n** .

On détermine $D_n = Max\left|F_n(x) - F(x)\right|$ et on compare cet écart à des valeurs critiques particulières données par **les** tables d(n).

On rejette H_0 si $D_n \geq d(n)$ à un seuil de signification α

Pour $\alpha = 0,05$ et $n > 80$ on a: $D_n > \frac{1,36}{\sqrt{n}}$

Pour $\alpha = 0,01$ et $n > 80$: $D_n > \frac{1,63}{\sqrt{n}}$

$D_n > \frac{1,36}{\sqrt{n}}$ et $D_n > \frac{1,36}{\sqrt{n}}$ définissent chacun une **région** critique pour le seuil α correspondant

VIII- Test d'indépendance

Le test vise à rechercher s'il existe une liaison **entre** deux variables aléatoires X et Y . On teste :

H_0 : X et Y sont indépendantes

H_1 : X et Y ne sont pas **indépendantes**

Soit le tableau de contingence représentant empiriquement la relation des deux variables aléatoires

X \ Y	y_1	y_2	....	y_j	....	y_q	Total
x_1	n_{11}	n_{12}	....	n_{1j}	....	n_{1q}	$n_{1.}$
x_2	n_{21}	n_{22}	....	n_{2j}	....	n_{2q}	$n_{2.}$
.	.	.		.		.	.
x_i	n_{i1}	n_{i2}	....	n_{ij}	....	n_{iq}	$n_{i.}$
.	.	.		.		.	.
x_p	n_{p1}	n_{p2}	....	n_{pj}	....	n_{pq}	$n_{p.}$
Total	$n_{.1}$	$n_{.2}$	....	$n_{.j}$	....	$n_{.q}$	n

L'effectif théorique au rang (i,j) est $C_{ij} = \dfrac{n_{i.}n_{.j}}{n}$

car si H_0 est vraie on a : $p_{ij} = p_{i.}p_{.j}$; $p_{i.}$ et $p_{.j}$ sont les probabilités marginales.

L'effectif théorique au rang (i,j) est : $np_{i.}p_{ij}$

La quantité $\chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\dfrac{\left(n_{ij} - np_{i.}p_{.j}\right)^2}{np_{i.}p_{.j}}$ est une réalisation du χ^2_{pq-1}

Sur l'échantillon on estime $p_{i.}$ et $p_{.j}$ par respectivement

$\hat{p}_{i.} = \dfrac{n_{i.}}{n}$ et $\hat{p}_{.j} = \dfrac{n_{.j}}{n}$ (on a $(p-1)+(q-1)$ estimations)

On cherche alors la quantité :

$$\chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left(n_{ij} - \dfrac{n_{i.}n_{.j}}{n}\right)^2}{\dfrac{n_{i.}n_{.j}}{n}}$$

qui est une réalisation du $\chi^2_{(p-1)(q-1)}$ à :
$(p-1)(q-1) = pq - 1 - (p-1) - (q-1)$ degrés de liberté.

Pour un seuil de signification α fixé on a :

$$p\left(\chi^2 < \chi^2_{(p-1)(q-1)}(\alpha)\right) = 1-\alpha$$

On accepte H_0 si $\chi^2 < \chi^2_{(p-1)(q-1)}(\alpha)$ et on rejettera H_0 dans le cas contraire.

La valeur $\chi^2_{(p-1)(q-1)}(\alpha)$ est lue dans la table du χ^2 en fonction du seuil α et du nombre de degrés de liberté.

IX- Analyse de la variance (effet d'un facteur)

Soient k échantillons de taille $n_1, n_2, \ldots, n_k$ respectivement $\left(n = \sum_{i=1}^{k} n_i\right)$. On suppose qu'un facteur A influe sur les moyennes de distribution (les variances ne sont pas influées) et que chaque échantillon issu d'une population normale suit une loi normale $N(m_i, \sigma)$.

On teste

$$\begin{cases} H_0 \ : \ m_1 = m_2 = \ldots = m_k \\ H_1 \ : \ \exists i, j \in \{1, 2, \ldots, k\}; \ m_i \neq m_j \end{cases}$$

Soit le tableau

Tableau 5.1

	x_{11} x_{12} x_{1n_1}	x_{21} x_{22} x_{2n_2}		x_{i1} x_{i2} x_{in_i}		x_{k1} x_{k2} x_{kn_k}
moyennes	$\overline{X}_1$	$\overline{X}_2$		$\overline{X}_i$		$\overline{X}_k$

En posant

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij} \quad \textbf{et} \quad X_{ij} - \overline{X} = X_{ij} - \overline{X}_i + \overline{X}_i - \overline{X}$$

On obtient l'équation d'analyse de la variance

$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2 + \frac{1}{n}\sum_{i=1}^{k} n_i\left(\overline{X}_i - \overline{X}\right)^2$$

ou $T^2 = R^2 + L^2$

R^2 : variance résiduelle

L^2 : variance due au facteur

La quantité $\frac{L^2}{\sigma^2}$ suit une loi du χ^2_{n-k} à $n-k$ degrés de liberté .

Les quantités $\frac{T^2}{\sigma^2}$ et $\frac{R^2}{\sigma^2}$ suivent des lois du χ^2_{n-1} et χ^2_{k-1} respectivement .

Si H_0 est vraie , $\frac{L^2}{R^2}\frac{n-k}{k-1} = F(k-1; n-k)$ suit une loi de Snedecor.

On décide H_0 si $\frac{L^2}{R^2}\frac{n-k}{k-1} = f < F^{(\alpha)}(k-1; n-k)$

On rejette H_0 si

$$f \geq F^{(\alpha)}(k-1; n-k) \quad \text{où} \quad F^{(\alpha)}(k-1; n-k)$$

est obtenue par lecture de la table de Fisher-Snedecor et α est le seuil de signification fixé .

EXERCICES

5.1. Quelle décision doit-on prendre dans le cas du test de la moyenne d'une loi normale $N(m,\sigma)$:

$$\begin{cases} H_0 & : \quad m = m_0 \\ H_1 & : \quad m = m_1 \end{cases}$$

σ connu . Calculer la puissance du test .
Application numérique :
$\sigma = 6$; $m_0 = 30$; $m_1 = 34$; $\alpha = 0,05$; $n = 25$
Quelle hypothèse doit-on retenir si l'échantillon donne :
a) $\bar{x} = 31$; b) $\bar{x} = 32$

5.2. Un échantillon de taille n=30 est extrait d'une population suivant une loi normale $N(m,\sigma)$ avec $\sigma = 1,6$:

1- Tester

$$\begin{cases} H_0 & : \quad m = m_0 = 3 \\ H_1 & : \quad m > m_0 \end{cases}$$

au seuil de signification $\alpha = 0,01$

2- Déterminer ß . Tracer la courbe de puissance .

3- Pour n=40 , construire la courbe de puissance correspondante .

5.3 Un fabriquant affirme que la proportion de pièces défectueuses produits par son usine s'élève à p=0,08 . Un utilisateur pour pouvoir choisir entre les hypothèses :

$$H_0 \; : \; p = p_0 = 0,08$$
$$H_1 \; : \; p > 0,08$$

prélève un échantillon aléatoire de taille n=110 . Il choisit de fixer le seuil de signification α à $\alpha = 0,05$.

1- Déterminer le seuil critique

2- Trouver l'erreur de 2° espèce ß .

5.4. Reprendre les données de l'exercice 5.2 et tester :

$$H_0 \; : \; m = m_0 = 3 \qquad ; \qquad H_1 \; : \; m \neq m_0$$

5.5. La durée de vie de composants électroniques provenant d'une usine A suit une loi normale $N(m_1, \sigma_1 = 30)$ et celle de ceux provenant d'une usine B une loi normale $N(m_2, \sigma_2 = 32)$. La durée de vie d'un échantillon de taille $n_1 = 150$ provenant de l'usine A a pour moyenne $\bar{x}_1 = 1852$ heures et celle d'un échantillon de taille $n_2 = 90$ provenant de l'usine B est $\bar{x}_2 = 1840$ heures .

Peut-on conclure , au seuil de signification de 5% , qu'il existe une différence significative entre la durée de vie des composants de l'usine A et celle de B ?

5.6. Une machine remplit des paquets de café . On prélève un échantillon de paquets de taille $n_1 = 120$ de poids moyen 48,53grs et d'écart-type 2,8 grs . Le lendemain on prélève un autre échantillon de taille $n_2 = 270$ de moyenne 50,08 grs et d'écart-type 3,1 grs .

Peut-on conclure , au seuil de signification de 1% qu'il existe une différence significative entre les poids moyens des paquets ?

5.7. Soient deux populations gaussiennes P_1 et P_2 . On suppose que $m_1 = m_2 = m$ et $\sigma_1 = \sigma_2 = \sigma$. On extrait de chaque population un échantillon de taille n_1 et n_2 respectivement . Les moyennes respectives sont $\overline{X}_1$ et $\overline{X}_2$.

Montrer que

$$\frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S'^2_1+(n_2-1)S'^2_2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

suit une loi de Student à $n_1 + n_2 - 2$ degrés de liberté .

5.8. 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 $\overline{x}_1 = 22,235$ grs et $\overline{x}_2 = 21,988$ grs et écarts type $S'_1 = 0,18$ grs et $S'_2 = 0,23$ grs respectivement .

Peut-on conclure que ces deux échantillons proviennent d'une même population pour un seuil de signification de 5% ?

5.9. Deux populations P_1 et P_2 d'étudiants se présentent à un même examen final de mathématiques . On prélève un échantillon de 13 notes de la population P_1 qui donne : 8; 12,5; 10,5; 11; 5; 13; 10; 9,5; 9,5; 11,5; 11; 7; 10. Un échantillon de 10 notes de la population P_2 donne : 12,5; 11; 8; 7; 6; 10; 10; 9,5; 9; 10,5.

Peut-on admettre au seuil de signification de 5% que l'écart-type entre les moyennes des deux populations est significatif?

5.10. Un magasin reçoit pour la vente deux importants lots de pièces électroniques ; l'un provenant d'un fournisseur A , l'autre d'un fournisseur B .
Du lot provenant du fournisseur A , on extrait un échantillon aléatoire de 140 pièces et on trouve 18 pièces défectueuses . Du lot provenant du fournisseur B , un échantillon de 58 pièces donne 8 défectueuses .

On veut savoir si le pourcentage de pièces défectueuses sont différents à un seuil de signification de 5% .

5.11. 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 eux ont contracté la maladie . Un autre échantillon aléatoire de taille $n_2 = 113$ animaux non vaccinés montre que 76 d'entre eux ont contacté la maladie .

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

5.12. Soient deux populations P_1 et P_2 suivant des lois normales de variance respective σ_1^2 et σ_2^2 . De chaque population P_i on extrait un échantillon de taille $n_i (i = 1, 2)$. Soient S'^2_1 et S'^2_2 les variances estimées .

Montrer que si $\sigma_1^2 = \sigma_2^2$, $F = \dfrac{S'^2_1}{S'^2_2}$ suit une loi de Snedecor à $n_1 - 1$ et $n_2 - 1$ degrés de liberté .

5.13. On prélève un échantillon de notes de physique de chacune des deux sections A et B d'étudiants . On obtient :

Tableau 5.2

Section A	12	11,5	6	10	11	5,5	10,5	8
	13	14	12	10,5	11	12,5	11	9

Section B	6	12	11	8,5	12	9,5	10,5	11	15	9,5	8

1- Calculer les valeurs des estimations S'^2_A et S'^2_B des variances inconnues des populations (notes des étudiants)

2- Peut-on dire que les variances σ_A^2 et σ_B^2 sont identiques au seuil de signification $\alpha = 0,05$?

5.14. On étudie le relevé annuel d'absences d'un groupe d'étudiants à un cours de mathématiques hebdomadaire .
On relève :

Tableau 5.3

Nombre d'étudiants absents x_i	Nombres de semaines
0	7
1	11
2	8
3	3
4	2
5	1

1- Trouver le nombre $\bar{x}$ d'absents par cours .

2- Trouver la variance S^2 empirique du nombre d'absents par cours .

3- Peut-on admettre que la distribution empirique peut-être ajustée par une loi de Poisson ?

4- Tester l'ajustement par un test du Khi-carré $\left(\chi^2\right)$ à *un seuil de signification* $\alpha = 0,05$.

5.15. Une machine fabrique des pièces en grand nombre . On effectue un contrôle en prélevant 100 échantillons aléatoires de 30 pièces chacun . On obtient le tableau :

Tableau 5.4

Nombres de pièces défectueuses par échantillon: x_i	0	1	2	3	4	5	6	7
Nombre de lots n_i	4	15	24	23	17	9	6	2

1- Quel est le nombre moyen $\bar{x}$ de pièces défectueuses ?
Trouver la variance empirique S^2 .

2- Peut-on ajuster cette répartition empirique par une loi binomiale ? Vérifier l'ajustement à un seuil de signification $\alpha = 0,05$.

5.16. Les notes de mathématiques d'un groupe d'étudiants lors d'un examen sont réparties selon le tableau :

Tableau 5.5

Notes x	Nombre d'étudiants n_i	Notes x	Nombre d'étudiants n_i
$2 \le x < 4$	2	$10 \le x < 12$	16
$4 \le x < 6$	5	$12 \le x < 14$	12
$6 \le x < 8$	8	$14 \le x < 16$	9
$8 \le x < 10$	14	$16 \le x < 18$	4

1- Par quelle loi de probabilité on peut procéder à un ajustement de la distribution empirique proposée . Trouver l'estimation des paramètres de cette loi .

2- Vérifier l'ajustement par un test du χ^2 à un seuil de signification $\alpha = 0,05$.

3- Vérifier l'ajustement par le test de Kolmogorov à un seuil de signification $\alpha = 0,05$.

5.17. On traite trois échantillons de malades atteints d'une maladie M , selon leur classe d'âge et leur sexe . On veut savoir si l'âge et le sexe peuvent être considéré comme des facteurs indépendants à un seuil de signification de 5% .
Nombre de malades suivant le sexe dans les classes d'âge:

Tableau 5.6

Age : A / Sexe	$A < 30$ans	$30 \le A < 60$	$60 \le A$	Total
Hommes	58	90	49	197
Femmes	14	17	14	45
Total	72	107	63	242

5.18. Soit le tableau suivant représentant les notes obtenues par quatre étudiants A , B , C , D en mathématiques (M) , physique (P) et en chimie (C) lors de trois examens . On suppose que les notes sont normalement distribuées et que les variances sont égales .

Tableau 5.7.

Etudiants / Matières	A	B	C	D
M	11	12	13	12
P	8	11	11	10
C	11	14	12	11

Tester l'hypothèse selon laquelle les moyennes des populations ne sont pas significativement différentes au seuil de signification de 5% .

SOLUTIONS DES EXERCICES

5.1. Soient $\bar{x}$ la moyenne de l'échantillon de taille n et $f(x,m)$ la densité de probabilité

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

Les fonctions de vraisemblance $L(x,m_0)$ et $L(x,m_1)$ sont :

$$L(x,m_0)=f(x_1,m_0)f(x_2,m_0).....f(x_n,m_0)$$

$$=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m_0)^2\right)$$

et

$$L(x,m_1)=f(x_1,m_1)f(x_2,m_1).....f(x_n,m_1)$$

$$=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-m_1)^2\right)$$

Il faut trouver l'ensemble des points tel que : $\dfrac{L(x,m_1)}{L(x,m_0)}\geq K_\alpha$

soit

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-m_i)^2-(x_i-m_0)^2\right]\right\}\geq K_\alpha$$

$$\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[2x_i(m_1-m_0)-(m_1^2-m_0^2)\right]\geq \ln K_\alpha=C$$

$$\frac{m_1 - m_0}{2\sigma^2}\left[2n\bar{x} - (m_1 + m_0)\right] \geq C$$

En supposant que $m_1 - m_0 > 0$ (la démarche reste la même si on suppose $m_1 - m_0 < 0$) on obtient :

$$\bar{x} \geq \frac{1}{2n}\left[m_1 + m_0 + \frac{2C\sigma^2}{m_1 - m_0}\right]$$

On désigne $\frac{1}{2n}\left[m_1 + m_0 + \frac{2C\sigma^2}{m_1 - m_0}\right]$ par A . La région critique W_0 est l'ensemble des points tel que $\bar{x} \geq A$. On décidera H_1 si la moyenne $\bar{x}$ de l'échantillon est supérieure à A .
A est détermine par la résolution de l'équation $\alpha = P\left(\overline{X}_{H_0} \geq A\right)$ où $\overline{X}_{H_0}$ est la variable aléatoire $\overline{X}$ dans l'hypothèse H_0 . On a aussi : $\overline{X}_{H_0}$ suit une loi normale $N\left(m_0, \frac{\sigma}{\sqrt{n}}\right)$ et $\frac{\overline{X}_{H_0} - m_0}{\sigma / \sqrt{n}}$ suit une loi normale N(0,1).
La lecture de la table de la loi normale N(0,1) donne

$$\frac{A - m_0}{\sigma / \sqrt{n}} = u_\alpha$$ On en déduit A .

La règle de décision est :

si $\bar{x} < A$ on décide H_0 c'est à dire $m = m_0$

si $\bar{x} \geq A$ on décide H_1 c'est à dire $m = m_1$

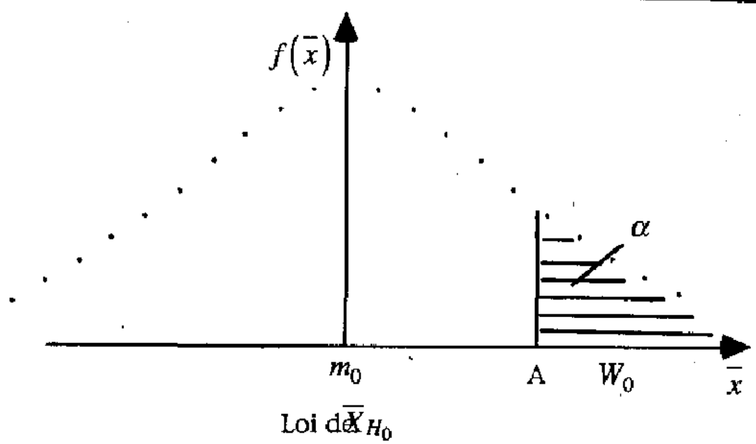


Fig. 5.1

La puissance du test est $1 - \beta = p(W_0 / H_1)$. C'est la probabilité de décider H_1 et H_1 est vraie , c'est à dire $\bar{x} \geq A$ mais $\overline{X}_{H_1}$ suit une loi normale $N\left(m_1, \frac{\sigma}{\sqrt{n}}\right)$

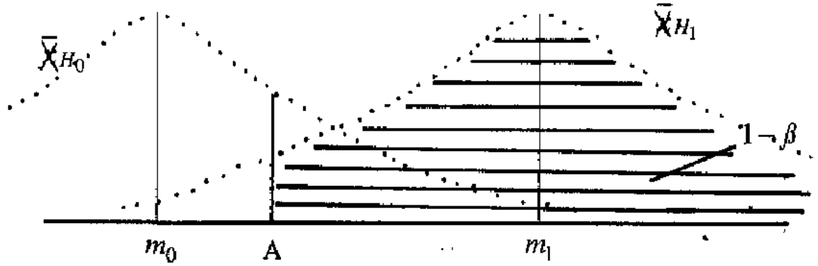


Fig. 5.2

On a :

$$P\left(\overline{X}_{H_1} \geq A\right) = P\left(\frac{\overline{X}_{H_1} - m_1}{\sigma / \sqrt{n}} \geq \frac{A - m_1}{\sigma / \sqrt{n}}\right)$$

$$= P\left(Z \geq u_\beta\right) = 1 - \beta$$

où Z suit une loi normale N(0,1) . La lecture de la table de la loi N(0,1) donne β.
Application numérique . On **détermine** W_0 c'est à dire A pour savoir si $\bar{x} \geq A$ ou $\bar{x} < A$.

On a :

$$A = m_0 + u_\alpha \frac{\sigma}{\sqrt{n}} = 30 + 1,65 \frac{6}{\sqrt{25}} = 31,98$$

Comme $\bar{x} = 31$ dans le 1er**cas** , on décide H_0 c'est à dire $m = m_0$ $\left(\text{car } \bar{x} = 31 < A = 31,98\right)$
Dans le second cas $\bar{x} = 32 > 31,98 = A$, on décide $H_1\left(m = m_1\right)$

Pour la puissance du test $1 - \beta$ **on a:**

$$u_\beta = \frac{A - m_1}{\sigma / \sqrt{n}} = \frac{31,98 - 34}{6 / \sqrt{26}} = -1,68$$

$$1 - \beta = P(Z \geq -1,68) = 1 - P(Z < -1,68) = 1 - \Phi(-1,68)$$

$= \Phi(1,68) = 0,9535$ **où Φ est** la fonction de répartition de Z .
et $\beta = 0,0465$.
La puissance du test $1 - \beta = 0,9535$ est forte . On décidera H_1 , quand elle est vraie , dans 95% des cas .

5.2. 1- Le test est paramétrique . Soit $\bar{x}$ la moyenne de l'échantillon . La règle de décision est :
On décide H_0 si $\bar{x} < A$

ou H_1 si $\bar{x} \geq A$, A est le seuil critique

On a : $\bar{x}$ suit une loi normale $N\left(m_0, \frac{\sigma}{\sqrt{n}}\right) = N\left(3; \frac{1,6}{\sqrt{30}}\right)$

$$P\left(\overline{X} > A\right) = \alpha \quad \text{ou} \quad P\left(\overline{X} \leq A\right) = 1 - \alpha$$

$$P\left(\frac{\overline{X} - 3}{1,6 / \sqrt{30}} \leq \frac{A - 3}{1,6 / \sqrt{30}}\right) = 0,99$$

$$P\left(Z \leq \frac{A - 3}{1,6 / \sqrt{30}}\right) = 0,99$$

où Z suit une loi normale N(0,1)
La lecture de la table de la loi normale N(0,1) donne :

$$\frac{A - 3}{1,6 / \sqrt{30}} = 2,33$$

On obtient A= 3,68 . On décideH_0 si $\bar{x} < 3,68$
ou H_1 si $\bar{x} \geq 3,68$.

2- La valeur du seuil critique A = 3,68 reste la même pour toute valeur de $m > m_0 = 3$ appartenant à l'hypothèse H_1 ; le test est un test upp .
Le risque β est une fonction de m: $\beta = \beta(m)$

$$\beta(m) = P\left(\overline{W}_0 / H_1\right) = P\left(\overline{X} < A / H_1\right)$$

La puissance du test $1 - \beta(m)$ devient alors :

$$1-\beta(m)=P\left(\overline{X}\geq A/H_1\right)$$

$$=P\left(\overline{X}\geq A/m>m_0\right)$$

$$=P\left(\overline{X}\geq 3,68/m>3\right)$$

Comme $\overline{X}$ suit une loi normale $N\left(m,\dfrac{1,6}{\sqrt{30}}\right)$ on a :

$$1-\beta(m)=P\left(\frac{\overline{X}-m}{1,6/\sqrt{30}}\geq\frac{3,68-m}{1,6/\sqrt{30}}\right)$$

et

$$t=\frac{3,68-m}{1,68/\sqrt{30}}=\frac{3,68-m}{0,292}$$

En donnant à m des valeurs particulières on dresse le tableau suivant puis on construit la courbe de puissance .

Tableau 5.8

m	t	$\beta(m)$	$1-\beta(m)$
3,2	1,64	0,9495	0,0505
3,4	0,96	0,8289	0,1711
3,6	0,27	0,6443	0,3557
3,68	0	0,5	0,5
3,8	-0,41	0,3409	0,6591
4	-1,09	0,1379	0,8621
4,2	-1,78	0,0375	0,9625
4,4	-2,47	0,0069	0,9931
4,6	-3,15	0,0008	0,99 2

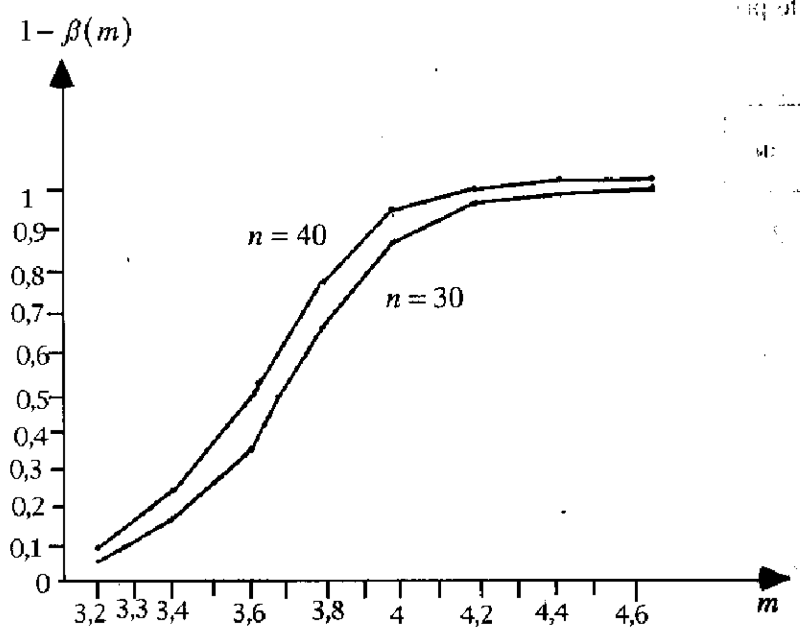


Fig. 5.3

Le test conduit à accepter H_1 quand m = 4,4 dans 99% des cas et à accepter H_1 quand m = 3,4 dans 17% de cas seulement .

3- Si n =40 on a $A=2,33\dfrac{1,6}{\sqrt{40}}+3=3,59$

et

$$1-\beta(m)=P\left(\overline{X}>3,59/m>3\right)$$

soit

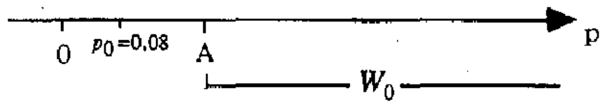
$$t=\frac{3,59-m}{1,6/\sqrt{40}}=\frac{3,59-m}{0,253}$$

On dresse le tableau donnant les valeurs de 1-ß(m) en fonction de m puis on construit la courbe de puissance (dans le précédent système d'axes)

Tableau 5.9

m	t	ß(m)	1-ß(m)	m	t	ß(m)	1-ß(m)
3,2	1,54	0,9382	0,0618	4	-1,62	0,0526	0,9474
3,4	0,75	0,7734	0,2266	4,2	-2,41	0,008	0,9920
3,59	0	0,5	0,5	4,4	-3,20	0,007	0,9993
3,6	-0,04	0,484	0,516	4,6	-3,89	0	1
3,8	-0,83	0,2033	0,7967				

5.3. 1- Soit f la proportion de pièces défectueuses trouvées dans l'échantillon aléatoire de taille **n = 110.**
On décide H_0 si $f < A$ et H_1 si $f \geq A$ où A est le seuil critique .



ON a $\alpha = P(W_0 / H_0)$

$$= P(f \geq A / p = 0,08)$$

On sait que f suit une loi normale $N\left(p_0, \sqrt{\dfrac{p_0(1-p_0)}{n}}\right)$

avec $p_0 = 0,08$

d'où $\alpha = P(f \geq A / p = 0,08)$

$$= P\left(\frac{f - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \geq \frac{A - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}\right)$$

$$= 1 - P\left(Z < \frac{A - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}\right) = 1 - 0,95 = 0,05$$

$$P\left(Z < \frac{A - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}\right) = 0,95$$

Z suit une loi normale N(0,1)

La lecture de la table de la loi normale N(0,1) donne

$$\frac{A - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = 1,645$$

et

$$A = 1,645\sqrt{\frac{0,08.0,92}{110}} + 0,08 = 0,1226$$

On décide

$$H_0 \quad \text{si} \quad f < 12,26\%$$

et

$$H_1 \quad \text{si} \quad f \geq 12,26\%$$

2- La valeur de A reste la même pour toute valeur de p>0,08 . Le test est un test UPP

Le risque de 2°espèce ß est une fonction de p:ß=ß(p)

$$\beta(p) = P(\overline{W}_0 / H_1) = P(f < A / p > 0,08)$$

et

$$1 - \beta(p) = P(f \geq A / p > 0,08)$$

Comme f suit une loi normale $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$ on a :

$$1 - \beta(p) = P\left(\frac{f-p}{\sqrt{\frac{p(1-p)}{n}}} \geq \frac{0,1226-p}{\sqrt{\frac{p(1-p)}{n}}}\right)$$

Soit

$$t = \frac{0,1226-p}{\sqrt{\frac{p(1-p)}{n}}}$$

avec n = 110

En donnant à p des valeurs particulières on trace la courbe de puissance

Tableau 5.10

p	t	ß(p)	1-ß(p)
0,09	1,21	0,8849	0,1151
0,10	0,80	0,7881	0,2119
0,11	0,42	0,6628	0,3372
0,12	0,08	0,5319	0,4681
0,1229	0	0,5	0,5
0,13	-0,22	0,4129	0,5871
0,14	-0,53	0,2981	0,7019
0,15	-0,80	0,2119	0,7881
0,16	-1,07	0,1423	0,8577
0,17	-1,32	0,0934	0,9066
0,18	-1,57	0,0594	0,9406

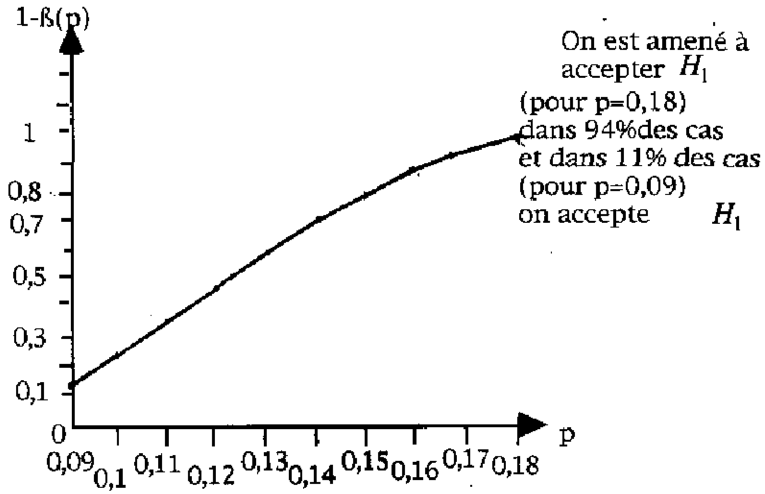


Fig. 5.4

5.4- L'hypothèse H_1 se décompose en deux hypothèses H'_1, H''_1

$$H'_1 : m < m_0$$

$$H''_1 : m > m_0$$

La région d'acceptation de H_0 est définie par :

$$\alpha' = P\left(\overline{X} \le A' / m = m_0\right)$$

relativement au test H'_1 et par :

$$\alpha'' = P\left(\overline{X} \ge A'' / m = m_0\right)$$

relativement au test H''_1

Avec $\alpha' = \alpha'' = \dfrac{\alpha}{2}$ on a donc :

$$\frac{0,01}{2} = P\left(\frac{\overline{X}-3}{1,6/\sqrt{30}} < \frac{A'-3}{1,6/\sqrt{30}}\right)$$

c'est à dire

$$P\left(Z < \frac{A'-3}{1,6/\sqrt{30}}\right) = 0,005$$

où Z suit une loi normale N(0,1) . La lecture de la table de la loi normale N(0,1) donne :

$$\frac{A'-3}{1,6/\sqrt{30}} = -2,58$$

d'où

$$A' = -2,58\frac{1,6}{\sqrt{30}} + 3 = 2,25$$

De la même façon on cherche A" .

On a:

$$\frac{0,01}{2} = P\left(\overline{X} \ge A''\right) = P\left(\frac{\overline{X}-3}{1,6/\sqrt{30}} \ge \frac{A''-3}{1,6\sqrt{30}}\right)$$

ou

$$P\left(Z < \frac{A''-3}{1,6/\sqrt{30}}\right) = 0,995$$

où Z suit une loi normale N(0,1). La lecture de la table de la loi normale N(0,1)donne :

$$\frac{A''-3}{1,6/\sqrt{30}} = 2,58$$

d'où

$$A'' = 2,58\,\frac{1,6}{\sqrt{30}} + 3 = 3,75$$

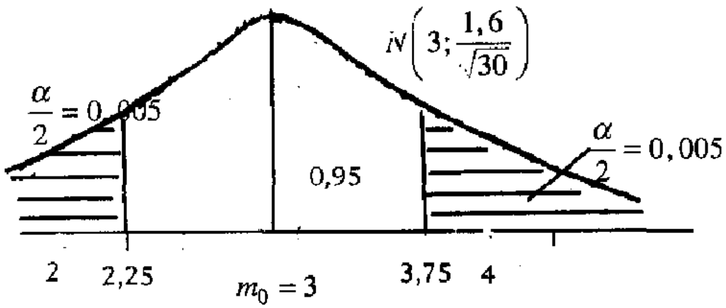


Fig. 5.5

5.5. Il s'agit du test

$$\begin{cases} H_0 \ : \ m_1 - m_2 = 0 \\ H_1 \ : \ m_1 - m_2 \neq 0 \end{cases}$$

La variable aléatoire $\overline{X}_1 - \overline{X}_2$ suit une loi normale

$$N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

On a:

$$\bar{x}_1 - \bar{x}_2 = 12$$

et

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{30^2}{150} + \frac{32^2}{90}} = 4,17$$

L'intervalle d'acceptation de H_0 est :

$$P(|Z| < u_\alpha) = 1 - \alpha$$

où

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

suit une loi normale N(0,1) avec $m_1 - m_2 = 0$

La règle de décision est :
On décide

$$H_0 \quad \text{si} \quad |Z| < u_\alpha$$

et

$$H_1 \quad \text{si} \quad |Z| \geq u_\alpha$$

On a $\alpha = 0,05$ **et** $P(|Z| < u_\alpha) = 0,95$. La lecture de la table de la loi normale **N(0,1)** donne $u_\alpha = 1,96$.
La valeur calculée **de** z **de** Z est :

$$z = \frac{12}{4,17} = 2,88$$

et

$$z = 2,88 > u_\alpha = 1,96$$

On rejette H_0 .

5.6. Les échantillons **de taille** n_1 et n_2 sont suffisamment élevés pour pouvoir **affirmer que** les poids moyens $\overline{X}_1$ et $\overline{X}_2$ suivent une **loi normale** $N(m_1, \sigma_1)$ et $N(m_2, \sigma_2)$ respectivement.

Il s'agit du test

$$\begin{cases} H_0 \ : \ m_1 - m_2 = 0 \\ H_1 \ : \ m_1 - m_2 \neq 0 \end{cases}$$

La variable aléatoire $\overline{X}_1 - \overline{X}_2$ suivent une loi normale :

$$N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

La valeur de $\overline{X}_1 - \overline{X}_2$ est :

$$\bar{x}_1 - \bar{x}_2 = 48,53 - 50,08 = -1,55$$

et l'écart-type

$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ est estimé par :

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(2,8)^2}{120} + \frac{(3,1)^2}{270}} \doteq 0,3176$$

L'intervalle d'acceptation de H_0 est :

$$P(|Z| < u_\alpha) = 1 - \alpha$$

où

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ suit une loi normale N(0,1)et $\alpha = 0,01$

La règle de décision est :

On décide

H_0 si $|Z| < u_\alpha$

et

H_1 si $|Z| \geq u_\alpha$

On a:

$$P(|Z| < u_\alpha) = 0,99$$

d'où $u_\alpha = 2,58$ obtenue par la table de la loi normale N(0,1).

La valeur calculée z de Z est :

$$z = \frac{-1,55}{0,3176} = -4,88$$

et

$$|z| = 4,88 > 2,58 = u_\alpha$$

On rejette H_0 . On conclut qu'il y a une différence significative entre les poids moyens des paquets .

5.7. On sait que si X suit une loi normale N(0,1) et Y suit une loi du Khi-carré (χ_k^2) à k degrés de liberté , $t = \frac{X\sqrt{k}}{\sqrt{Y}}$ suit une loi de Student à k degrés de liberté .

On a:

$\overline{X}_1$ suit une loi normale $N\left(m_1, \frac{\sigma_1}{\sqrt{n_1}}\right)$

et

$\overline{X}_2$ suit une loi normale $N\left(m_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$

donc

$\overline{X}_1 - \overline{X}_2$ suit une loi normale $N\left(m_1 - m_2, \sqrt{\frac{\sigma_2^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

Comme $m_1 = m_2 = m$ et $\sigma_1 = \sigma_2 = \sigma$, $\overline{X}_1 - \overline{X}_2$ suit une loi normale :

$$N\left(0, \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right)$$

et par conséquent :

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{suit une loi normale N(0,1)}$$

Les échantillons suivent des lois normales et ils sont indépendants ; on déduit que :

$$\frac{(n_1-1)}{\sigma_1^2}S'^2_1 = Y_1 \quad \text{et} \quad \frac{(n_2-1)}{\sigma_2^2}S'^2_2 = Y_2$$

suivent des lois du $\chi^2_{n_1-1}$ et $\chi^2_{n_2-1}$ où S'^2_1 et S'^2_2 sont les estimations de σ_1^2 et σ_2^2 , $S'^2_i = \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(X_{ij}-\overline{X}_i\right)^2$

$$Y_1 + Y_2 = \frac{(n_1-1)}{\sigma_1^2}S'^2_1 + \frac{(n_2-1)}{\sigma_2^2}S'^2_2 = \frac{(n_1-1)S'^2_1+(n_2-1)S'^2_2}{\sigma^2}$$

suit une loi du $\chi^2_{n_1+n_2-2}$.

On conclut que :

$$t = \frac{Z\sqrt{n_1+n_2-2}}{\sqrt{Y_1+Y_2}}$$

suit une loi de Student à n_1+n_2-2 degrés de liberté .

et

$$t = \frac{\dfrac{\overline{X}_1-\overline{X}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sqrt{n_1+n_2-2}}{\sqrt{\dfrac{(n_1-1)S'^2_1+(n_2-1)S'^2_2}{\sigma^2}}} =$$

$$\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\dfrac{(n_1-1)S'^2_1+(n_2-1)S'^2_2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

5.8. La distribution étant une loi normale $N(m,\sigma)$, on désire savoir s'il y a une différence significative entre les moyennes inconnues m_1 et m_2 . On suppose que l'écart-type inconnu σ est tel que $\sigma = \sigma_1 = \sigma_2$

On teste :

$$\begin{cases} H_0 \ : \ m_1 - m_2 = 0 \\ H_1 \ : \ m_1 - m_2 \neq 0 \end{cases}$$

La variable aléatoire $\overline{X}_1 - \overline{X}_2$ suit une loi normale

$$N\left(m_1-m_2, \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right) = N\left(m_1-m_2, \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

La variable aléatoire :

$$t = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\dfrac{(n_1-1)S'^2_1+(n_2-1)S'^2_2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \quad \text{suit une loi de}$$

Student à $n_1+n_2-2 = 12+18-2 = 28$ degrés de liberté

Pour k = 28 et $\alpha = 0,05$, la valeur de u_α (tabulée) est :

$$u_\alpha = 2,048 \quad \text{car} \quad P(|t| < u_\alpha) = 1 - \alpha = 0,95$$

On décide

$$H_0 \ : \ \text{si} \quad |t| < u_\alpha$$

et

$$H_1 \ : \ \text{si} \quad |t| \geq u_\alpha$$

Comme la valeur de t calculée est :

$$t = \frac{22,235 - 21,988}{\sqrt{\frac{(12-1)0,0324 + (18-1)0,0529}{12+18-2}}\sqrt{\frac{1}{12}+\frac{1}{18}}} = 3,132$$

On a : $t = 3,123 > u_\alpha = 2,048$; donc on rejette H_0 . La différence est significative entre les moyennes . Les deux échantillons ne proviennent pas de même population .

5.9. On suppose que les notes des étudiants sont distribuées suivant une loi normale $N(m, \sigma)$. On désire savoir s'il y a une différence significative entre les moyennes inconnues m_1 et m_2 des deux populations respectives P_1 et P_2 . On suppose que les écarts-type σ_1 et σ_2 sont tels que $\sigma_1 = \sigma_2 = \sigma$.

Les notes dans les deux populations suivent des lois normales $N(m_1, \sigma_1)$ et $N(m_2, \sigma_2)$ respectivement .

$\overline{X}_1$ suit une loi normale $N\left(m_1, \frac{\sigma}{\sqrt{n_1}}\right)$

$\overline{X}_2$ suit une loi normale $N\left(m_2, \frac{\sigma}{\sqrt{n_2}}\right)$

avec

$n_1 = 13$ et $n_2 = 10$

et

$\overline{X}_1 - \overline{X}_2$ suit une loi normale $N\left(m_1 - m_2, \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$

La variable aléatoire

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S'^2_1 + (n_2-1)S'^2_2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

suit une loi de Student à $n_1 + n_2 - 2 = 21$ degrés de liberté .

La règle de décision est :

on décide

$$H_0 \ \text{si} \ |t| < u_\alpha$$

et

$$H_1 \ \text{si} \ |t| \geq u_\alpha$$

avec $u_\alpha = 2,080$ (valeur tabulée pour k nombre de degrés de liberté égal à 21 et α seuil de signification égal à 0,05)

Comme

$$S'^2_1 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2 = \frac{1}{12}\left[\sum_{i=1}^{13} X_i^2 - 13\overline{X}^2\right] = 4,84$$

et

$$S'^2_2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}\left(X_j - \overline{X}\right)^2 = \frac{1}{9}\left[\sum_{j=1}^{10} X_i^2 - 10\overline{X}^2\right] = 3,725$$

La variable aléatoire t prend alors la valeur :

$$t = \frac{9,88 - 9,35}{\sqrt{\frac{12.4,84 + 9.3,725}{21}}\sqrt{\frac{1}{13} + \frac{1}{10}}} = 0,603$$

On décide H_0 . Au seuil de signification de 5% , la différence n'est pas significative entre les moyennes m_1 et m_2 des deux populations .

5.10. La variable aléatoire $f_1 - f_2$ suit une loi normale :

$$N\left(p_1 - p_2, \sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

où f_1 et f_2 sont les proportions de pièces défectueuses dans les échantillons provenant des lots des fournisseurs A et B respectivement .

Comme le test est :

$$\begin{cases} H_0 \ : \ p_1 - p_2 = 0 \\ H_1 \ : \ p_1 - p_2 \neq 0 \end{cases}$$

où p_1 et p_2 sont les proportions de pièces défectueuses dans les populations (lots fournis par A et B) respectives .

On déduit $f_1 - f_2$ suit une loi normale

$$N\left(0, \sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

La valeur prise par $f_1 - f_2$ est:

$$f_1 - f_2 = \frac{18}{140} - \frac{8}{58} = -0,009$$

celle de f est :

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = 0,132$$

La valeur z de $|Z| = \dfrac{|f_1 - f_2|}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ est :

$$z = \frac{|-0,009|}{\sqrt{0,132(0,868)\left(\frac{1}{140} + \frac{1}{58}\right)}} = 0,17$$

Comme

$$P(|Z| < u_\alpha) = 1 - \alpha = 0,95$$

On décide

H_0 si $|Z| < u_\alpha$

et

H_1 si $|Z| \geq u_\alpha$

La lecture de la table de la loi normale N(0,1) donne :

$$u_\alpha = 1,96 \quad \text{pour} \quad \alpha = 0,05$$

On a :

$$z = 0,17 < 1,96 = u_\alpha$$

La différence n'est pas significative entre les proportions p_1 et p_2 . On accepte H_0 .

5.11. La règle de décision est

On décide

$$H_0 \; : \; p_1 - p_2 = 0$$

ou

$$H_1 \; : \; p_1 - p_2 \neq 0$$

Où p_1 et p_2 sont les proportions d'animaux malades parmi les animaux vaccinés et non vaccinés respectivement .
La variable aléatoire $f_1 - f_2$ suit une loi normale

$$N\left(p_1 - p_2, \sqrt{f(1-f)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right)$$

où f_1 et f_2 sont les proportions d'animaux malades parmi les animaux de l'échantillon d'animaux vaccinés et celui d'animaux non vaccinés respectivement , $p_1 - p_2 = 0$ et où

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

On a donc :

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad \text{suit une loi normale N(0,1)}$$

On décide

$$H_0 \quad \text{si} \quad |Z| < u_\alpha$$

$$H_1 \quad \text{si} \quad |Z| \geq u_\alpha$$

Comme

$$P(|Z| < u_\alpha) = 1 - \alpha = 0,95 \; ,$$

La lecture de la table de la loi N(0,1) donne $u_\alpha = 1,96$

La valeur de z de Z donne :

$$\frac{-0,148}{\sqrt{0,611(0,389)\left(\frac{1}{80}+\frac{1}{113}\right)}} = -2,078$$

On a

$$|z| = 2,078 > 1,96 = u_\alpha$$

On rejette H_0.
Au seuil de signification de 5% , la différence entre les proportions est significative . Il y a 95 chances sur 100 pour que cette différence soit significative .

5.12. On sait que :

$$\frac{(n_1 - 1)}{\sigma_1^2} S'^2_1 = \chi^2_{n_1 - 1}$$

et

$$\frac{(n_2 - 1)}{\sigma_1^2} S'^2_2 = \chi^2_{n_2 - 1}$$

suivent une loi du Khi-carré et que la variable aléatoire F a pour forme :

$$F_{k_1,k_2} = \frac{\chi^2_{k_1} / k_1}{\chi^2_{k_2} / k_2}.$$

où k_1 et k_2 sont les nombres de degrés de liberté .

Le rapport $\dfrac{\chi^2_{n_1-1}}{\chi^2_{n_2-1}}$ vaut :

$$\frac{\chi^2_{n_1-1}}{\chi^2_{n_2-1}} = \frac{(n_1-1)S'^2_1}{\sigma_1^2} \frac{\sigma_2^2}{(n_2-1)S'^2_2}$$

et

$$\frac{S'^2_1}{S'^2_2} = \frac{\chi_{n_1-1} / (n_1-1)}{\chi^2_{n_2-1} / (n_2-1)} = F_{n_1-1,n_2-1} (\text{ car } \sigma_1^2 = \sigma_2^2)$$

5.13. 1- On a :

$$\bar{x}_A = \frac{1}{16}\sum_{i=1}^{16} x_i = \frac{1}{16}(12+11,5+\ldots+9)$$

$$= 10,47$$

$$S'^2_A = \frac{1}{15}\sum_{i=1}^{16}\left(x_i - \bar{x}_A\right)^2 = 5,52$$

et

$$\bar{x}_B = \frac{1}{11}\sum_{j=1}^{11} y_j = \frac{1}{11}(6+12+\ldots+8)$$

$$= 10,27$$

$$S'^2_B = \frac{1}{10}\sum_{j=1}^{11}\left(y_j - \bar{x}_B\right)^2 = 5,72$$

2- on suppose que les notes sont distribuées suivant une loi normale .
On teste au seuil de signification α :

$$\begin{cases} H_0 \ : \ \sigma_A^2 = \sigma_B^2 \\ H_1 \ : \ \sigma_A^2 \neq \sigma_B^2 \end{cases}$$

On a :

$$P\left(F < F^{(\alpha)}_{n_1-1,n_2-1}\right) = 1-\alpha = 0,95$$

La règle de décision est :

On accepte

$$H_0 \text{ si } F < F^{(\alpha)}_{n_1-1,n_2-1}$$

et

$$H_1 \text{ si } F \geq F^{(\alpha)}_{n_1-1,n_2-1}$$

La valeur $F^{(\alpha)}_{15,10}$ obtenue par lecture de la table de Fisher-Snedecor au seuil $\alpha = 0,05$ et à 15 et 10 degrés de liberté est :

$$F^{\alpha}_{15,10} = 2,54$$

On obtient donc :

$$f = 1,036 < 2,54 = F^{\alpha}_{15,10}$$

On accepte H_0 . Au seuil de signification de 5% la **différence** entre les variances σ_A^2 et σ_B^2 n'est pas significative .

5.14. 1- On a :

$$\bar{x} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{7.0+11.1+8.2+3.3+2.4+1.5}{32}$$

$$= 1{,}53 \text{ absences}$$

2-

$$S^2 = \frac{\sum_i n_i x_i^2}{\sum_i n_i} - \bar{x}^2$$

$$= \frac{7.0^2+11.1^2+8.2^2 3.3^2+2.4^2+1.5^2}{32} - (1{,}53)^2$$

$$= 1{,}62 \text{ absence}$$

3- On sait que si une variable aléatoire X suit une loi de Poisson . On a : $E(X) = V(X)$

On peut admettre que la distribution empirique peut-être ajustée par une loi de Poisson de paramètre $\lambda = 1{,}53$.

4- Pour $\lambda = 1{,}53$ on a :

$$P(X=k) = e^{-1{,}53}\frac{(1{,}53)^2}{k!}$$

On dresse le tableau (avec $n = \sum_i n_i = 32$)

Tableau 5.11

x_i	$P(X=k)=p_i$	n_i	np_i
0	0,2165	7	6,928
1	0,3312	11	10,598
2	0,2534	8	8,1088
3	0,1292	3	4,1344
4	0,0494	2	1,5808
5	0,0151	1	0,4832

Les effectifs np_i doivent être supérieurs à 5. On **regroupe les** dernières classes .

On dresse un autre tableau :

Tableau 5.12

x_i	n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
0	7	6,928	0,072	0,0052	0,00075
1	11	10,5984	0,4016	0,1613	0,01521
2	8	8,1088	-0,1088	0,0118	0,00145
3,4,5	6	6,1984	-0,1984	0,0394	0,00635
Total	32	———	———	———	0,02376

La valeur du χ_C^2 calculée est :

$$\chi_C^2 = \sum_i \frac{(n_i - np_i)^2}{np_i} = 0,02376$$

On a :

$$P\left(\chi_C^2 < \chi_k^2(\alpha)\right) = 1 - \alpha = 0,95$$

La règle de décision est :

On accepte l'hypothèse d'ajustement si:

$$\chi_C^2 < \chi_k^2(\alpha)$$

Sinon on la rejette .

Le nombre k de degrés de liberté est : $k = 4 - 1 - 1 = 2$ (4 est le nombre de classes statistiques et on a estimé un seul paramètre λ)

$\sum_i n_i = 32$ et $\frac{\sum_i n_i x_i}{32} = \bar{x} = \lambda$ paramètre estimé

Pour $\alpha = 0.05$ et $k = 2$, la lecture de la table du χ_k^2 donne :

$$\chi_k^2(0,05) = 5,991$$

On obtient

$$\chi_C^2 = 0,02376 < 5,991 = \chi_2^2(0,05)$$

On accepte l'hypothèse d'ajustement par une loi de Poisson.

5.15.

1- On a :

$$\bar{x} = \frac{\sum_i n_i x_i}{\sum_i n_i} = \frac{4.0 + 15.1 + 24.2 + 23.3 + \ldots + 2.7}{100} = 2,95$$

$$\text{et } S^2 = \frac{\sum_i n_i x_i^2}{\sum_i n_i} - \bar{x}^2 = \frac{4.0^2 + 15.1^2 + 24.2^2 + \ldots + 2.7^2}{100} - 2,96^2$$

$$S^2 = 2,59$$

2- On choisit $\underline{p}$ la proportion de pièces détachées défectueuses $p = \frac{x}{30} = 0,0983$. Si l'on doit utiliser directement la table de la loi binomiale, on utilise la loi

B(30;0,10) . Il reste évident qu'on peut utiliser la loi B(30;0,0983) et utiliser une machine programmable .

On a:

$$P(X=k)=C_{30}^{k}(0,10)^{k}(0,90)^{30-k}=p_i$$

Tableau 5.13

x_i	n_i	p_i	np_i	n_i-np_i	$(n_i-np_i)^2$	$\frac{(n_i-np_i)^2}{np_i}$
0	4	0,0424	4,24	-0,24	0,0576	0,0136
1	15	0,1413	14,13	0,87	0,7569	0,0536
2	24	0,2277	22,77	1,23	1,5129	0,0664
3	23	0,2361	23,61	-0,61	0,3721	0,0158
4	17	0,1771	17,71	-0,71	0,5041	0,0284
5	9	0,1023	10;23	-1,23	1,5129	0,1479
6	6	0,0474	4,74	1,26	1,5876	0,3349
7	2	0,0180	1,80	0,2	0,04	0,0222
Total	100	———	———	———	———	———

Les effectifs calculés np_i doivent être supérieurs à 5 . On regroupe les classes .

Tableau 5.14

x_i	n_i	np_i	$(n_i-np_i)^2$	$\frac{(n_i-np_i)^2}{np_i}$
0;1	19	18,37	0,3969	0,0216
2	24	22,77	1,5129	0,0664
3	23	23,61	0,3721	0,0158
4	17	17,71	0,5041	0,0284
5	9	10,23	1,5129	0,1479
6;7	8	6,54	2,1316	0,3259
Total	100	———	———	0,606

La valeur du χ_C^2 calculée est :

$$\chi_C^2=\sum_i \frac{(n_i-np_i)^2}{np_i}=0,606$$

Le nombre k de degrés de liberté tient compte des deux relations qui déterminent la loi binomiale :

$$\sum_i n_i=100 \quad \text{et} \quad \frac{\sum_i n_i x_i}{\sum_i n_i}=2,95$$

(en supposant que dans tous les calculs on a utilisé p = 0,0983 et non 0,1)
Le nombre de degrés de liberté est donc k = 6 - 2 = 4

On a :

$$\chi^2_C = 0,606 < \chi^2_4(0,05) = 9,488$$

On accepte l'hypothèse d'ajustement par une loi binomiale .

5.16. 1- La forme de l'histogramme des fréquences permet d'affirmer qu'on peut proposer un ajustement par une loi normale

Tableau 5.15.

Notes x	Centre de classe x_i	n_i	f_i
$2 \le x < 4$	3	2	0,029
$4 \le x < 6$	5	5	0,071
$6 \le x < 8$	7	8	0,114
$8 \le x < 10$	9	14	0,2
$10 \le x < 12$	11	16	0,229
$12 \le x < 14$	13	12	0,171
$14 \le x < 16$	15	9	0,129
$16 \le x < 18$	17	4	0,057
Total	———	70	1

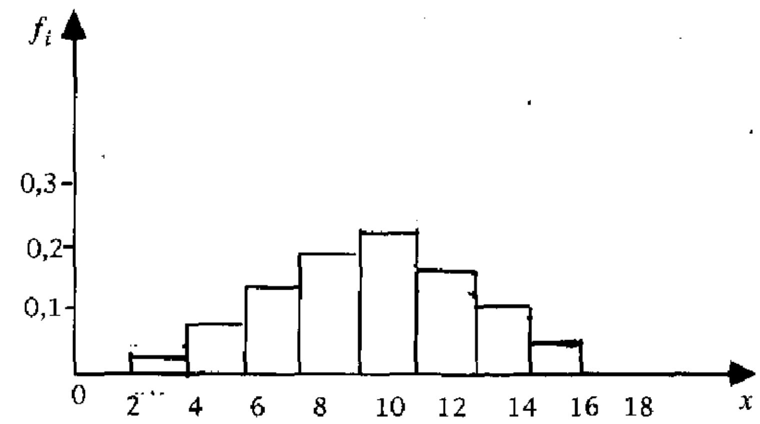


Fig. 5.6.

Les paramètres de la loi normale m et σ sont estimés par $\bar{x}$ et S'.

$$\bar{x} = \frac{\sum_i n_i x_i}{\sum_i n_i} = 10,69$$

$$S'^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 = (3,45)^2 = 11,90$$

L'ajustement est fait par une loi normale N(10,69 ; 3,45)

2- On teste la validité de l'ajustement par une loi normale de densité de probabilité :

$$f(x) = \frac{1}{3,45\sqrt{2\pi}} e^{-\frac{(x-10,69)^2}{2.(3,45)^2}}$$

Les probabilités $P(a \le X \le b) = p_i$ sont calculées au moyen de la table de la loi normale . .

Tableau 5.16.

Notes x	Effectif n_i	$P(a \le X \le b) = p$	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
$2 \le x < 4$	2	0,0203	1,421	0,579	0,2359
$4 \le x < 6$	5	0,0607	4,249	0,751	0,1327
$6 \le x < 8$	8	0,1337	9,359	-1,359	0,1973
$8 \le x < 10$	14	0,2001	14,007	-0,007	0,000003
$10 \le x < 12$	16	0,2273	15,911	0,089	0,00049
$12 \le x < 14$	12	0,1835	12,845	-0,845	0,0556
$14 \le x < 16$	9	0,1067	7,49	1,531	0,3129
$16 \le x < 18$	4	0,0448	3,136	0,864	0,238

Les effectifs calculés np_i doivent être supérieurs à 5 .
On regroupes les classes :

Tableau 5.17

Notes x	n_i	p_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
$2 \le x < 6$	7	0,081	5,67	0,3119
$6 \le x < 8$	8	0,1337	9,359	0,1973
$8 \le x < 10$	14	0,2001	14,007	0,000003
$10 \le x < 12$	16	0,2273	15,911	0,00049
$12 \le x < 14$	22	0,1835	12,845	0,0556
$14 \le x < 18$	13	0,1515	10,605	0,5409
Total	70	-----	-----	1,1062

La valeur du χ_C^2 calculée est:

$$\chi_C^2 = 1,1062$$

Le nombre de degrés de liberté k est :

$$k = (6-1) - 2 = 3$$

car on a deux paramètres estimés m et σ^2 par $\bar{x}$ et S'^2 respectivement et $\sum_i n_i = 70$

A un seuil de signification $\alpha = 0,05$ on a :

$$\chi_3^2(\alpha) = 7,81$$

On a :

$$\chi_C^2 = 1,1062 < 7,81 = \chi_3^2(\alpha)$$

On accepte l'hypothèse d'ajustement par une loi normale N(10,69 ; 3,45)

3- On vérifie l'ajustement par le test de Kolmogorov.

Tableau 5.18

centre de classe x_i	$F_n(x)$	$F(x) = P(X \le x)$	$\lvert F_n(x) - F(x) \rvert$
3	0,0286	0,0203	0,0083
5	0,1	0,081	0,019
7	0,2143	0,2147	0,0004
9	0,4143	0,4148	0,0005
11	0,6429	0,6421	0,0008
13	0,8143	0,8256	0,0113
15	0,9429	0,9323	0,0106
17	1	0,9771	0,0229

L'écart-type maximum est :

$$D_n = |F_n(x) - F(x)| = 0,0229$$

La valeur critique d_n correspondante est (pour n = 70) :

$$d_n = 0,15975$$

d_n est donnée par la table du test de Kolmogorov - Smirnov pour $\alpha = 0,05$ et $n = 70$.

On a :

$$D_n = 0,0229 < 0,15975 = d_n$$

On conclut que l'hypothèse de normalité ne doit pas être rejetée .

5.17. On forme un tableau d'effectifs théoriques . Il y a $\frac{197}{242} = 81,14\%$ d'hommes (ou le total des malades) . On cherche donc les 81,4% de 72;107 et 63 (nombre total de malades selon la tranche d'âge) et on complète le tableau puisque les sommes marginales restent les mêmes . On trouve:

$$\frac{197}{242}.72 = 58,62 \quad ; \quad \frac{197}{242}107 = 87,1 \quad ; \quad \ldots$$

Tableau 5.19

Age=A / Sexe	A< 30	$30 \leq A < 60$	$A \geq 60$	Total
Hommes	58,62	87,1	51,28	197
Femmes	13,38	19,9	11,72	45
Total	72	107	63	242

Le même tableau aurait pu être trouvé en concluant les $\frac{45}{242} = 18,6\%$ de femmes (ou le nombre total de malades) puis on complète le tableau de la même façon que précédemment .

On calcule :

$$\chi_C^2 = \sum_i \sum_j \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}}$$

$$\chi_C^2 = \frac{(58-58,62)^2}{(58,62)} + \frac{(14-13,38)^2}{13,38} + \ldots + \frac{(14-11,72)^2}{11,72}$$

$$= 1,099$$

Le nombre de degrés de liberté est :

$$(p-1)(q-1) = (2-1)(3-1) = 2$$

où p est le nombre de lignes du tableau et q le nombre de colonnes .

La lecture de la table du $\chi_2^2(\alpha)$ où $\alpha = 0,05$ donne :

$$\chi_2^2(\alpha) = 5,991$$

On accepte l'hypothèse suivant laquelle la répartition des malades par classes d'âge n'est pas affectée par leur sexe .

5.18. On a:

$$\overline{X}_1 = \frac{1}{3}(11+8+11) = 10$$

$$\overline{X}_2 = \frac{1}{3}(12+11+14) = 12,33$$

$$\overline{X}_3 = \frac{1}{3}(13+11+12) = 12$$

$$\overline{X}_4 = \frac{1}{3}(12+10+11) = 11$$

et

$$\overline{X} = \frac{\sum_i \sum_j X_{ij}}{12} = 11,33$$

On calcule :

$$\sum_i \left(\overline{X}_i - \overline{X}\right)^2 = (10-11,33)^2 + (12,33-11,33)^2 +$$

$$+(12-11,33)^2 + (11-10,33)^2 = 3,3267$$

puis

$$\sum_j \left(X_{1j} - \overline{X}_i\right)^2$$

Pour cela on fait varier i de 1 à 4 .

$$\sum_j \left(X_{1j} - \overline{X}_1\right)^2 = (11-10)^2 + (8-10)^2 + (11-10)^2 = 6$$

$$\sum_j \left(X_{2j} - \overline{X}_1\right)^2 = (12-12,33)^2 + (11-12,33)^2 + (14-12,33)^2$$

$$= 4,6667$$

$$\sum_j \left(X_{3j} - \overline{X}_3\right)^2 = (13-12)^2 + (11-12)^2 + (12-12)^2 = 2$$

$$\sum_j \left(X_{4j} - \overline{X}_4\right)^2 = (12-11)^2 + (10-11)^2 + (11-11)^2 = 2$$

d'où

$$\sum_i \sum_j \left(X_{ij} - \overline{X}_i\right)^2 = 6 + 4,6667 + 2 + 2 = 14,6667$$

On forme le tableau (avec $\alpha = 0,05$)

Tableau 5.20.

Source de variation	Somme des carrés	Degrés de liberté	Carré moyen	Rapport F
Expliquée	9,98	k-1=3	$\frac{9,98}{k-1} = 3,33$	$F_{3;8} = 1,82$
Résiduelle	14,67	n-k=8	$\frac{14,67}{n-k} = 1,83$	
Totale	24,66	n-1=11		

Le test est:

$$\begin{cases} H_0 \ : \ m_1 = m_2 = m_3 = m_4 \\ H_1 \ : \ \text{Les moyennes } m_1, m_2, m_3 \text{ et } m_4 \text{ ne} \\ \qquad \text{pas égales .} \end{cases}$$

La valeur tabulée $F_{3;8}^{(\alpha)} = 4,07$ est supérieure à la valeur f calculée ($f = F_{3;8} = 1,82$) . On accepte l'hypothèse H_0 . On peut considérer les quatre échantillons comme provenant de la même population (puisque les populations suivent une loi normale et ont même variance) .

CHAPITRE 6

RÉGRESSION SIMPLE - RÉGRESSION MULTIPLE

I- Régression entre deux variables

1- Régression simple .

Soient deux variables X et Y . On recherche une fonction f telle que $f(X)$ soit aussi proche que possible de Y en moyenne quadratique . Le cas le plus utilisé est le modèle linéaire :

$$Y = f(X) = \alpha X + \beta + \varepsilon$$

où ε est la variable résiduelle représentant l'écart entre la valeur ajustée et la valeur observée . On estime les cœfficients α et β de telle sorte à remplacer la série d'observations (x_i, y_i) $i = 1,\ldots,n$ par une équation du type :

$$Y = aX + b$$

En admettant que le modèle $Y = \alpha X + b + \varepsilon$ est vrai , on estime les cœfficients α et β de ce modèle . Pour chaque observation (x_i, y_i) , la valeur calculée $y_i^* = ax_i + b$ par le modèle estimé ne coïncide pas en général avec la valeur $y_i = \alpha x_i + \beta + \varepsilon_i$ observée .

2- Méthode des moindres carrés

Soient deux variables X et Y liées par une certaine relation . Les points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, où x_i et y_i $(i = 1, 2, \ldots, n)$ sont les valeurs prises par X etY respectivement et n la taille de l'échantillon , représentés dans R^2 forment un nuage de points (diagramme de dispersion) . La représentation graphique se fait par une courbe continue (dite courbe d'ajustement) approchant le nuage de points . Entre les deux variables X (variable explicative) et Y (variable expliquée) la relation $y = f(x)$ peut être linéaire ou non linéaire .

a) Droite de régression .

Parmi toutes les droites d'équation $Y = aX + b$ qui approchent le nuage de points , celle qui vérifient la propriété :

$$\sum_{i=1}^{n} (y_i - ax_i - b)^2 \text{ minimale}$$

donne le meilleur ajustement au sens des moindres carrés . Les constantes a et b sont données par la résolution du système :

$$\begin{cases} bn + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

On obtient :

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

et

$$b = \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

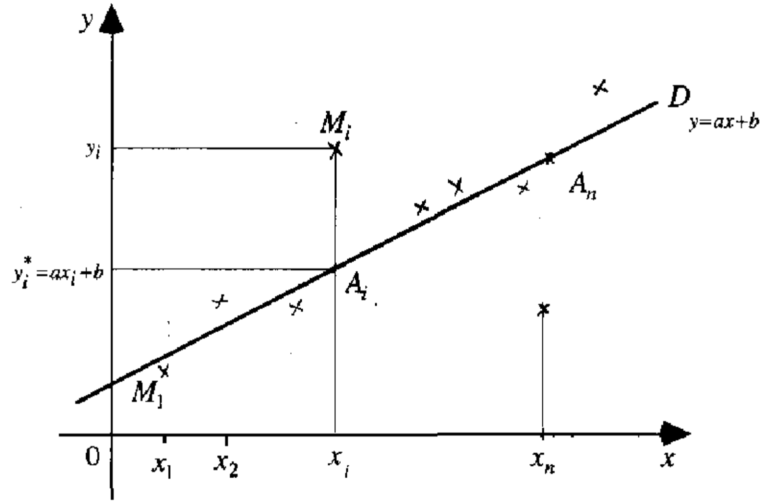


Fig. 6.1

La droite de régression (D) de Y en X a pour équation :

$$Y = aX + b$$

où

$$a = \frac{Cov(X,Y)}{V(X)}$$

et

$$b = \overline{y} - a\overline{X}$$

La droite de régression (D') de X en Y a pour équation :

$$X = a'\,Y + b'$$

où

$$a' = \frac{Cov(X,Y)}{V(Y)}$$

et

$$b' = \overline{X} - a'\,\overline{Y}$$

Les droites (D) et (D') passent toutes les deux par le point $\left(\overline{X},\overline{Y}\right)$

b) Ajustement parabolique .

La parabole d'ajustement du nuage de points $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ a pour équation :

$$Y = aX^2 + bX + c$$

où les constantes a, b et c sont données par la résolution du système :

$$\begin{cases} Cn + b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ C\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + a\sum_{i=1}^{n} x_i^3 = \sum x_i y_i \\ C\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + a\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i \end{cases}$$

c) Autres ajustements .

On utilise souvent des ajustements autres que linéaire . On peut citer :

$$* \quad Y = a_0 + a_1X + a_2X^2 + \ldots + a_nX^n$$

Fonction du n^{ieme} degrés . Les cœfficients $a_0, a_1, \ldots, a_n$ sont obtenus par la résolution du système de n équations à n inconnues .

$$\begin{cases} a_0 + a_1\sum x_i + a_2\sum x_i^2 + \ldots + a_n\sum x_i^n = \sum y_i \\ a_0\sum x_i + a_1\sum x_i^2 + a_2\sum x_i^3 + \ldots + a_n\sum x_i^{n+1} = \sum x_i y_i \\ . \\ . \\ a_0\sum x_i^n + a_1\sum {}_i^{n+1} + a_2\sum x_i^{n+2} + \ldots + a_n\sum x_i^{2n} = \sum x_i^n y_i \end{cases}$$

$$* \quad Y = ab^X$$

Fonction exponentielle . Ce modèle se linéarise en posant :

$$Z = \log Y$$

On obtient :

$$Z = \log a + bX$$

$$* \quad Y = \frac{1}{aX + b}$$

Hyperbole : Ce modèle se linéarise en posant :

$$Z = \frac{1}{Y}$$

On obtient

$$Z = aX + b$$

* $Y = ax^b$

Fonction puissance . Ce modèle se linéarise en posant :

$$Z = \log Y \text{ et } T = \log X$$

On obtient

$$Z = \log a + bT$$

3- Cœfficient de corrélation .

Dans toute la suite , il n'est question que de liaison fonctionnelle linéaire .
La régression de Y en X donne l'équation :

$$Y = aX + b$$

Le cœfficient de corrélation **r** est défini par :

$$r = r_{xy} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2\right]^{\frac{1}{2}}}$$

C'est la mesure de dépendance entre les variables X et Y .

Remarques : On a :

a) $-1 \le r \le 1$

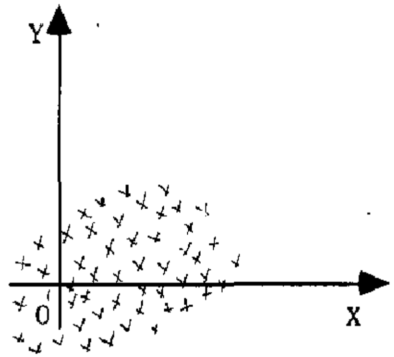
b) $r = 1$ ou $r = -1$ si tous les points observés se trouvait sur une même droite .

c) $r \approx 1$ ou $r \approx -1$ si tous les points observés se trouvent prés d'une même droite .

d) $r = 0$ dans le cas où il y a indépendance entre X et Y . La réciproque est fausse

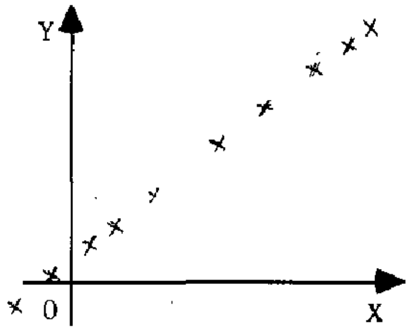
$S_{XY} = Cov(X, Y)$ est dite covariance de X et Y

S_X et S_Y sont les écarts-type de X et Y respectivement .



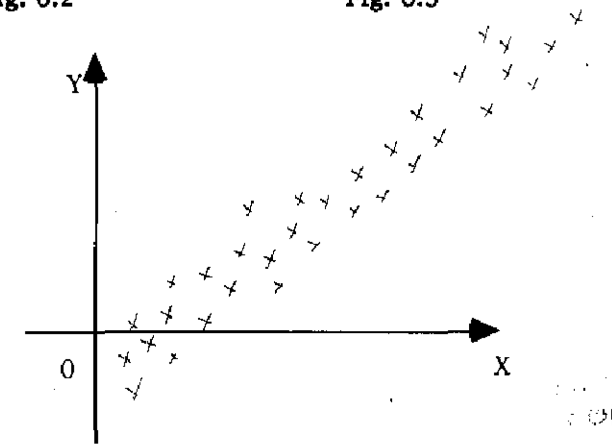
Pas de liaison entre X et Y

Fig. 6.2



Liaison fonctionnelle parfaite

Fig. 6.3



Existence de liaison fonctionnelle entre X et Y

Fig. 6.4

Pour le calcul de r_{xy} on utilise la formule :

$$r_{xy} = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]\left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}$$

L'équation de la droite d'ajustement , au sens des moindres carrés , de Y en X , $Y = aX + b$ s'écrit :

$$Y - \bar{y} = \frac{rS_Y}{S_X}\left(X - \bar{x}\right)$$

L'équation de la droite de X en Y , $X = a'\,Y + b'$ s'écrit :

$$X - \bar{x} = \frac{rS_X}{S_Y}\left(y - \bar{y}\right)$$

Pour les données d'un tableau d'effectifs à double entrée on utilise la formule :

$$r_{xy} = \frac{\frac{1}{n}\sum_{i,j} n_{ij}\left(x_i - \bar{x}\right)\left(y_j - \bar{y}\right)}{\sqrt{\left[\frac{1}{2}\sum_{i=1}^{k} n_{i.}\left(x_i - \bar{x}\right)^2\right]\left[\frac{1}{2}\sum_{j=1}^{l} n_{.j}\left(y_j - \bar{y}\right)^2\right]}}$$

4- Construction d'intervalles de confiance des cœfficients de la droite de régression

Soit le modèle linéaire à une équation :

$$Y = \alpha x + \beta + \varepsilon$$

où Y est la variable expliquée , x la variable explicative et ε la variable résiduelle représentant l'écart entre la valeur ajustée et la valeur observée .
Les cœfficients a et b estimateurs de α et ß sont obtenus par la méthode des moindres carrés . La précision de l'ajustement est définie par la somme des carrés :

$$\sum_i \left(y_i - ax_i - b\right)^2 = \sum_i \varepsilon_i^2$$

Le modèle calculé est :

$$y = ax + b \ ;$$

Le modèle théorique initial est :

$$y = \alpha x + \beta + \varepsilon$$

On suppose :

a) X et Y observées sans erreur ; X_i et ε_i ne sont pas corrélés .

b) $E\left(\varepsilon_i\right) = 0 \quad , \quad \forall_i$

c) $E\left(\varepsilon_i \varepsilon_j\right) = 0 \quad , \quad \text{si } i \neq j$

d) ε_i suit une loi normale $N(0, \sigma)$

e) $V\left(\varepsilon_i\right) = E\left(\left(\varepsilon_i - E\left(\varepsilon_i\right)\right)^2\right) = E\left(\varepsilon_i^2\right) = V(\varepsilon)$ indépendante de i

f) $E\left(\varepsilon_i / X\right) = 0 \ ; \ V\left(\varepsilon_i / X\right) = \sigma^2 \quad i = 1, \ldots, n$

On a :

$$E(a) = \alpha$$

$$E(b) = \beta$$

L'estimateur a est une variable aléatoire normale . La variable aléatoire $\frac{a - E(a)}{\sigma(a)}$ suit une loi normale N (0,1) avec $\sigma^2(a) = V(a)$.
On construit un intervalle de confiance pour le cœfficient α si $\sigma(a)$ était connu . On estime $\sigma(a)$ par $S(a)$.

$$S(a) = \frac{S}{\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{\frac{1}{2}}}$$

où

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} \varepsilon_i^2$$

Le rapport $\frac{a - E(a)}{S(a)}$ suit une loi de Student à n-2 degrés de liberté ; il en est de même pour $\frac{b - E(b)}{S(b)}$ où $S(b)$ est estimateur de $\sigma(b)$.

$$S(b) = S\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right]^{\frac{1}{2}}$$

On construit un intervalle bilatéral symétrique au seuil de signification α :

$$p\left(-u_{\frac{\alpha}{2};n-2} < \frac{a - \alpha}{S(a)} < u_{\frac{\alpha}{2};n-2}\right) = 1 - \alpha$$

et

$$p\left(-u_{\frac{\alpha}{2};n-2} < \frac{b - \beta}{S(b)} < u_{\frac{\alpha}{2};n-2}\right) = 1 - \alpha$$

La valeur $u_{\frac{\alpha}{2};n-2}$ est lue dans la table de la loi de Student à n-2 degrés de liberté .

5- Test du cœfficient de corrélation

a) Première méthode : Pour une distribution normale de X et une distribution normale de Y on teste l'hypothèse ρ=0 (ρ cœfficient de corrélation entre les variables X et Y) caractérisant l'indépendance totale comme suit :
ρ est estimé par r et l'hypothèse ρ =0 sera rejetée lorsque :

$$|t| = / \frac{r}{\sqrt{1-r^2}} / \sqrt{n-2} \geq u_{\frac{\alpha}{2};n-2}$$

La valeur $u_{\frac{\alpha}{2};n-2}$ est lue dans la table de la loi de Student .

b) Deuxième méthode : S'il existe pas de liaison fonctionnelle entre X et Y , la variable aléatoire $Z = Argthr$ suit une loi normale $N\left(0, \frac{1}{\sqrt{n-3}}\right)$ .On a :

$$p\left(|Z| < u_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

L'hypothèse de l'existence d'une liaison fonctionnelle entre X et Y est accepté si :

$$|Z| \geq u_{\frac{\alpha}{2}}$$

Les valeurs $u'_{\frac{\alpha}{2}}$ sont lues dans la table de la loi normale N(0,1)

$$u'_{\frac{\alpha}{2}} = u_{\frac{\alpha}{2}} \sqrt{n-3}$$

car $Z\sqrt{n-3}$ suit une loi normale N(0,1)

II- RÉGRESSION MULTIPLE - CORRÉLATIONS MULTIPLES ET PARTIELLES:

1- Regression multiple

Soient p+1 variables $Y, X_1, X_2, \ldots, X_p$ où Y est la variable à expliquer et les $X_i(i=1,\ldots,p)$, variables explicatives , sont linéairement indépendantes . La liaison est décrite sous la forme :

$$Y = \beta + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n + \varepsilon$$

On estime les paramètres de l'équation par la méthode des moindres carrés pour obtenir une équation du type :

$$Y = b + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n$$

2- Equation de régression La corrélation multiple est la corrélation qui existe entre au moins trois variables . Le modèle linéaire à trois variables s'exprime sous la forme :

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + b$$

Où Y, X_1 et X_2 sont les variables considérées et a_1, a_2 et b sont des constantes .
Cette équation est dite équation de la régression linéaire de Y en X_1 et X_2 .
On note par $x_{11}, x_{22}, \ldots, x_{1k}, \ldots$ les valeurs prises par x_1 et par $x_{21}, x_{22}, \ldots, x_{2l}, \ldots$ les valeurs prises par x_2 . Les valeurs prises par Y sont notées $y_1, y_2, \ldots, y_m, \ldots$
On estime les cœfficients a_1, a_2 et b par la méthode des moindres carrés en cherchant le minimum de la quantité :

$$H(a_1, a_2, b) = \sum_{i=1}^{n} (y_i - a_1 x_{1i} - a_2 x_{2i} - b)^2$$

où n est le nombre d'observations .
Le minimum de $H(a_1, a_2, b)$ est atteint pour :

$$\begin{cases} \dfrac{\partial H(a_1, a_2, b)}{\partial a_1} = 0 \\ \dfrac{\partial H(a_1, a_2, b)}{\partial a_2} = 0 \\ \dfrac{\partial H(a_1, a_2, b)}{\partial b} = 0 \end{cases}$$

Il en résulte les trois équations suivantes :

$$\begin{cases} nb + a_1 \sum_{i=1}^{n} x_{1i} + a_2 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i \\ b \sum_{i=1}^{n} x_{1i} + a_1 \sum_{i=1}^{n} x_{1i}^2 + a_2 \sum_{i=1}^{n} x_{1i} x_{2i} = \sum_{i=1}^{n} x_{1i} y_i \\ b \sum_{i=1}^{n} x_{2i} + a_1 \sum_{i=1}^{n} x_{1i} x_{2i} + a_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} x_{2i} y_i \end{cases}$$

La résolution de ce système donne :

$$a_1 = \frac{Cov(X_1, X_2)Cov(X_2, Y) - Cov(X_1, Y)V(X_2)}{Cov^2(X_1, X_2) - V(X_1)V(X_2)}$$

$$a_1 = \frac{Cov(X_1, X_2)Cov(X_1, Y) - Cov(X_2, Y)V(X_1)}{Cov^2(X_1, X_2) - V(X_1)V(X_2)}$$

$$b = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{x}_2$$

2- Cœfficient de corrélations multiple et partielle

Le cœfficient de corrélation **linéaire** multiple est défini par :

$$ryx_1x_2 = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 r_{yx_1} r_{yx_2} rx_1x_2}{1 - r_{x_1x_2}^2}}$$

où r_{yx_1}, r_{yx_2} et $r_{x_1x_2}$ sont les cœfficient de corrélation linéaire des variables Y et X_1, Y et X_2 et X_1 et X_2 respectivement . Le cœfficient de corrélation partielle entre Y et X_1 noté r_{yx_1,x_2} est défini par :

$$r_{yx_1,x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_2}^2\right)\left(1 - r_{x_1x_2}^2\right)}}$$

Le cœfficient de corrélation partielle entre Y et X_2 noté r_{yx_2,x_1}^2 est défini par :

$$r_{yx_1,x_2} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{\left(1 - r_{y_{x_1}}^2\right)\left(1 - r_{x_2x_1}^2\right)}}$$

Le cœfficient de corrélation partielle entre X_1 et X_2 noté $r_{x_1\ x_2,y}$ est défini par :

$$r_{x_1\ x_2,y} = \frac{r_{x_1x_2} - r_{yx_1} r_{y_{x_2}}}{\sqrt{\left(1 - r_{y_{x_1}}^2\right)\left(1 - r_{yx_2}^2\right)}}$$

Les relations entre les cœfficients de corrélation multiple et partielle sont données par :

$$1 - r_{yx_1x_2}^2 = \left(1 - r_{yx_1}^2\right)\left(1 - r_{yx_2,x_1}^2\right)$$

$$1 - r_{yx_1x_2}^2 = \left(1 - r_{yx_2}^2\right)\left(1 - r_{yx_1,x_2}^2\right)$$

$$1 - r_{yx_1x_2}^2 = \left(1 - r_{x_1x_2}^2\right)\left(1 - r_{x_1\ x_2,y}^2\right)$$

3- Intervalles de confiance Le modèle de **régression** multiple décrit une liaison de la forme :

$$Y = \alpha_1 X_1 + \ldots + \alpha_k X_k + \beta + \varepsilon$$

où $\alpha_i (i = 1,\ldots,k)$ sont des paramètres fixes , ß une constante et ε un terme aléatoire .

Le modèle à trois variables est :

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \beta + \varepsilon$$

A un échantillon de taille n correspond un système de n équations :

$$y_1 = \alpha_1 X_{11} + \alpha_2 X_{21} + \beta + \varepsilon_1$$

$$\vdots$$

$$y_n = \alpha_1 X_{1n} + \alpha_2 X_{2n} + \beta + \varepsilon_n$$

On suppose :

a) Les variables Y, X_1, X_2 sont observées sans erreur .

b) $E(\varepsilon_i) = 0 \quad , \quad \forall_i = 1,\ldots,n$

c) $V(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2 \quad , \quad \forall_i = 1,\ldots,n$

d) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall_i \neq j \ ; \ i,j = 1,\ldots,n$

e) $E(\varepsilon_i / X) = 0 \ ; \ V(\varepsilon_i / X) = \sigma^2 \ ; \ \forall_i = 1,\ldots,n$

f) ε_i suit une loi normale N(0, σ)

On note par b, a_1, a_2 les estimateurs de β, α_1 et α_2 . Ces estimateurs sont choisis par la méthode des moindres carrés .
On a:

$$E(b) = \beta \ ; \ E(a_1) = \alpha_1 \ \text{ et } \ E(a_2) = \alpha_2$$

Comme pour le modèle de régression simple , les variables :

$$T_1 = \frac{a_1 - \alpha_1}{S(a_1)} \quad \text{et} \quad T_2 = \frac{a_2 - \alpha_2}{S(a_2)}$$

suivent une loi de Student à n-2-1 degrés de liberté et

$$p\left(-u_{\frac{\alpha}{2};n-3} < T_k < u_{\frac{\alpha}{2};n-3} \right) = 1 - \alpha \ ; \ k = 1,2$$

où $u_{\frac{\alpha}{2};n-3}$ est donné par la table de la loi de Student .

Les valeurs $S(a_1)$ et $S(a_2)$ des estimateurs se calculent par les formules :

$$S(a_1) = \sqrt{\frac{S^2_{y.x_1x_2}}{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 (1 - r^2_{x_1x_2})}}$$

et

$$S(a_2) = \sqrt{\frac{S^2_{y.x_1x_2}}{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2 (1 - r^2_{x_1x_2})}}$$

Avec

$$S^2_{y.x_1x_2} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-3} \left(1 - \frac{r^2_{x_1y} + r^2_{x_2y} - 2\, r_{x_1x_2} r_{x_1y} r_{x_2y}}{1 - r^2_{x_1x_2}} \right)$$

$$= \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n-3} \cdot \left(1 - r^2_{yx_1x_2}\right)$$

$S^2_{y.x_1x_2}$ variance résiduelle estimée .

EXERCICES

6.1 Montrer que le cœfficient de corrélation linéaire r est tel que : $-1 \leq r \leq 1$

6.2 Soit un nuage de points :

$$M_i\left(x_i, y_i\right) \ ; \ i = 1, \ldots, n$$

Montrer que la droite d'ajustement de Y en X au sens des moindres carrés s'écrit :

$$Y = aX + b$$

où

$$a = \frac{Cov(X,Y)}{V(X)}$$

et

$$b = \overline{Y} - a\overline{X}$$

6.3 Trouver les cœfficient a et b de la droite d'ajustement $Y = aX + b$ au sens des moindres carrés pour un nuage de points $\left(x_i, y_i\right)$; chaque point ayant pour nombre d'observations n_{ij} ; $i, j = 1, \ldots, k$.

6.4 Soit un échantillon aléatoire de taille n . Montrer que la variable aléatoire

$$\frac{a - E(a)}{S(a)} = \frac{a - \alpha}{S(a)}$$

suit une loi de Student à n-2 degrés de liberté .

6.5 Montrer que :

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

suit une loi de Student à n-2 degrés de liberté si les variables aléatoires X et Y sont distribuées normalement et indépendantes . r est le cœfficient de corrélation linéaire .

6.6 Soit la série des 12 mesures suivantes :

Tableau 6.1

x_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	1,5	2,5	2	2	3	4	4	3,5	4,5	4	5	5,5

1- Trouver l'équation (D) d'ajustement , de Y en X , au sens des moindres carrés.

2- Trouver l'équation (D') d'ajustement , de X en Y , au sens des moindres carrés .

3- Calculer le cœfficient de corrélation linéaire entre X et Y .

4- Trouver un intervalle de confiance bilatéral symétrique pour les cœfficients a et b (de la droite (D)) à un seuil de signification $\alpha = 0,05$.

6.7 Le tableau suivant donne pour 100 étudiants les notes de mathématiques (X) et de physique (Y)

Tableau 6.2

Y \ X	$4 \le X < 6$	$6 \le X < 8$	$8 \le X < 10$	$10 \le X < 12$	$12 \le X < 14$	$14 \le X < 16$	Total
$4 \le Y < 6$	2	1					3
$6 \le Y < 8$	4	7		1			12
$8 \le Y < 10$	9	9	1	1			20
$10 \le Y < 12$	9	18	8	10			45
$12 \le Y < 14$		3	1	7	1		12
$14 \le Y < 16$		1	1	5		1	8
Total	24	39	11	24	1	1	100

1- Calculer le cœfficient de corrélation linéaire entre X et Y

2- Peut-on envisager une liaison linéaire entre Y et X .

3- Indiquer l'équation de la droite d'ajustement linéaire de Y en X .

4- Tester l'hypothèse d'indépendance des deux variables aléatoires X et Y en utilisant le test du cœfficient de corrélation à un seuil de signification $\alpha = 0,05$. On supposera que les notes de mathématiques ou de physique sont normalement distribuées .

6.8 Déterminer un ajustement exponentiel , au sens des moindres carrés , de la série suivante :

Tableau 6.3

x_i	52	60	71	90	115	200
y_i	60	50	35	30	20	10

6.9 Un mélange d'huile et de solvant (furfural) est préparé à divers taux (en pourcentage volumique) pour déterminer sa température critique de dissolution . On obtient les résultats suivants :

Tableau 6.4

Taux de solvant %volumique	10	20	30	40	50	60	70	80	90
Température critique de dissolution	65	100	125	140	147	142	133	115	55

1- Représenter le nuage de points .

2- Trouver l'équation de la parabole d'ajustement au sens des moindres carrés .

3- Déterminer les valeurs de tendance pour les différents taux et comparer avec les valeurs réelles .

4- Estimer la température critique de dissolution si on utilise un taux de solvant égal à 65% .

6.10 On mesure l'indice de viscosité en fonction de la température lors de l'extraction de certains composés dans l'huile moteur . On obtient le tableau :

Tableau 6.5

Température	80	90	100	110
Indice de viscosité	82	89	96	97

1- Trouver un ajustement , au sens des moindres carrés , par une fonction puissance .

2- Estimer l'indice de viscosité à température 100 et comparer avec la valeur réelle (96)

6.11 Soient trois variables Y, X_1 et X_2 . Montrer que les cœfficient a_1, a_2 et b de l'équation d'ajustement , au sens des moindres carrés , $Y = a_1X_1 + a_2X_2 + b$ s'écrivent sous la forme:

$$a_1 = \frac{Cov(X_1,X_2)Cov(X_1,Y) - Cov(X_1,Y)V(X_2)}{Cov^2(X_1,X_2) - V(X_1)V(X_2)}$$

$$a_2 = \frac{Cov(X_1,X_2)Cov(X_1,Y) - Cov(X_2,Y)V(X_1)}{Cov^2(X_1,X_2) - V(X_1)V(X_2)}$$

$$b = \overline{Y} - a_1 \overline{X}_1 - a_2 \overline{X}_2$$

6.12 Trouver les cœfficients a_1, a_2, a_3 et b de l'équation d'ajustement au sens des moindres carrés de Y en X_1, X_2 et X_3:

$$Y = a_1 X_1 + a_2 X_2 + a_3 X_3 + b$$

6.13 On mesure les propriétés physico-chimiques des raffinats obtenues de l'extraction avec l'ajout du condensat au solvant (furfural) à température constante . On obtient le tableau suivant :

Tableau 6.6

Condensat / fulfural Poids/poids (Y)	0	0,05	0,10	0,05
Rendement (%) (X_1)	77,8	81	83,5	85
Indice de viscosité (X_2)	96	97	99	100

1- Trouver l'équation d'ajustement , au sens des moindres carrés , de Y en X_1 et X_2:

$$Y = a_1 X_1 + a_2 X_2 + b$$

2- Calculer les cœfficients de corrélation multiple .

SOLUTIONS DES EXERCICES

6.1. On calcule la quantité $E\left[(X-E(X))-\lambda(Y-E(Y))\right]^2$

$$E\left[(X-E(X))-\lambda(Y-E(Y))\right]^2 =$$

$$E\left[(X-E(X))^2-2\lambda(X-E(X))(Y-E(Y))+\lambda^2(Y-E(Y))^2\right]$$

$$= E\left[(X-E(X))^2\right]-2\lambda E\left[(X-E(X))(Y-E(Y))\right]$$

$$+\lambda^2 E\left[(Y-E(Y))^2\right]$$

Cette quantité positive est considérée comme un trinôme du 2° degrés en λ .
Le discriminant $\Delta' \leq 0$ s'écrit :

$$\Delta' = E\left[(X-E(X))(Y-E(Y))\right]^2 -$$

$$E\left[(X-E(X))^2\right]E\left[(Y-E(Y))^2\right]\leq 0$$

On obtient

$$\frac{Cov(X,Y)}{V(X)V(Y)}\leq 1$$

et finalement

$$|r| = \left|\frac{Cov(X,Y)}{S_X S_Y}\right| \leq 1$$

Il y a égalité , c'est à dire :

$$r=\pm 1 \;;(Cov(X,Y)=S_X S_Y)$$

si

$$X-E(X) \text{ et } Y-E(Y)$$

sont proportionnelles soit :

$$X-E(X) = k[Y-E(Y)]$$

6.2 On veut minimiser

$$H(a,b)=\sum_{i=1}^{n}(y_i-ax_i-b)^2$$

Ce minimum est atteint si les dérivées partielles de $H(a,b)$ par rapport à a et à b sont nulles .

$$\frac{\partial H(a,b)}{\partial a}=\frac{\partial \sum (y_i-ax_ib)^2}{\partial a}=0$$

$$-2\sum(y_i-ax_i-b)x_i=0$$

$$\sum x_i y_i - a\sum x_i^2 - b\sum x_i = 0$$

et

$$\frac{\partial H(a,b)}{\partial b}=\frac{\partial \sum (y_i-ax_i-b)^2}{\partial b}=0$$

$$-2\sum(y_i-ax_i-b)=0$$

$$\sum y_i - a\sum x_i - nb = 0$$

On obtient le système de deux équations à deux inconnues :

$$\begin{cases} nb + a\sum x_i = \sum y_i \\ b\sum x_i + a\sum x_i^2 = \sum x_i y_i \end{cases}$$

La première équation du système donne :

$$b + a\frac{\sum x_i}{n} = \frac{\sum y_i}{n}$$

C'est à dire :

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

La deuxième équation s'écrit :

$$b\bar{x} + a\frac{\sum x_i^2}{n} = \frac{\sum x_i y_i}{n}$$

ou

$$(\bar{y} - a\bar{x})\bar{x} + a\frac{\sum x_i^2}{n} = \frac{\sum x_i y_i}{n}$$

$$a\left(\frac{1}{n}\sum x_i^2 - \bar{x}^2\right) = \frac{1}{n}\sum x_i y_i - \bar{x}\,\bar{y}$$

Sachant que :

$$Cov(X,Y) = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n}\sum x_i y_i - -\bar{x} - \bar{y}$$

et

$$V(X) = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

On déduit :

$$aV(X) = Cov(X,Y)$$

La fonction :

$$H(a,b) = \sum (y_i - ax_i - b)^2$$

est minimale pour :

$$a = \frac{Cov(X,Y)}{V(X)} \qquad \text{et} \qquad b = \bar{y} - a\bar{x}$$

En effet : $\forall (a,b) \in R^2$

$$H(a,b) \geq H\left(\frac{Cov(X,Y)}{V(X)}, \bar{y} - a\bar{X}\right) = 0$$

6.3 La droite cherchée est telle que :

$$H(a,b) = \sum_{i,j} (y_j - ax_i - b) n_{ij} \quad \text{minimale}$$

On doit avoir :

$$\begin{cases} \dfrac{\partial H(a,b)}{\partial a} = 0 \\ \dfrac{\partial H(a,b)}{\partial b} = 0 \end{cases}$$

ou bien :

$$\begin{cases} b\sum_{i,j} n_{ij} + a\sum_{i,j} n_{ij}x_i = \sum_{i,j} n_{ij}y_j \\ b\sum_{i,j} n_{ij}x_i + a\sum_{i,j} n_{ij}x_i^2 = \sum_{i,j} n_{ij}x_iy_i \end{cases}$$

Sachant que :

$$\bar{x} = \sum_{i,j} \frac{n_{ij}}{n} x_i \quad ; \quad \bar{y} = \sum_{i,j} \frac{n_{ij}}{n} y_i$$

$$V(X) = \sum_{i,j} \frac{n_{ij}}{n} x_i^2 - \bar{x}^2 \quad ; \quad Cov(X,Y) = \sum_{i,j} \frac{n_{ij}}{n} x_iy_i - \bar{x}\,\bar{y}$$

en divisant par n (nombre total d'observations) les deux équations du système , on obtient :

$$\begin{cases} b + a\bar{x} = \bar{y} \\ b\bar{x} + a\sum_{i,j} \frac{n_{ij}x_i^2}{n} = \frac{1}{n}\sum_{i,j} n_{ij}x_iy_j \end{cases}$$

La première équation donne $b = \bar{y} - a\bar{x}$

La deuxième équation s'écrit :

$$(\bar{y} - a\bar{x})\bar{x} + \frac{a}{n}\sum_{i,j} n_{ij}x_i^2 = \frac{1}{n}\sum_{i,j} x_iy_j$$

ou

$$a\left(\frac{1}{n}\sum_{i,j} n_{ij}x_i^2 - \bar{x}^2\right) = \frac{1}{n}\sum_{i,j} n_{ij}x_iy_j - \bar{x}\,\bar{y}$$

ou encore

$$aV(X) = Cov(X,Y)$$

d'où le résultat :

$$a = \frac{Cov(X,Y)}{V(X)}$$

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

6.4 On sait que $\frac{a-\alpha}{\sigma(a)}$ suit une loi normale N(0,1) . Mais comme l'écart-type $\sigma(a)$ est inconnu , on l'estime à partir de la variable aléatoire :

$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n} \varepsilon_i^2$$

On a : La variable $\sum_{i=1}^{n} \frac{\varepsilon_i^2}{\sigma^2}$ suit une loi du χ^2_{n-2} (Khi-carré à n-2 degrés de liberté) c'est à dire :

$$\frac{n-2}{\sigma^2} S^2 \text{ suit une loi du } \chi^2_{n-2}$$

Mais :

$$\frac{n-2}{\sigma^2} S^2 = \frac{(n-2)S^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2 / \sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{(n-2)S^2(a)}{\sigma^2(a)}$$

D'autre part la variable aléatoire $T = \frac{X\sqrt{k}}{\sqrt{Y}}$ suit une loi de Student à k degrés de liberté si X suit une loi normale N(0,1) et Y suit une loi du χ_k^2.
On conclut que la variable aléatoire

$$\frac{(a-\alpha)/\sigma(a)}{\sqrt{(n-2)S^2(a)/\sigma^2(a)}}\sqrt{n-2} = \frac{a-\alpha}{S(a)}$$

suit une loi de Student à n-2 degrés de liberté .

6.5 On a :

$$\sum(y_i-\bar{y})^2 = \sum(y_i - y + y - \bar{y})^2$$

$$= \sum(y_i-y)^2 + \sum(y-\bar{y})^2 + 2\sum(y_i-y)(y-\bar{y})$$

Par définition de la droite d'ajustement $y = ax+b$, on a :

$$\sum(y_i - y) = 0$$

d'où

$$\sum(y_i-\bar{y})^2 = \sum(y_i-y)^2 + \sum(y-\bar{y})^2$$

En remplaçant y et $\bar{y}$ par leur valeur en fonction de x_i et $\bar{x}$ on peut écrire :

$$\sum(y_i-\bar{y})^2 = \sum(ax_i+b-a\bar{x}-b)^2 = a^2\sum(x_i-\bar{x})^2$$

Sachant que :

$$a = \frac{\sum(y_i-\bar{y})^2(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

et

$$r^2 = \frac{\left[\sum(y_i-\bar{y})(x_i-\bar{x})\right]^2}{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}$$

On obtient :

$$\sum(y-\bar{y})^2 = \frac{\left[\sum(y_i-\bar{y})(x_i-\bar{x})\right]^2}{\sum(x_i-\bar{x})^2} = r^2\sum(y_i-\bar{y})^2$$

On peut écrire donc :

$$\sum(y-\bar{y})^2 = \sum(y_i-y)^2 + r^2\sum(y_i-\bar{y})^2$$

ou

$$\sum(y-\bar{y})^2 = (1-r^2)\sum(y_i-\bar{y})^2$$

La variable aléatoire

$$\frac{\sum(y_i-\bar{y})^2}{n-1}$$

suit une loi du χ_{n-1}^2

La variable aléatoire

$$\frac{\sum(y-\bar{y})^2}{1} = r^2\sum(y_i-\bar{y})^2$$

suit une loi du χ_1^2

La variable aléatoire

$$\frac{\sum(\dot{y}_i - y)^2}{n-2} = \frac{1-r^2}{n-2}\sum\left(y_i - \bar{y}\right)^2$$

suit une loi du χ^2_{n-2}

On conclut que le rapport

$$\frac{\dfrac{\sum(y-\bar{y})^2}{1}}{\dfrac{\sum(y_i - y)^2}{n-2}}$$

(si les variables X et Y suivent des lois normales et sont indépendantes) suit une loi de Snedecor à $k_1 = 1$ et $k_2 = n-2$ degrés de liberté : $F_{1,n-2}$.
Mais ce rapport vaut :

$$\frac{\dfrac{\sum(y-\bar{y})^2}{1}}{\dfrac{\sum(y_i - y)^2}{n-2}} = \frac{r^2(n-2)}{1-r^2}$$

Donc la variable aléatoire

$$\frac{r^2(n-2)}{1-r^2}$$

suit une loi de Snedecor $F_{1,n-2}$; mais comme

$$F_{1,n-2} = \left(t_{n-2}\right)^2$$

où t_k suit une loi de Student à k degrés de liberté , on déduit que :

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2}$$

suit une loi de Student à n-2 degrés de liberté

6.6 1- L'équation de la droite (D) s'écrit : $y = aX + b$. Les cœfficients a et b sont donnés par la résolution du système :

$$\begin{cases} bn + a\sum x_i = \sum y_i \\ b\sum x_i + a\sum x_i^2 = \sum x_i y_i \end{cases}$$

Pour la commodité des calculs , on dresse le tableau :

Tableau 6.7

x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	1,5	1	1,5	2,25
2	2,5	4	5	6,25
3	2	9	6	4
4	2	16	8	4
5	3	25	15	9
6	4	36	24	16
7	4	49	28	16
8	3,5	64	28	12,25
9	4,5	81	40,5	20,25
10	4	100	40	16
11	5	121	55	25
12	5,5	144	66	30,25
$\sum x_i = 78$	$\sum y_i = 41,5$	$\sum x_i^2 =$ 650	$\sum x_i y_i =$ 317	$\sum y_i^2 =$ 161,25

Le système devient :

$$\begin{cases} 12b + 78a = 41,5 \\ 78b + 650a = 317 \end{cases}$$

d'où l'estimation a et b :

$$a = \frac{12.37 - 78.41,5}{12.650 - 78.78} = 0,33$$

$$b = \frac{41,5.650 - 78.317}{12.650 - 78.78} = 1,31$$

et l'équation de la droite (D) s'écrit :

$$y = 0,33x + 1,31$$

2- L'équation de la droite (D') s'écrit :

$$X = a_1 Y + b$$

Les cœfficient a_1 et b_1 sont donnés par la résolution du système :

$$\begin{cases} b_1 n + a_1 \sum y_i = \sum x_i \\ b_1 \sum y_i + a_1 \sum y_i^2 = \sum x_i y_i \end{cases}$$

On estime a_1 et b_1 par :

$$a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum y_i^2 - \left(\sum y_i\right)^2}$$

$$b_1 = \frac{\sum x_i \sum y_i^2 - \sum y_i \sum x_i y_i}{n \sum y_i^2 - \left(\sum y_i\right)^2}$$

d'où

$$a_1 = \frac{12.37 - 78.41,5}{12.161,25 - (41,5)^2} = 2,67$$

$$b_1 = \frac{78.161,25 - 41,5.317}{12.161,25 - (41,5)^2} = 3,72$$

L'équation de la droite (D') s'écrit donc :

$$x = 2,67y - 2,72$$

3- Le cœfficient de corrélation linéaire r_{xy} est :

$$r_{xy} = \frac{n \sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{n \sum x_i^2 - \left(\sum x_i\right)^2} \sqrt{n \sum y_i^2 - \left(\sum y_i\right)^2}}$$

d'où

$$r_{xy} = \frac{12.317 - 78.41,5}{\sqrt{12.650 - 78^2} \sqrt{12.161,25 - (41,5)^2}} = 0,938$$

4- On cherche d'abord :

$$S = \frac{1}{n-2} \sum_{i=1}^{n} \varepsilon_i^2$$

Tableau 6.8

y_i	$y = ax_i + b$	$\varepsilon_i^2 = (y_i - y)^2$
1,5	1,64	0,0196
2,5	1,97	0,2809
2	2,3	0,09
2	2,63	0,3969
3	2,96	0,0016
4	3,29	0,5041
4	3,62	0,1444
3,5	3,95	0,2025
4,5	4,28	0,0484
4	4,61	0,3721
5	4,94	0,0036
5,5	5,27	0,0529
Total	-------	2,117

On a :

$$S = \frac{1}{10} 2,117 = 0,2117$$

et

$$\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2 = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 650 - \frac{1}{12}(78)^2 = 143$$

On détermine ensuite $S(a)$ et $S(b)$

$$S(a) = \frac{S}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} = \frac{0,2117}{\sqrt{143}} = 0,0177$$

$$S(b) = S\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right]^{\frac{1}{2}}$$

$$= 0,2117\left[\frac{1}{12} + \frac{42,25}{143}\right]^{\frac{1}{2}} = 0,1303$$

On désire un intervalle de confiance bilatéral symétrique au seuil $\alpha = 0,05$. On doit avoir donc :

$$P\left(\left|T_{\frac{\alpha}{2};n-2}\right| < u_\alpha\right) = 1 - \frac{\alpha}{2} = 0,975$$

où u_α est donné par la lecture de la table de la loi de Student (nombre de degrés de liberté $k = n$-2 = 10) : $u_\alpha = 2,228$.

On construit les intervalles de confiance au seuil de confiance 0,95 pour a et b :

$$0,33-2,228.0,0177 < a < 0,33+2,228.0,0177$$
$$0,29 < a < 0,37$$

$$1,31-2,228.0,1303 < b < 1,31+2,228.0,1303$$
$$1,02 < b < 1,6$$

6.7 Etant dans un cas de données groupées , on considère que les différentes valeurs de X et de Y coïncident avec les centres de classe.
On obtient le tableau :

Tableau 6.9

Y \ X	5	7	9	11	13	15	$n_{i.}$
5	2	1					3
7	4	7		1			12
9	9	9	1	1			20
11	9	18	8	10			45
13		3	1	7	1		12
15		1	1	5		1	8
$n_{.j}$	24	39	11	24	1	1	100
$n_{.j}x_i$	120	273	99	264	13	15	784
$n_{.j}x_i^2$	600	1911	891	2904	169	225	6700
$x_i\sum_j n_{ij}y_j$	1090	2709	1125	3212	169	225	8530

Y \ X	$n_{i.}y_i$	$n_{i.}y_i^2$	$y_j\sum_i n_{ij}x_i$
5	15	75	85
7	84	588	560
9	180	1620	1152
11	495	5445	3883
13	156	2028	1560
15	120	1800	1290
$n_{.j}$	1050	11556	8530

On a :

$$Cov(X,Y)=\frac{1}{n}\sum_{i,j}n_{ij}\left(x_i-\bar{x}\right)\left(y_j-\bar{y}\right)$$

$$=\frac{1}{n}\sum_{i,j}n_{ij}x_iy_j-\frac{1}{n}\left(\sum_{i,j}n_{ij}x_i\right)\left(\frac{1}{n}\sum_{i,j}n_{ij}y_j\right)$$

$$=\frac{1}{100}8530-\frac{1}{100}784.\frac{1}{100}1050$$

$$=2{,}98$$

$$V(X)=\frac{1}{n}\sum_i n_{i.}\left(x_i-\bar{x}\right)^2=\frac{1}{n}\sum_i n_{i.}x_i^2-\bar{x}^2$$

$$=\frac{1}{100}6700-(7,84)^2$$

$$=5{,}53$$

$$V(Y)=\frac{1}{n}\sum_j n_{.j}\left(y_j-\bar{y}\right)^2=\frac{1}{n}\sum_j n_{.j}y_j^2-\bar{y}^2$$

$$=\frac{1}{100}11556-(10,5)^2$$

$$=5{,}31$$

d'où:

$$r_{xy}=\frac{2,98}{\sqrt{5,53}\sqrt{5,31}}=0,55$$

2- Il n'y a pas de raison de penser qu'il y a une relation linéaire entre X et Y (r=0,55)

3- Il existe toujours une droite , ajustée par la méthode des moindres carrés , d'équation $y = aX + b$. avec :

$$b = \bar{y} - a\bar{x} \quad \text{et} \quad a = \frac{Cov(X,Y)}{V(X)}$$

On obtient alors :

$$a = \frac{2,98}{5,53} = 0,54 \qquad ; \qquad b = 10,5 - 0,54.7,84 = 6,27$$

L'équation de la droite cherchée est :

$$y = 0,54x + 6,27$$

4- On utilise deux méthodes .
a) On a :

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{n-2} = \frac{0,55}{\sqrt{1-0,55^2}}\sqrt{100-2} = 6,52$$

Pour $\alpha = 0,05$, la valeur $u_{\frac{\alpha}{2};n-2}$ lue dans la table de la loi de Student est $u_{\frac{\alpha}{2};n-2}$ =0,1966.

et $|t| = 6,52 > 0,1966 = u_{\frac{\alpha}{2};n-2}$

L'hypothèse ρ =0 (caractérisant l'indépendance totale) est rejetée

b) La variable aléatoire $Z = Argth \ r$ suit une loi normale $N\left(0, \frac{1}{\sqrt{100-3}}\right)$. On a :

$$p\left(\left|Z\sqrt{100-3}\right| < u_\alpha\right) = 1 - \alpha = 0,95$$

On obtient par lecture de la table de la loi normale N(0,1), $u_\alpha = 1,96$. Comme $Z = Argthr = 0,62$ (pour r = 0,55) , on déduit $0,62.\sqrt{100-3} = 6,11 > 1,96$.
On conclut que l'existence d'une liaison fonctionnelle entre X et Y est une hypothèse acceptée

6.8 La fonction d'ajustement est de la forme : $Y = BA^x$ qu'on peut écrire aussi : $LogY = LogB + xLogA$. En posant : $LogY = V$; $LogA = a$ et $LogB = b$, on obtient :

$$V = ax + b$$

On forme le tableau :

Tableau 6.10

x_i	$v_i = Logy_i$	x_i^2	x_iv_i
52	1,778	2704	92,464
60	1,699	3600	101,938
71	1,544	5041	109,629
90	1,477	8100	132,941
115	1,301	13225	149,618
200	1	40000	200
$\sum x_i = 588$	$\sum v_i =$ 8,799	$\sum x_i^2 =$ 72670	$\sum x_iv_i =$ 786,59

Les cœfficients a et b sont donnés par la résolution du système :

$$\begin{cases} bn + a\sum x_i = \sum v_i \\ b\sum x_i + a\sum x_i^2 = \sum x_i v_i \end{cases}$$

qui s'écrit :

$$\begin{cases} 6b + 588a = 8,799 \\ 588b + 72670a = 786,59 \end{cases}$$

d'où les estimations a et b :

$$a = \frac{6.786,59 - 588.8,799}{6.72670 - (588)^2} = -0,005$$

$$b = \frac{8,799.72670 - 786,59.588}{6.72670 - (588)^2} = 1,96$$

Après avoir déterminé a et b on déduit A et B .

$$A = 10^{-0,05} = 0,99 \quad ; \quad B = 10^{1,96} = 91,2$$

L'équation de l'ajustement exponentiel s'écrit alors :

$$Y = 91,2(0,99)^x$$

$$= 91,2e^{-0,01x}$$

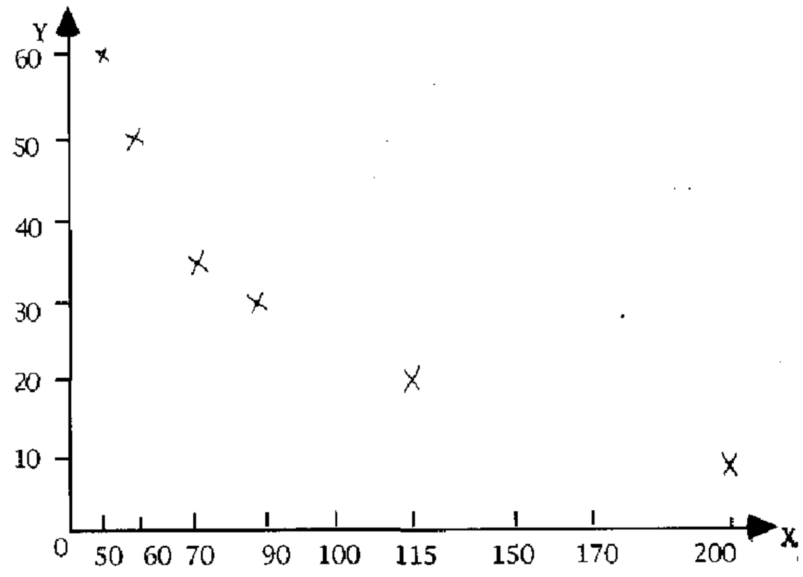


Fig. 6.5

6.9 1- Représentation du nuage de points

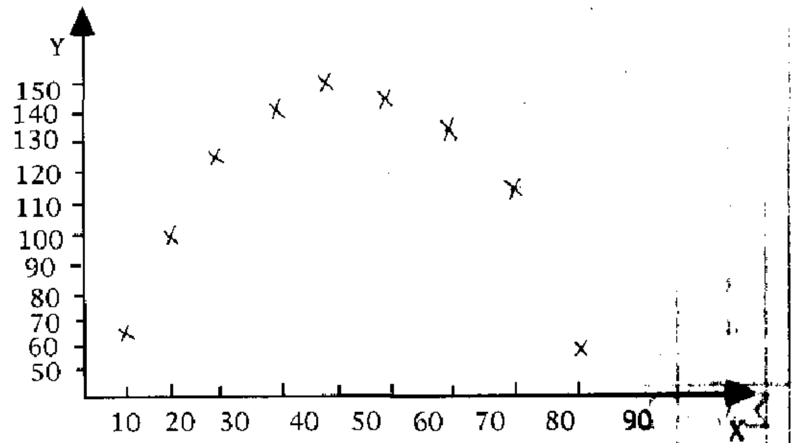


Fig. 6.6

2- On désigne par X le taux de solvant et par Y la température critique de dissolution.
L'équation cherchée est de la forme $Y = aX^2 + bx + c$ où a,b et c sont les solutions du système d'équations :

$$\begin{cases} cn + b\sum x_i + a\sum x_i^2 = \sum y_i \\ c\sum x_i + b\sum x_i^2 + a\sum x_i^3 = \sum x_i y_i \\ c\sum x_i^2 + b\sum x_i^3 + a\sum x_i^4 = \sum x_i^2 y_i \end{cases}$$

Pour la commodité des calculs , on choisit la valeur médiane (taux de solvant à 50%) pour origine des abscisses , c'est à dire X=0 .

On obtient le tableau :

Tableau 6.11

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
-4	65	16	-64	256	-260	1040
-3	100	9	-27	81	-300	900
-2	125	4	-8	16	-250	500
-1	140	1	-1	1	-140	140
0	147	0	0	0	0	0
1	142	1	1	1	142	142
2	133	4	8	16	266	532
3	115	9	27	81	345	1035
4	55	16	64	256	220	880
$\sum x_i =$ 0	$\sum y_i =$ 1022	$\sum x_i^2 =$ 60	$\sum x_i^3 =$ 0	$\sum x_i^4 =$ 708	$\sum x_i y_i$ =23	$\sum x_i^2 y_i$ =5169

La résolution du système :

$$\begin{cases} 9c + 60a = 1022 \\ 60b \qquad = 23 \\ 60c + 708a = 5169 \end{cases}$$

donne
a=5,339 ; b=0,383 ; c=149,147

L'équation cherchée s'écrit alors :

$$Y = -5,339X^2 + 0,383X + 149,147$$

où l'on a pris l'origine X=0 au temps de 50% variant de 10% en 10% .

3-.On dresse le tableau de comparaison suivant :

Tableau 6.12

Taux de solvant (%volumique)	x_i	y_i (réel) Température critique de dissolution	y estimé
10	-4	65	62,191
20	-3	100	99,947
30	-2	125	127,025
40	-1	140	143,425
50	0	147	149,147
60	1	142	144,191
70	2	133	128,557
80	3	115	102,245
90	4	55	65,255

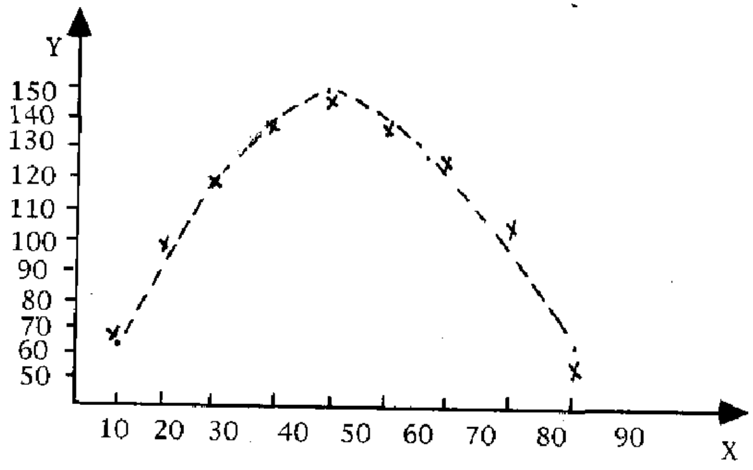


Fig. 6.7

4- Si le taux de solvant vaut 65% , il correspond à une valeur x=1,5 d'où

$$Y = -5,339(1,5)^2 + 0,383(1,5) + 149,147$$

$$Y = 137,71$$

6.10 1- On note X la température et par Y l'indice de viscosité. On veut trouver une fonction de la forme : $Bx^A = Y$. On a :

$$\ln Y = \ln B + A \ln x$$

En posant :

$$\ln Y = u \quad ; \quad \ln B = b \quad \text{et} \quad \ln x = v$$

On obtient :

$$u = AV + b$$

On forme le tableau :

Tableau 6.13

$v_i = \ln x_i$	$u_i = \ln y_i$	v_i^2	$v_i u_i$
4,382	4,407	19,202	19,311
4,5	4,49	20,25	20,205
4,605	4,564	21,206	21,017
4,7	4,575	22,09	21,503
$\sum v_i = 18,187$	$\sum u_i = 18,036$	$\sum v_i^2 = 82,748$	$\sum v_i u_i = 82,036$

Les constantes A et B sont données par la résolution du système :

$$\begin{cases} bn + A\sum v_i = \sum u_i \\ b\sum v_i + A\sum v_i^2 = \sum v_i u_i \end{cases}$$

C'est à dire :

$$\begin{cases} 4b + 18,187A = 18,036 \\ 18,187b + 82,748A = 82,036 \end{cases}$$

On obtient :

$$a = \frac{4.82,036 - 18,187.18,036}{4.82,748 - (18,187)^2} = 0,55$$

et

$$b = \frac{18,036.82,748 - 82,036.18,187}{4.82,748 - (18,187)^2} = 2,02$$

L'équation s'écrit alors : $u = 0,55v + 2,02$

La fonction d'ajustement cherchée est : $Y = 7,54x^{0,55}$

2- A température x=120 , on trouve une estimation de Y:

$$Y = 7,54(120)^{0,55} = 104,94$$

Cette valeur ne correspond pas avec celle de Y réelle expérimentale qui vaut 96 . Ceci montre qu'une relation peut ne pas être satisfaite pour une valeur de x n'appartenant pas à l'intervalle d'étude .

Dans ce cas l'intervalle d'étude est : $I = [80,110]$

6.11 En calculant et en annulant les trois dérivées partielles par rapport à a_1 , à a_2 et à b de :

$$H(a_1, a_2, b) = \sum_{i=1}^{n} (y_i - a_1 x_{1i} - a_2 x_{2i} - b)^2$$

où n est le nombre d'observations , on obtient le système de trois équations :

$$\begin{cases} nb + a_1 \sum x_{1i} + a_2 \sum x_{2i} = \sum y_i \\ b \sum x_{1i} + a_1 \sum x_{1i}^2 + a_2 \sum x_{1i} x_{2i} = \sum x_{1i} y_i \\ b \sum x_{2i} + a_1 \sum x_{1i} x_{2i} + a_2 \sum x_{2i}^2 = \sum x_{2i} y_i \end{cases}$$

En divisant par n la première équation du système on a directement :

$$b = \overline{Y} - a_1 \overline{X}_1 - a_2 \overline{X}_2$$

Les deux autres équations sont divisées par n et résolues simultanément :

$$\begin{cases} (\overline{y} - a_1 \overline{x}_1 - a_2 \overline{x}_2)\overline{x}_1 + a_1 \dfrac{\sum x_{1i}^2}{n} + a_2 \dfrac{\sum x_{1i} x_{2i}}{n} = \dfrac{\sum x_{1i} y_i}{n} \\ (\overline{y} - a_1 \overline{x}_1 - a_2 \overline{x}_2)\overline{x}_2 + a_1 \dfrac{\sum x_{1i}^2 x_{2i}}{n} + a_2 \dfrac{\sum x_{2i}^2}{n} = \dfrac{\sum x_{2i} y_i}{n} \end{cases}$$

ou

$$\begin{cases} a_1 \left(\dfrac{\sum x_{1i}^2}{n} - \overline{x}_1^2 \right) + a_2 \left(\dfrac{\sum x_{1i} x_{2i}}{n} - \overline{x}_1 \overline{x}_2 \right) = \dfrac{\sum x_{1i} y_i}{n} - \overline{y}\,\overline{x}_1 \\ a_1 \left(\dfrac{\sum x_{1i} x_{2i}}{n} - \overline{x}_1 \overline{x}_2 \right) + a_2 \left(\dfrac{\sum x_{2i}^2}{n} - \overline{x}_2^2 \right) = \dfrac{\sum x_{2i} y_i}{n} - \overline{y}\,\overline{x}_2 \end{cases}$$

$$\begin{cases} a_1 V(X_1) + a_2 Cov(X_1, X_2) = Cov(X_1, Y) \\ a_1 Cov(X_1, X_2) + a_2 V(X_2) = Cov(X_2, Y) \end{cases}$$

d'où:

$$a_1 = \frac{Cov(X_1, X_2) Cov(X_1, Y) - Cov(X_1, Y) V(X_2)}{Cov^2(X_1, X_2) - V(X_1) V(X_2)}$$

$$a_2 = \frac{Cov(X_1, X_2) Cov(X_1, Y) - Cov(X_2, Y) V(X_1)}{Cov^2(X_1, X_2) - V(X_1) V(X_2)}$$

6.12 On cherche à minimiser la fonction :

$$H(a_1, a_2, a_3, b) = \sum_{i=1}^{n} (y_i - a_1 x_{1i} - a_2 x_{2i} - a_3 x_{3i} - b)^2$$

où $x_{1i}, x_{2i}, x_{3i}\,(i=1,\dots,n)$ sont les valeurs prises par les variables x_1, x_2 et x_3 respectivement ; n étant le nombre d'observations .
En annulant les dérivées partielles de $H(a_1, a_2\ , a_3, b)$ par rapport à $a_1, a_2\ , a_3$ et b:

$$\frac{\partial H}{\partial a_1}=0 \quad , \quad \frac{\partial H}{\partial a_2}=0 \quad , \quad \frac{\partial H}{\partial a_3}=0 \quad , \quad \frac{\partial H}{\partial b}=0$$

On obtient :

$$\begin{cases} \sum y_i x_{1i} - a_1 \sum x_{1i}^2 - a_2 \sum x_{1i}x_{2i} - a_3 \sum x_{1i}x_{3i} - b \sum x_{1i} = 0 \\ \sum y_i x_{2i} - a_1 \sum x_{1i}x_{2i} - a_2 \sum x_{2i}^2 - a_3 \sum x_{2i}x_{3i} - b \sum x_{2i} = 0 \\ \sum y_i x_{3i} - a_1 \sum x_{1i}x_{3i} - a_2 \sum x_{2i}x_{3i} - a_3 \sum x_{3i}^2 - b \sum x_{3i} = 0 \\ \sum y_i - a_1 \sum x_{1i} - a_2 \sum x_{2i} - a_3 \sum x_{3i} - bn = 0 \end{cases}$$

ou bien

$$\begin{cases} bn + a_1 \sum x_{1i} + a_2 \sum x_{2i} + a_3 \sum x_{3i} = \sum y_i \\ b \sum x_{1i} + a_1 \sum x_{1i}^2 + a_2 \sum x_{1i}x_{2i} + a_3 \sum x_{1i}x_{3i} = \sum x_{1i}y_i \\ b \sum x_{2i} + a_1 \sum x_{1i}x_{2i} + a_2 \sum x_{2i}^2 + a_3 \sum x_{2i}x_{3i} = \sum x_{2i}y_i \\ b \sum x_{3i} + a_1 \sum x_{1i}x_{3i} + a_2 \sum x_{2i}x_{3i} + a_3 \sum x_{3i}^2 = \sum x_{3i}y_i \end{cases}$$

Il reste à résoudre ce système de 4 équations à 4 inconnues .

6.13 1- L'équation d'ajustement au sens des moindres carrés s'écrit : $Y = a_1X_1 + a_2X_2 + b$.
Les constantes a_1, a_2 et b sont la solution du système d'équations :

$$\begin{cases} nb + a_1 \sum x_{1i} + a_2 \sum x_{2i} = \sum y_i \\ b \sum x_{1i} + a_1 \sum x_{1i}^2 + a_2 \sum x_{1i}x_{2i} = \sum x_{1i}y_i \\ b \sum x_{2i} + a_1 \sum x_{1i}x_{2i} + a_2 \sum x_{2i}^2 = \sum x_{2i}y_i \end{cases}$$

On forme le tableau :

Tableau 6.14

y_i	x_{1i}	x_{2i}	x_{1i}^2	x_{2i}^2	$x_{1i}y_i$	$x_{2i}y_i$	$x_{1i}x_{2i}$	y_i^2
0	77,8	96	6052,84	9516	0	0	7468,8	0
0,05	81	97	6561	9409	4,05	4,85	7857	0,0025
0,10	83,6	99	6988,96	9801	8,36	9,9	8276,4	0,01
0,15	85	100	7225	10000	12,75	15	8500	0,0225
$\sum y_i$ =0,3	$\sum x_{1i}$ = 327,4	$\sum x_{2i}$ =392	$\sum x_{1i}^2$ = 26827,8	$\sum x_{2i}^2$ = 38426	$\sum x_{1i}y_i$ = 25,16	$\sum x_{2i}y_i$ = 29,75	$\sum x_{1i}x_{2i}$ = 32102,2	$\sum y_i^2$ = 0,035

Le système d'équations devient :

$$\begin{cases} 4b + 327,4a_1 + 392a_2 = 0,3 \\ 327,4b + 26827,8a_1 + 32102,2a_2 = 25,16 \\ 392b + 32102,2a_1 + 38426a_2 = 29,75 \end{cases}$$

La solution du système est :

$$a_1 = 0,008 \quad ; \quad a_2 = 0,021 \quad ; \quad b = -2,64$$

L'équation cherchée s'écrit :

$$Y = 0,008x_1 + 0,021x_2 - 2,64$$

2- Le cœfficient de corrélation linéaire multiple est :

$$r_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2\,r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}}$$

On calcule d'abord les cœfficients de corrélation linéaire r_{yx_1}, r_{yx_2} et $r_{x_1x_2}$.

On a :

$$r_{yx_1} = \frac{n\sum y_i x_{1i} - \left(\sum y_i\right)\left(\sum x_{1i}\right)}{\sqrt{\left(n\sum y_i^2 - \left(\sum y_i\right)^2\right)\left(n\sum x_{1i}^2 - \left(\sum x_{1i}\right)^2\right)}}$$

$$= \frac{4.25,16 - 0,3.327,4}{\sqrt{\left(4.0,035 - (0,3)^2\right)\left(4.26827,8 - (327,4)^2\right)}} = 0,986$$

$$r_{yx_2} = \frac{n\sum y_i x_{2i} - \left(\sum y_i\right)\left(\sum x_{2i}\right)}{\sqrt{\left(n\sum y_i^2 - \left(\sum y_i\right)^2\right)\left(n\sum x_{2i}^2 - \left(\sum x_{2i}\right)^2\right)}}$$

$$= \frac{4.29,75 - 0,3.392}{\sqrt{\left(4.0,035 - (0,3)^2\right)\left(4.38426 - (392)^2\right)}} = 0,989$$

$$r_{x_1x_2} = \frac{n\sum x_{1i} x_{2i} - \left(\sum x_i\right)\left(\sum x_{2i}\right)}{\sqrt{\left(n\sum x_{1i}^2 - \left(\sum x_{1i}\right)^2\right)\left(n\sum x_{2i}^2 - \left(\sum x_{2i}\right)^2\right)}}$$

$$= \frac{4.32102,2 - 327,4.392}{\sqrt{\left(4.26827,8 - (327,4)^2\right)\left(4.38426 - (392)^2\right)}} = 0,979$$

Par conséquent :

$$r_{yx_1x_2} = \sqrt{\frac{0,986^2 + 0,989^2 - 2.0,986.0,989.0,979}{1 - 0,979^2}}$$

$$= 0,9857$$

2- Les cœfficient de corrélation partielle sont :

$$r_{yx_1,x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_2}^2\right)\left(1 - r_{x_1x_2}^2\right)}} = 0,5893$$

$$r_{yx_2,x_1} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_1}^2\right)\left(1 - r_{x_2x_1}^2\right)}} = 0,6974$$

et

$$r_{x_1x_2,y} = \frac{r_{x_1x_2} - r_{yx_1}r_{yx_2}}{\sqrt{\left(1 - r_{yx_1}^2\right)\left(1 - r_{x_2}^2\right)}} = 0,1559$$

TABLES STATISTIQUES

Table 1

Loi de Poisson

$$P(X = x) = \frac{e^{-m} m^x}{x!.}$$

x \ m	0,5	1,0	1,5	2	2,5	3	3,5	4	4,5	5	5,5
0	0,6065	0,3679	0,223	0,1353	0,0821	0,0498	0,0302	0,0183	0,0111	0,0067	0,0041
1	0,3033	0,3679	0,3347	0,2707	0,2052	0,1494	0,1057	0,0733	0,0500	0,0337	0,0225
2	0,0758	0,1839	0,2510	0,2707	0,2565	0,2240	0,1850	0,1465	0,1125	0,0842	0,0618
3	0,0126	0,0613	0,1255	0,1804	0,2138	0,2240	0,2158	0,1954	0,1687	0,1404	0,1033
4	0,0016	0,0153	0,0471	0,0902	0,1336	0,1680	0,1888	0,1954	0,1898	0,1755	0,1558
5	0,0002	0,0031	0,0141	0,0361	0,0668	0,1008	0,1322	0,1563	0,1708	0,1755	0,1714
6		0,0005	0,0035	0,0120	0,0278	0,0504	0,0771	0,1042	0,1281	0,1462	0,1571
7		0,0001	0,0008	0,0034	0,0099	0,0216	0,0385	0,0595	0,0824	0,1044	0,1234
8			0,0001	0,0009	0,0031	0,0081	0,0169	0,0298	0,0463	0,0653	0,0849
9				0,0002	0,0009	0,0027	0,0066	0,0132	0,0232	0,0363	0,0519
10					0,0002	0,0008	0,0023	0,0053	0,0104	0,0181	0,0285
11					0,0001	0,0002	0,0007	0,0019	0,0043	0,0082	0,0143
12						0,0001	0,0002	0,0006	0,0016	0,0034	0,0065
13							0,0001	0,0002	0,0006	0,0013	0,0028
14								0,0001	0,0002	0,0005	0,0011
15									0,0001	0,0002	0,0004
16										0,0001	0,0001

Table 2

Fonction répartition de la loi normale réduite
Probabilité de trouver une valeur inférieure à x

$$\Phi(x)=\frac{1}{\sqrt{2\Pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$$

x	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	0.7290	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9291	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	0.9779	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	0.9959	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table 3

Loi de χ^2 (K.Pearson)

$p(\chi_\nu^2 \geq x) = \alpha$ .pour $\nu = 10$; $\alpha = 0,01$ on a:

$$p(\chi_{10}^2 \geq 23,21) = 0,01$$

ν \ α	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,0158	0,455	2,71	3,84	5,02	6,63	7,88
2	0,211	1,386	4,61	5,99	7,38	9,21	10,60
3	0,584	2,366	6,25	7,81	9,35	11,34	12,84
4	1,064	3,357	7,78	9,49	11,14	13,28	14,86
5	1,61	4,251	9,24	11,07	12,83	15,09	16,75
6	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	3,49	7,34	13,36	15,51	17,53	20,09	21,96
9	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	4,87	9,34	15,99	18,31	20,48	23,21	25,19
11	5,58	10,34	17,28	19,68	21,92	24,73	26,76
12	6,30	11,34	18,55	21,03	23,34	26,22	28,30
13	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	8,55	14,34	22,31	25,00	27,49	30,58	32,80
16	9,31	15,34	23,54	26,30	28,85	32,00	34,27
17	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	10,86	17,34	25,99	28,87	31,53	34,81	37,16
19	11,65	18,34	27,20	30,14	32,85	36,19	38,58
20	12,44	19,34	28,41	31,41	34,17	37,57	40,00
21	13,24	20,34	29,62	32,67	35,48	38,93	41,40
22	14,04	21,34	30,81	33,92	36,78	40,29	42,80
23	14,85	22,34	32,01	35,17	38,08	41,64	44,18
24	15,66	23,34	33,20	36,42	39,36	42,98	45,56
25	16,47	24,34	34,38	37,65	40,65	44,31	46,93
26	17,29	25,34	35,56	38,89	41,92	45,64	48,29
27	18,11	26,34	36,74	40,11	43,19	46,96	49,64
28	18,94	27,34	37,92	41,34	44,46	48,28	50,99
29	19,77	28,34	39,09	42,56	45,72	49,59	52,34
30	20,60	29,34	40,26	43,77	46,98	50,89	53,67

Table 4

Loi t de Student

ν \ α	0,10	0,05	0,025	0,01	0,005
1	3,087	6,314	12,706	30,821	63,657
2	1,886	2,920	4,303	6,905	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,385	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921
17	1,333	1,740	2,110	2,5672	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,085	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
60	1,296	1,671	2,000	2,390	2,660
120	1,289	1,658	1,980	2,358	2,617
∞	1,282	1,645	1,960	2,326	2,576

La table donne : $p(t_\nu \geq x) = \alpha$

Exemple: Pour $\alpha = 0,025$ et $\nu = 10$ on a:

$p(t_{10} \geq 2,228) = 0,025$; $p(t_{10} < -2,228) = 0,025$;

$p(|t_{10}| \geq 2,228) = 0,05$

Table 5

Loi de Fisher - Snedecor

$$P\big(F(\nu_1, \nu_2) \geq x\big) = \alpha$$

Exemple pour $\alpha = 0,05$

$\nu_1 = 5$, $\nu_2 = 10$ on a $x = 3,33$

$\alpha = 0,05$

ν_2 \ ν_1	1	2	3	4	5
1	161,4	199,5	215,7	224,6	230,2
2	18,51	19,00	19,16	19,25	19,30
3	10,13	9,55	9,28	9,12	9,01
4	7,71	6,94	6,59	6,39	6,26
5	6,61	5,79	5,41	5,19	5,05
6	5,99	5,14	4,79	4,53	4,39
7	5,59	4,74	4,35	4,12	3,97
8	5,32	4,46	4,07	3,84	3,69
9	5,12	4,26	3,86	3,63	3,48
10	4,96	4,10	3,71	3,48	3,33
11	4,84	3,98	3,59	3,36	3,20
12	4,75	3,88	3,49	3,26	3,11
13	4,67	3,80	3,41	3,18	3,02
14	4,60	3,74	2,34	3,11	2,96
15	4,54	3,68	3,29	3,06	2,90
16	4,49	3,63	3,24	3,01	2,85
17	4,45	3,59	3,20	2,96	2,81
18	4,41	3,55	3,16	2,93	2,77
19	4,38	3,52	3,13	2,90	2,74
20	4,35	3,49	3,10	2,87	2,71
21	4,32	3,47	3,07	2,84	2,68
22	4,30	3,44	3,05	2,82	2,66
23	4,28	3,42	3,03	2,80	2,64
24	4,26	3,40	3,01	2,78	2,62
25	4,24	3,38	2,99	2,76	2,60
26	4,22	3,37	2,98	2,74	2,59
27	4,21	3,35	2,96	2,73	2,57
28	4,20	3,34	2,95	2,71	2,56
29	4,18	3,33	2,93	2,70	2,54
30	4,17	3,32	2,92	2,69	2,53
40	4,08	3,23	2,84	2,61	2,45
60	4,00	3,15	2,76	2,52	2,37
120	3,92	3,07	2,68	2,45	2,29
∞	3,84	2,99	2,60	2,37	2,21

Table 5 (suite)

ν_2 \ ν_1	6	8	12	24	∞
1	234,0	238,9	243,9	249,0	254,3
2	19,33	19,37	19,41	19,45	19,50
3	8,94	8,84	8,74	8,64	8,53
4	6,16	6,04	5,91	5,77	5,63
5	4,95	4,82	4,68	4,53	4,36
6	4,28	4,15	4,00	3,84	3,67
7	3,87	3,73	3,57	3,41	3,23
8	3,58	3,44	3,28	3,12	2,93
9	3,37	3,23	3,07	2,90	2,71
10	3,22	3,07	2,91	2,74	2,54
11	3,09	2,95	2,79	2,61	2,40
12	3,00	2,85	2,69	2,50	2,30
13	2,92	2,77	2,60	2,42	2,21
14	2,85	2,70	2,53	2,35	2,13
15	2,79	2,64	2,48	2,29	2,07
16	2,74	2,59	2,42	2,24	2,01
17	2,70	2,55	2,38	2,19	1,96
18	2,66	2,51	1,34	2,15	1,92
19	2,63	2,48	2,31	2,11	1,88
20	2,60	2,45	2,28	2,08	1,84
21	2,57	2,42	2,25	2,05	1,81
22	2,55	2,40	2,23	2,03	1,78
23	2,53	2,38	2,20	2,00	1,76
24	2,51	2,36	2,18	1,98	1,73
25	2,49	2,34	2,16	1,96	1,71
26	2,47	2,32	2,15	1,95	1,69
27	2,46	2,30	2,13	1,93	1,67
28	2,44	2,29	2,12	1,91	1,65
29	2,43	2,28	2,10	1,90	1,64
30	2,42	2,27	2,09	1,89	1,62
40	2,34	2,18	2,00	1,79	1,51
60	2,25	2,10	1,92	1,70	1,39
120	2,17	2,01	1,83	1,61	1,25
∞	2,09	1,94	1,75	1,52	1,00

Table 5(suite 2)

$\alpha = 0,01$

ν_2 \ ν_1	1	2	3	4	5
1	4052	4999	5403	5625	5794
2	98,49	99,00	99,17	99,25	99,30
3	34,12	30,81	29,46	28,71	28,24
4	21,20	18,00	16,69	15,98	15,52
5	16,26	13,27	12,06	11,39	10,97
6	13,74	10,91	9,78	9,15	8,75
7	12,25	9,55	8,45	7,85	7,45
8	11,26	8,65	7,59	7,01	6,63
9	10,56	8,02	6,99	6,42	6,06
10	10,04	7,58	6,55	5,99	5,64
11	9,65	7,20	6,22	5,67	5,32
12	9,33	6,93	5,95	5,41	5,06
13	9,07	6,70	5,74	5,20	4,86
14	8,86	6,51	5,56	5,03	4,69
15	8,68	6,38	5,42	4,89	4,56
16	8,53	6,23	5,29	4,77	4,44
17	8,40	6,11	5,18	4,67	4,34
18	8,28	6,01	5,05	4,58	4,25
19	8,18	5,93	5,01	4,50	4,17
20	8,10	5,85	4,94	4,43	4,10
21	8,02	5,78	4,87	4,37	4,04
22	7,94	5,72	4,82	4,31	3,99
23	7,88	5,66	4,76	4,26	3,94
24	7,82	5,61	4,72	4,22	3,90
25	7,77	5,57	4,68	4,18	3,86
26	7,72	5,53	4,64	4,14	3,82
27	7,68	5,49	4,60	4,11	3,78
28	7,64	5,45	4,57	4,07	3,75
29	7,60	5,42	4,54	4,04	3,73
30	7,56	5,39	4,51	4,02	3,70
40	7,31	5,18	4,31	3,83	3,51
60	7,08	4,98	4,13	3,65	3,34
120	6,85	4,79	3,95	3,48	3,17
∞	6,64	4,60	3,78	3,32	3,02

Table 5 (suite3)

ν_2 \ ν_1	6	8	12	24	∞
1	5859	5981	6106	6234	6366
2	99,33	99,36	99,42	99,46	99,50
3	27,91	27,49	27,05	26,60	26,12
4	15,21	14,80	14,37	13,93	13,46
5	10,67	10,27	9,89	9,47	9,02
6	8,47	8,10	7,72	7,31	6,88
7	7,19	6,84	6,47	6,07	5,65
8	6,37	6,03	5,67	5,28	4,86
9	5,80	5,47	5,11	4,73	4,31
10	5,39	5,06	4,71	4,33	3,91
11	5,07	4,74	4,40	4,02	3,60
12	4,82	4,50	4,16	3,78	3,36
13	4,62	4,30	3,96	3,59	3,16
14	4,46	4,14	3,80	3,43	3,00
15	4,32	4,00	3,67	3,29	2,87
16	4,20	3,89	3,55	3,18	2,75
17	4,10	3,79	3,45	3,08	2,65
18	4,01	3,71	3,37	3,00	2,57
19	3,94	3,63	3,30	2,52	2,49
20	3,87	3,56	3,23	2,86	2,42
21	3,81	3,51	3,17	2,80	2,36
22	3,78	3,45	3,12	2,75	2,31
23	3,71	3,41	3,07	2,70	2,28
24	3,67	3,36	3,03	2,68	2,21
25	3,63	3,32	2,99	2,62	2,17
26	3,59	3,29	2,96	2,58	2,13
27	3,56	3,26	2,93	2,55	2,10
28	3,53	3,23	2,90	2,52	2,06
29	3,50	3,20	2,87	2,49	2,03
30	3,47	3,17	2,84	2,47	2,01
40	3,29	2,99	2,66	2,29	1,80
60	3,12	2,82	2,50	2,12	1,60
120	2,96	2,66	2,34	1,95	1,38
∞	2,80	2,51	2,18	1,79	1,00

Table 6

Transformation $Z = Argthr$.

$r = thz$ exemple $r = 0,4219$ pour $z = 0,450$

$z = 1,570$ pour $r = 0,9170$

z	0,01	0,02	0,03	0,04	0,05
0.0	0.0100	0.0200	0.0300	0.0400	0.0500
0.1	0.1096	0.1194	0.1293	0.1391	0.1489
0.2	0.2070	0.2165	0.2260	0.2355	0.2449
0.3	0.3004	0.3095	0.3185	0.3275	0.3364
0.4	0.3885	0.3969	0.4053	0.4136	0.4219
0.5	0.4699	0.4777	0.4854	0.4930	0.5005
0.6	0.5441	0.5511	0.5580	0.5649	0.5717
0.7	0.6107	0.6169	0.6231	0.6291	0.6351
0.8	0.6696	0.6751	0.6805	0.6858	0.6911
0.9	0.7211	0.7259	0.7306	0.7352	0.7398
1.0	0.7658	0.7699	0.7735	0.7779	0.7818
1.1	0.8041	0.8076	0.8110	0.8144	0.8178
1.2	0.8367	0.8397	0.8426	0.8455	0.8483
1.3	0.8643	0.8668	0.8692	0.8717	0.8741
1.4	0.8875	0.8896	0.8917	0.8937	0.8957
1.5	0.9069	0.9087	0.9104	0.9121	0.9138
1.6	0.9232	0.9246	0.9261	0.9275	0.9289
1.7	0.9366	0.9379	0.9391	0.9402	0.9414
1.8	0.94763	0.94884	0.94983	0.95080	0.95175
1.9	0.95709	0.95742	0.95873	0.95953	0.96032
2.0	0.96473	0.96541	0.96609	0.96675	0.96739
2.1	0.97103	0.97159	0.97215	0.97269	0.97323
2.2	0.97622	0.97668	0.97714	0.97752	0.97803
2.3	0.98049	0.98087	0.98124	0.98161	0.98197
2.4	0.98399	0.98431	0.98462	0.98492	0.98522
2.5	0.98888	0.98714	0.98739	0.98764	0.98788
2.6	0.98924	0.98945	0.98966	0.98987	0.99007
2.7	0.99118	0.99136	0.99153	0.99170	0.99185
2.8	0.99278	0.99292	0.99306	0.99320	0.99333
2.9	0.99408	0;99420	0.99431	0.99443	0.99454

table 6 (suite)

z	0,06	0,07	0,08	0,09	0,10
0.0	0.0599	0.0699	0.0798	0.0898	0.0997
0.1	0.1586	0.1684	0.1781	0.1877	0.1974
0.2	0.2548	0.2636	0.2729	0.2821	0.2913
0.3	0.3452	0.3540	0.3627	0.3714	0.3800
0.4	0.4301	0.4382	0.4462	0.4542	0.4621
0.5	0.5080	0.5154	0.3227	0.5299	0.5370
0.6	0.5784	0.5850	0.5915	0.5980	0.6044
0.7	0.6411	0.6169	0.6527	0.6584	0.6840
0.8	0.6963	0.7014	0.7064	0.7114	07163
0.9	0.7443	0.7487	0.7531	0.7574	0.7616
1.0	0.7857	0.7895	0.7932	0.7969	0.8005
1.1	0.8210	0.8243	0.8275	0.8306	0.8337
1.2	0.8511	0.8536	0.8565	0.8591	0.8617
1.3	0.8764	0.8787	0.8810	0.8832	0.8854
1.4	0.8977	0.8996	0.9015	0.9033	0.9051
1.5	0.9154	0.9170	0.9186	0.9201	0.9217
1.6	0.9302	0.9316	0.9329	0.9341	0.9354
1.7	0.9425	0.9436	0.9447	0.9458	0.94681
1.8	0.95268	0.95359	0.95449	0.95537	0.95624
1.9	0.96109	0.96185	0.96259	0.96331	0.96403
2.0	0.96803	0.96865	0.96926	0.96984	0.97045
2.1	0.97375	0.97426	0.97477	0.97526	0.97574
2.2	0.97846	0.97888	0.97929	0.97970	0.98010
2.3	0.98233	0.98267	0.98301	0.98335	0.98367
2.4	0.98551	0.98579	0.98607	0.98635	0.98861
2.5	0.98812	0.98835	0.98858	0.98881	0.98903
2.6	0.99026	0.99045	0.99064	0.99083	0.99101
2.7	0.99202	0.99218	0.99233	0.99248	0.99263
2.8	0.99346	0.99359	0.99372	0.99384	0.99396
2.9	0.99464	0.99475	0.99485	0.99495	0.99505

Table 7

Test de Kolmogorov-Smirnov

$$D_n = \sup\left|F_n^*(X) - F(X)\right|$$

Valeurs de D_n telles que $\alpha = P\left(D_n < d_n\right)$

n \ α	0,80	0,90	0,95	0,99
1	0,90000	0,95000	0,97500	0,99500
2	0,68377	0,77639	0,84189	0,92929
3	0,56481	0,63604	0,70760	0,82900
4	0,49265	0,56522	0,62394	0,73424
5	0,44698	0,50945	0,56328	0,66853
6	0,41037	0,46799	0,51926	0,61661
7	0,38148	0,43607	0,48342	0,57581
8	0,35831	0,40962	0,45427	0,54179
9	0,33910	0,38746	0,43001	0,51332
10	0,32260	0,36866	0,40925	0,48893
11	0,30829	0,35242	0,39122	0,46770
12	0,29577	0,33815	0,37543	0,44995
13	0,28470	0,32549	0,36143	0,43247
14	0,27481	0,31417	0,34890	0,41762
15	0,26588	0,30397	0,33760	0,40420
15	0,25778	0,29472	0,32733	0,39201
17	0,25039	0,28627	0,31796	0,38086
18	0,24360	0,27851	0,30936	0,37062
19	0,23735	0,27136	0,30143	0,36117
20	0,23156	0,26473	0,29408	0,35241
21	0,22617	0,25858	0,28724	0,34427
22	0,22115	0,25283	0,28087	0,33666
23	0,21645	0,24746	0,27490	0,32954
24	0,21205	0,24242	0,26931	0,32286
25	0,20790	0,23768	0,26404	0,31657
26	0,20399	0,23320	0,25907	0,31064
27	0,20030	0,22868	0,25438	0,30502
28	0,19680	0,22497	0,24993	0,29971
29	0,19348	0,22117	0,24571	0,29466
30	0,19032	0,21756	0,24170	0,28987
31	0,18732	0,21412	0,23788	0,28530
32	0,18445	0,21085	0,23424	0,28094
33	0,18171	0,20771	0,23076	0,27677
34	0,17909	0,20472	0,22743	0,27279
35	0,17659	0,20185	0,22425	0,26897
36	0,17418	0,19910	0,22119	0,26532
37	0,17188	0,19646	0,21826	0,26180
38	0,16966	0,19392	0,21544	0,25843
39	0,16753	0,19148	0,21273	0,25518
40	0,16547	0,18913	0,21012	0,25205

table 7(suite)

n \ α	0,80	0,90	0,95	0,99
41	0,16349	0,18687	0,20760	0,24904
42	0,16158	0,18408	0,20517	0,24613
43	0,15974	0,18257	0,20283	0,24332
44	0,15796	0,18053	0,20056	0,24060
45	0,15623	0,17856	0,19837	0,23798
46	0,15457	0,17665	0,19625	0,23544
47	0,15295	0,17481	0,19420	0,23208
48	0,15139	0,17302	0,19221	0,23059
49	0,14987	0,17128	0,19028	0,22828
50	0,14840	0,16959	0,18841	0,22604
51	0,14697	0,16796	0,18659	0,22386
52	0,14558	0,16637	0,18482	0,22174
53	0,14423	0,16483	0,18311	0,21968
54	0,14292	0,16332	0,18144	0,21768
55	0,14164	0,16186	0,17981	0,21574
56	0,14040	0,16044	0,17823	0,21384
57	0,13919	0,15906	0,17669	0,21199
58	0,13801	0,15771	0,17519	0,21019
59	0,13686	0,15639	0,17373	0,20844
60	0,13573	0,15511	0,17231	0,20673
61	0,13464	0,15385	0,17091	0,20506
62	0,13357	0,15263	0,16956	0,20343
63	0,13253	0,15144	0,16823	0,20184
64	0,13151	0,15027	0,16693	0,20029
65	0,13052	0,14913	0,16567	0,19877
66	0,12954	0,14802	0,16443	0,19729
67	0,12859	0,14693	0,16322	0,19584
68	0,12766	0,14587	0,16204	0,19442
69	0,12675	0,14483	0,16088	0,19303
70	0,12586	0,14381	0,15975	0,19167
71	0,12499	0,14281	0,15864	0,19034
72	0,12413	0,14183	0,15755	0,18903
73	0,12329	0,14087	0,15649	0,18776
74	0,12247	0,13993	0,15544	0,18650
75	0,12167	0,13901	0,15442	0,18528
76	0,12088	0,13811	0,15342	0,18408
77	0,12011	0,13723	0,15244	0,18290
78	0,11935	0,13636	0,15147	0,18174
79	0,11860	0,13551	0,15052	0,18060
80	0,11787	0,13467	0,14960	0,17949
81	0,11716	0,13385	0,14868	0,17840
82	0,11645	0,13305	0,14779	0,17732
83	0,11576	0,13226	0,14691	0,17627
84	0,11508	0,13148	0,14605	0,17523
85	0,11442	0,13072	0,14520	0,17421

table 7(suite 2)

n \ α	0,80	0,90	0,95	0,99
86	0,11376	0,12997	0,14437	0,17321
87	0,11311	0,12923	0,14355	0,17223
88	0,11248	0,12850	0,14274	0,17126
89	0,11186	0,12779	0,14195	0,17031
90	0,11125	0,12709	0,14117	0,16938
91	0,11064	0,12640	0,14040	0,16846
92	0,11005	0,12572	0,13965	0,16755
93	0,10947	0,12506	0,13891	0,16666
94	0,10889	0,12440	0,13818	0,16579
95	0,10883	0,12375	0,13746	0,16493
96	0,10777	0,12312	0,13675	0,16408
97	0,10722	0,12249	0,13606	0,16324
98	0,10668	0,12187	0,13537	0,16242
99	0,10615	0,12126	0,13469	16161
100	0,10563	0,12067	0,13403	0,16081
$n > 100$	$1.073\sqrt{n}$	$1.223\sqrt{n}$	$1.358\sqrt{n}$	$1.629\sqrt{n}$

BIBLIOGRAPHIE

1- AIVAZAN S. , ENUKOV I. , MECHLKINE L.
Eléments de modélisation et traitement primaire des données .
MIR. Moscou . 1986

2- ALLAIN J.Y. , MALLER F.
Statistiques
OPU. Alger . 1986

3- BAILLY P.
Exercices corrigés de statistique descriptive
OPU. Alger . 1993

3- BENYAKHLEF M.
Probabilités et statistiques mathématiques t.2
Editions Maghrébines . Casablanca . 1977

5- DAGNELIE P.
Théorie et méthodes statistiques t .1 et t .2
Presses agronomiques . Gembloux . 1985

6- FOURASTIE J. , SAHLER B.
Probabilités et statistiques
Dunod . Paris . 1981

7- JAFFARD P.
Initiation aux méthodes de la statistique et du calcul des des probabilités .
Masson . Paris . 1976

8- KHALDI K.
Probabilités . Rappels de cours . Exercices corrigés .
OPU . Alger . 1994

9- LABROUSSE C.
Statistique . Exercices corrigés avec rappels de cours .
t . 1 , t .3
Dunod . Paris . 1975

10- LEBART L. , MORINEAU A. , FENELON J.P.
Traitement des données statistiques .
Dunod . Paris . 1982

11- MOREAU M. , MATHIEU A.
Statistique appliquée à l'expérimentation .
Eyrolles . Paris . 1979

12- MOUCHOT C.
Exercices pédagogiques de statistique et économétrie .
Economia . Paris . 1989

13- MURRAY R. SPIEGEL
Théorie et application de la statistique .
Mac Graw . Hill . Paris . 1985

14- PUPION G.
Statistiques descriptives et décisionnelles appliquées à l'économétrie et à la gestion .
Presses universitaires de France . Paris . 1988

15- QUITTARD PINON F. , LIGNELET P.
Eléments de statistique . t. 2 .
OPU . Alger . 1983

16- ROSANOV Y.A.
Théorie des probabilités , processus aléatoires et statistique mathématique .
Naouka . Moscou . 1985

17- SALVATORE D.
Econométrie et statistique appliquées .
Mac Graw. Hill . Paris . 1986

18- SAPORTA G.
Théories et méthodes de la statistique .
Technip . Paris . 1978

Achevé d'imprimer sur les presses de
l'OFFICE DES PUBLICATIONS UNIVERSITAIRES
Imprimerie Régionale de Constantine