Exercice 2

On observe 100 fois le nombre d'arrivées (variable X) de clients à un bureau de poste pendant un intervalle de temps (10 minutes) et on obtient les valeurs suivantes :

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	4	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	6	6

a. Dresser le tableau statistique de la distribution de la variable X (effectifs cumulés, …).

b. Calculer les valeurs de tendance centrale de la distribution : la moyenne, le mode et les trois quartiles Q_1, Q_2 et Q_3.

c. Calculer les valeurs de la dispersion de la distribution : variance, l'écart type et l'intervalle interquartile.

d. Tracer le diagramme en bâtons et la boite à moustaches de cette distribution.

Correction de l'exercice 2

a. Tableau statistique

X	ni	fi	Fi	xi*fi	xi^2*fi
1	15	0.15	0.15	0.15	0.15
2	25	0.25	0.4	0.5	1
3	26	0.26	0.66	0.78	2.34
4	20	0.2	0.86	0.8	3.2
5	7	0.07	0.93	0.35	1.75
6	7	0.07	1	0.42	2.52
$\sum$	100	1		3	10.96

b. Les valeurs de tendance centrale

La moyenne : $\bar{X} = 3$

Le mode= 3

Indice de Q1 est n/4=25 ➔ Q1=2

Indice de Q2 est n/2=50 ➔ Q2=3

Indice de Q3 est 3n/4=75 ➔ Q3=4

c. Les valeurs de la dispersion de la distribution : variance, l'écart type et l'intervalle interquartile :

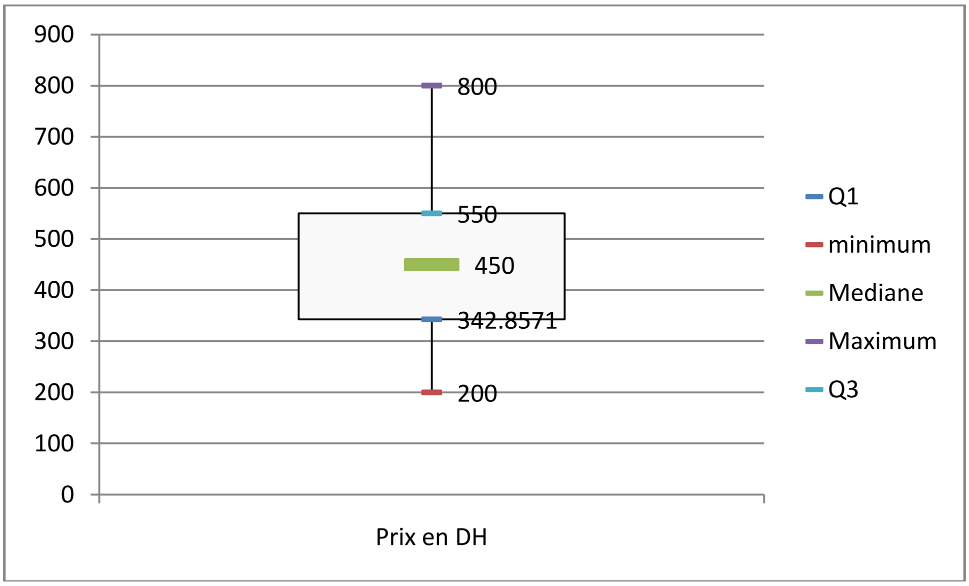
Var(X)= 10.96 – 3^2= 1.96

$\partial_x = \sqrt{var\ (X)} = 1.4$

IQ = Q3-Q1=4 – 2 = 2

Q1-1.5.IQ=2 - 1.5 . 2= -1

Q3+1.5 . IQ= 4+1.5 . 2=7



Exercice 3

On dispose des résultats d'une enquête concernant les loyers annuels des appartements dans un quartier de la ville.

Montant du loyer (x 1000)	Effectifs
[4; 6[	20
[6; 8[	40
[8; 10[	80
[10; 15[	30
[15; 20[	20
[20; 30[	10

a. Compléter le tableau statistique (valeurs centrales, effectifs cumulés, fréquence, fréquences cumulés)
b. Déterminez les valeurs de tendance centrale de la distribution : moyenne, mode et les quartiles.
c. Mesurez la dispersion de la distribution au moyen de : l'étendue, l'écart type et de l'intervalle interquartile.
d. Tracez l'histogramme et la boite à moustaches de cette distribution.

Correction de l'exercice 3

Montant x 1000	n_i	x_i	N_i	f_i	F_i	$f_i\ x_i$	d_i
[4; 6[	20	5	20	0.1	0.1	0.5	10
[6; 8[	40	7	60	0.2	0.3	1.4	20
[8; 10[	80	9	140	0.4	0.7	3.6	40
[10; 15[	30	12.5	170	0.15	0.85	1.875	6
[15; 20[	20	17.5	190	0.1	0.95	1.75	4
[20; 30[	10	25	200	0.05	1	1.25	1
$\sum$				1		10.375 x 1000	

$x_i = \frac{a_i + a_{i+1}}{2}$

$di = \frac{ni}{a_{i+1} - a_i}$ densité parce que on a pas la même amplitude.

Mode :

La classe modale = [8; 10[x 1000 (la classe qui a la plus grande densité)

Mode M= $a_i + \frac{\Delta i}{\Delta_i + \Delta_{i+1}} \ (a_{i+1} - a_i)$

Δ_i = 40-20

Δ_{i+1} = 40-6

$$M== 8000 + \frac{20}{34} \times (10000\text{-}8000)$$

$$M== 8000 + \frac{20}{34} \times 2000$$

$$M== 8000 + \frac{40000}{34}$$

M = 9176,470

$Q_1 = a_i + (a_{i+1} - a_i) \times \frac{0{,}25 - F_i}{F_{i+1} - F_i}$ (a_i =8000 ; $a_{i+1} = 10000$)

$Q_2 = a_i + (a_{i+1} - a_i) \times \frac{0{,}5 - F_i}{F_{i+1} - F_i}$ (a_i =8000 ; $a_{i+1} = 10000$)

$Q_3 = a_i + (a_{i+1} - a_i) \times \frac{0{,}75 - F_i}{F_{i+1} - F_i}$ (a_i =10000 ; $a_{i+1} = 15000$)

Q_1 = 7500

Q_2 = 9000

Q_3 = 11666

L'étendu W = (30-4) x 1000 = 26000

IQ = $Q_3 - Q_1$ = 11666 -7500 = 4166

Ecart type $\partial_x = \sqrt{var\ (X)}$

var(x) =$(\frac{1}{n} \sum n_i x_i^2) - \overline{x^2} = \sum f_i\, x_i^2 - \overline{x^2}$

$$= \frac{1}{200} \times 26002{,}5\ .10^6 - (10375)$$

$\partial_x = \sqrt{var\ (x)}$ = 8062,25

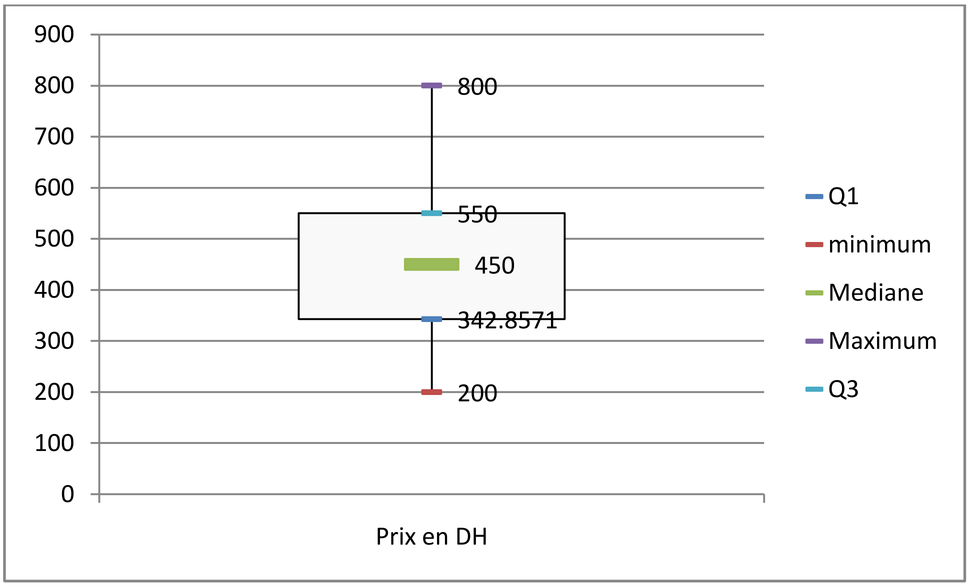
Q_1 – 1,5 IQ = 7500 – 1,5 x 4166 = 1251

Q_3 – 1,5 IQ = 11666 + 1,5 x 4166 = 17915

Q_1 = 7500

Q_2 = 9000

Q_3 = 11666



Exercice 4

Une société immobilière dispose de 600 appartements dont les surfaces sont données par le tableau suivant :

Surface (en mm^2)	[25; 50[	[50; 60[	[60; 80[	[80; 100[	[100; 120[	[120; 145[
fréquence	0,02	0,15	0,13	0,22	0,28	0,20

a. Compléter le tableau statistique suivant :

Classes	Centres x_i	Effectifs n_i	Densités	Effectifs cumulés N_i	Fréquences f_i	Fréquences cumulés F_i	fi * x_i	$f_i * x_i * x_i$
[25; 50[								
[50; 60[								
[60; 80[								
[80; 100[								
[100; 120[								
[120; 145[								
Total								

b. Calculer les indicateurs de position et ceux de dispersion et compléter le tableau suivant :

Moyenne =	Q_1 =
Classe modale =	Q_2 =
Mode =	Q_3 =
Variance =	$Q_3 - Q_1$ =
Ecart – type =	$Q_1 - 1,5\ (Q_3 - Q_1)$ =
Coefficient de variation =	$Q_3 - 1,5\ (Q_3 - Q_1)$ =

c. Tracer la boite à moustaches (boîte de Tukey) de cette série statistique,
d. Donner l'histogramme correspond à cette série statistique
e. Tracer la courbe cumulative des fréquences.

Correction de l'exercice 4

a. Compléter le tableau statistique suivant :

Classes	Centres x_i	Effectifs n_i	Densités	Effectifs cumulés N_i	Fréquences f_i	Fréquences cumulés F_i	fi * x_i	$f_i * x_i * x_i$
$[25; 50[$	37.5	12	0.48	12	0.02	0.02	0.75	28.125
$[50; 60[$	55	90	9	102	0.15	0.17	8.25	453.75
$[60; 80[$	70	78	3.9	180	0.13	0.3	9.1	637
$[80; 100[$	90	132	6.6	312	0.22	0.52	19.8	1782
$[100; 120[$	110	168	8.4	480	0.28	0.8	30.8	3388
$[120; 145[$	82.5	120	4.8	600	0.2	1	16.5	1361.25
Total		600			1		85.2	7650.125

b. Calculer les indicateurs de position et ceux de dispersion et compléter le tableau suivant :

Moyenne = 85.2	Q_1 =72.12598
Classe modale = $[50; 60[$	Q_2 =98.18182
Mode = 56.30178	Q_3 =116.4286
Variance = 391.085	$Q_3 - Q_1$ = 44.30262
Ecart – type = 19.77587	$Q_1 - 1{,}5\ (Q_3 - Q_1)$ = 5.67205
	$Q_3 - 1{,}5\ (Q_3 - Q_1)$ = 182.8825

$Q_1 = a_i + (a_{i+1} - a_i) \times \frac{0.25 - F_i}{F_{i+1} - F_i}$ $\quad (a_i = 60 ; a_{i+1} = 80)$

$Q1 = 60 + (80-60) \times \frac{0.25 - 0.173}{0.3 - 0.173} = 72.12598$

$Q_2 = a_i + (a_{i+1} - a_i) \times \frac{0.5 - F_i}{F_{i+1} - F_i}$ $\quad (a_i = 80 ; a_{i+1} = 100)$

$Q2 = 80 + (100-80) \times \frac{0.5 - 0.3}{0.52 - 0.3} = 98.18182$

$Q_3 = a_i + (a_{i+1} - a_i) \times \frac{0.75 - F_i}{F_{i+1} - F_i}$ $\quad (a_i = 100 ; a_{i+1} = 120)$

$Q3 = 100 + (120-100) \times \frac{0.75 - 0.52}{0.8 - 0.52} = 116.4286$

Mode : $M = a_i + \frac{\Delta i}{\Delta_i + \Delta_{i+1}} (a_{i+1} - a_i)$

$\Delta_i = 9 - 0.48 = 8.52$

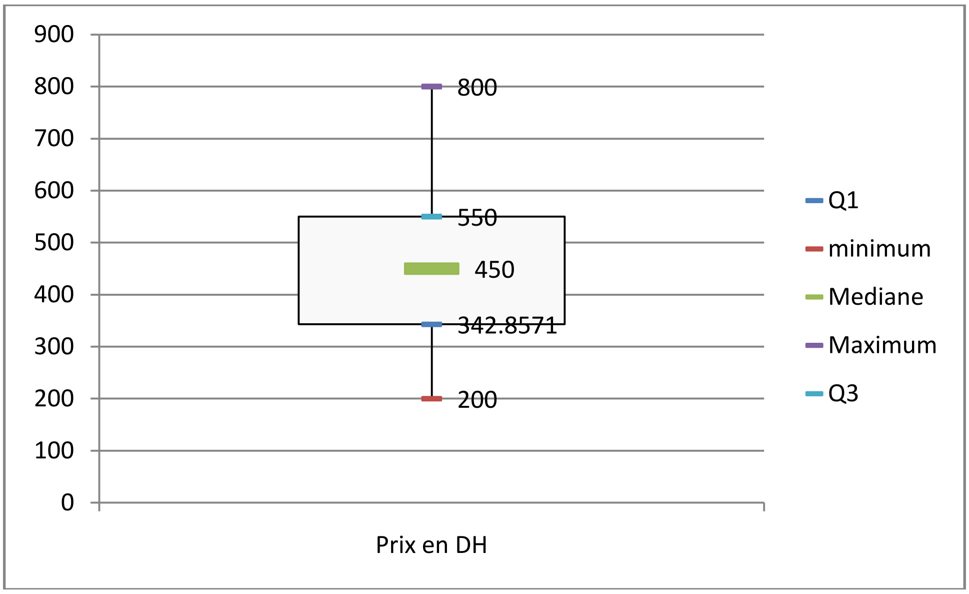
$\Delta_{i+1} = 9 - 3.9 = 5.1$

$M = 50 + \frac{8.52}{13.52} \times (60-50) = 56.30178$

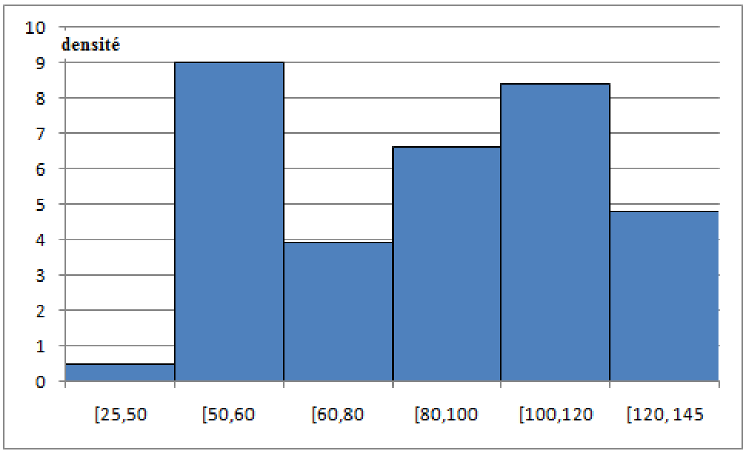
$Var(X) = 7650.125 - 85.2^2 = 391.085$

Ecart type $\partial_x = \sqrt{var\ (X)}$ = 19.77587

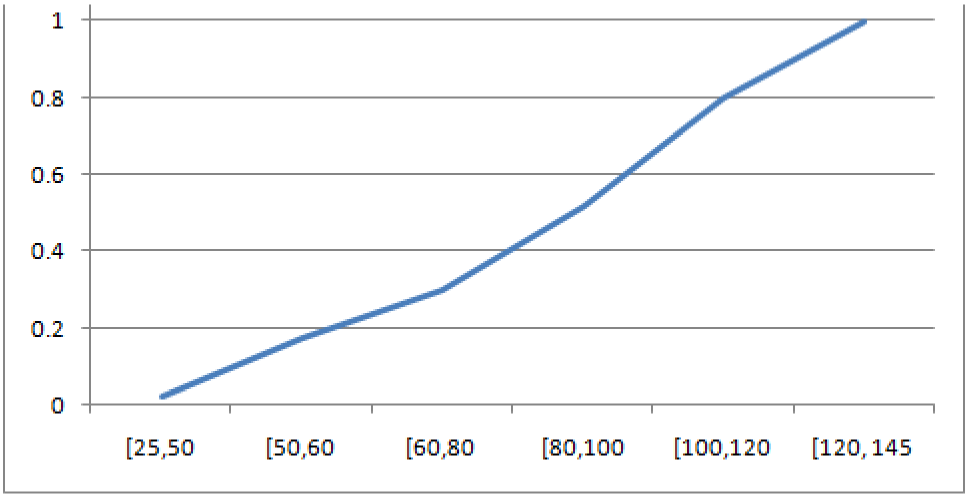
a.



d.



d.



II. Statistique descriptive bivariée

Exercice 1

On considère la série double suivante

x_i	2	5	6	10	12
y_i	83	70	70	54	49

1) Calculer la covariance,
2) Déterminer l'équation de la droite de régression Y = aX + b
3) Le coefficient de corrélation linéaire,
4) Le coefficient de détermination

Correction de l'exercice 1

1. $\text{Cov}(x, y) = \frac{1}{4} \sum x_i y_i - \bar{X}\ \bar{Y}$

$\text{Cov}(x, y) = 412.8 - 7 \times 65{,}2 = -43{,}6$

$\bar{Y} = \frac{1}{n} \sum y_i = 65{,}2$

$\bar{X} = \frac{1}{n} \sum x_i = 7$

$$\boxed{\text{Cov}(X, Y) = -43{,}6}$$

2. Y = aX + b

$$\begin{cases} a = \dfrac{cov\ (X,Y)}{var\ (X)} = -3{,}4 \quad (\text{var}(X) = (\frac{1}{n}\sum x_i^2) - \overline{X^2} = 61.8 - 49 = 12{,}8) \\ b = \bar{Y} - a\,\bar{X} = 65{,}2 + 3{,}4 \times 7 = 89 \end{cases}$$

$$\begin{cases} a = \dfrac{-43{,}6}{12{,}8} = -3{,}4 \\ b = 89 \end{cases}$$

$$\boxed{Y = -3{,}4\,X + 89}$$

3. Le coefficient de corrélation:

$r = \dfrac{cov\ (X,Y)}{\sigma X\ \sigma Y}$

$\text{Var}(Y) = \frac{1}{n}\sum y_i^2 - \overline{Y^2} = \frac{1}{5}\,20006 - 65{,}2$

$\text{Var}(Y) = 17754{,}96$

$\sigma Y = 133{,}24$

$$r = \frac{-43,6}{\sqrt{12,8\ x\ 133,24}} = -0,009$$

R= 0,008 proche de 0

Il n’ y a pas de corrélation linéaire entre X et Y

Exercice 2

Une expérience a été réalisée sur 250 personnes pour étudier la relation qui existe entre l’âge X et le temps de sommeil Y. le tableau suivant a été obtenu :

X \ Y	[5,7[	[7,9[	[9,11[	[11,15[
[1,3[	0	0	2	36
[3,11[	0	3	12	26
[11,19[	2	8	35	16
[19,31[	0	26	22	3
, [31,59[	22	15	6	0

1) Calculer les moyennes marginales et les écarts types marginaux de X et Y,
2) Déterminer la covariance et le coefficient de corrélation linéaire,
3) Déterminer la droite de régression de Y en fonction de X
4) Estimer le temps de sommeil d’une personne de 66 ans

Correction de l’exercice 2

1)

	y_i	6	8	10	13	
x_i	X \ Y	[5,7[	[7,9[	[9,11[	[11,15[	$\sum$
2	[1,3[	0	0	2	36	38
7	[3,11[	0	3	12	26	41
15	[11,19[	2	8	35	16	60
25	[19,31[	0	26	22	3	61
75	[31,59[	22	15	6	0	43
	$\sum$	52	77	81	81	243

Déterminer la covariance Cov (X,Y) et le coefficient linéaire

Cov (X ,Y) = $\frac{1}{n}\sum_{i=1}^{5}\sum_{y=1}^{4}(x_i - \bar{X})\ (y\ \ - \bar{Y})\ x_y$

Cov (X ,Y) = $(\frac{1}{n}\sum_{i=1}^{5}\sum_{y=1}^{4} x_{ij}\ \ x_i y_j)$ - $\bar{X}\,\bar{Y}$

$\bar{X} = \frac{1}{n}\sum_{i=1}^{5} x_i\ \ n_i$

= $\frac{1}{243}$ ((38 x 2) + (7 x 41) + (15 x 60) + (25 x 61) + (45 x 43))

$\bar{X}$ = 19,43

X_i	n_i	x_i
[1,3[	38	2
[3,11[	41	7
[11,19[	60	15
[19,31[	61	25
[31,59[	43	45
$\sum$	243	

y	n_i	y_i
[5,7[	33	6
[7,9[	52	8
[9,11[	77	10
[11,19[	81	13
$\sum$	243	

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{4}(\ y_i n_i)$

= $\frac{1}{243}$ ((33 . 6) + (52 . 8) + (77 . 10) + (81 . 13))

$\bar{Y}$ = 10,02

Cov (X ,Y) = = $(\frac{1}{n}\sum_{i=1}^{5}\sum_{j=1}^{4} x_i\ \ x_i y_1)$ - $\bar{X}\,\bar{Y}$

= $\frac{1}{243}$ ((2 x 10 x 2) + (36 x 13 x 2) + (3 x 8 x 7)

+ (12 x 10 x 7) + (26 x 13 x 7)
+ (26 x 13 x 7) + (2 x 6 x 15)
+ (8 x 8 x 15) + (35 x 10 x 15)
+ (16 x 8 x 25) + (22 x 10 x 25)
+ (3 x 13 x 25) + (22 x 6 x 45)
+ (15 x 8 x 45) + (6 x 10 x 45)
- (10,02 x 19,43)

Cov (X ,Y) = -24,3

Calculer le coefficient de corrélation linéaire

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

σx = $\sqrt{\text{var(x)}}$ σy = $\sqrt{\text{var(y)}}$

Var (x) = $\frac{1}{n}(\sum_{i=1}^{5} x_i\ \ x_i^2) - \overline{X^2}$ = 200,5

Var (y) = $\frac{1}{n}(\sum_{j=1}^{4} x_i\ \ y_i^2) - \bar{Y}^2$= 4,05

r = $\frac{cov\ (x,y)}{\sigma x\ \ \sigma y}$ = - 0,85

R = r^2 = 0,72

Existe une corrélation linéaire forte entre X et Y

- Déterminer la droite de la corrélation linéaire

  Y = aX +b

a = $\frac{Cov\ (X,Y)}{Var\ \ (X)}$ b= $\overline{Y}$ - a $\overline{X}$

= 10,2 + 0,12 . 19,43

a = $\frac{24,3}{200,5}$ = -0,12

b= 12,35

Y = -0,12X + 12,35

- Estimer le temps de sommeil d'une personne de 66 ans

Y= 0,12 x66 + 15,35

Y = 4,43 heures

Exercice 3

Le tableau ci-dessous donne la répartition de 200 naissances en fonction de la parité de la mère et du poids du nouveau-né.

	Primipares	multipares
poids inférieur à 3 kg	26	20
entre 3 et 4 kg	61	63
Supérieur à 4 kg	8	22

Existe les deux caractères, parité de la mère et poids du nouveau-né, sont-ils statistiquement reliés ?

Correction de l'exercice 3

X \ Y	Primipare	Multipare	$\sum$
Poids $< 3\ kg$	26	20	46
[3,4[Kg	61	63	124
$>$ 4 kg	8	22	30
$\sum$	95	105	200

Hypothèse H_0
Supposent que x et y sont indépendants
Effectifs théorique

X \ Y	Primipare	Multipare
< 3 Kg	21,85	24,15
$3Kg < P < 4$Kg	58,9	65,1
> 4 Kg	14,25	15,75

Colonne 1 $\begin{cases} n^*_{11} = \frac{95 \, x \, 46}{200} \\ n^*_{21} = \frac{95 \, x \, 124}{200} \\ n^*_{31} = \frac{95x \, 30}{200} \end{cases}$

Colonne 2 $\begin{cases} n^*_{12} = \frac{46 \, x \, 105}{200} \\ n^*_{22} = \frac{105 \, x \, 124}{200} \\ n^*_{32} = \frac{30 \, x \, 105}{200} \end{cases}$

Khi-2 = $X^2 = \sum_{i,j} \frac{(n^*_{ij} - n_{ij})^2}{n^*_{ij}}$

$= \frac{1}{21,85}(26 - 21,85)^2 + \frac{1}{58,9}(61 - 58,9)^2 + \frac{1}{14,25}$

$(8-14,25)^2 + \frac{1}{24,15}(20-24,15)^2 + \frac{1}{65,1}(65+1-63)^2 + \frac{1}{15,75}(15,75-22)^2 = 6,88$

Si $X^2 = 0$ alors X ,Y sont indépendant

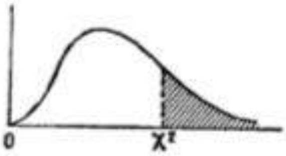
- On va choisir le seuil de signification 5% (0,05)
- Le degré de liberté = (nombre ligne - 1) (nombre colonne -1)=(3-1) (2-1) = 2

$X^2_{\text{Critique}} = 5,99$ (voir table de Khi-2)

Si $X^2 > X^2_{\text{Critique}}$ alors donc il existe un lien entre X et Y
Dans cet exercice existe une corrélation entre X et Y.

Table de χ^2 (*).

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



d.d.l. \ α	0,90	0,50	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,0158	0,455	1,074	1,642	2,706	3,841	5,412	6,635	10,827
2	0,211	1,386	2,408	3,219	4,605	5,991	7,824	9,210	13,815
3	0,584	2,366	3,665	4,642	6,251	7,815	9,837	11,345	16,266
4	1,064	3,357	4,878	5,989	7,779	9,488	11,668	13,277	18,467
5	1,610	4,351	6,064	7,289	9,236	11,070	13,388	15,086	20,515
6	2,204	5,348	7,231	8,558	10,645	12,592	15,033	16,812	22,457
7	2,833	6,346	8,383	9,803	12,017	14,067	16,622	18,475	24,322
8	3,490	7,344	9,524	11,030	13,362	15,507	18,168	20,090	26,125
9	4,168	8,343	10,656	12,242	14,684	16,919	19,679	21,666	27,877
10	4,865	9,342	11,781	13,442	15,987	18,307	21,161	23,209	29,588
11	5,578	10,341	12,899	14,631	17,275	19,675	22,618	24,725	31,264
12	6,304	11,340	14,011	15,812	18,549	21,026	24,054	26,217	32,909
13	7,042	12,340	15,119	16,985	19,812	22,362	25,472	27,688	34,528
14	7,790	13,339	16,222	18,151	21,064	23,685	26,873	29,141	36,123
15	8,547	14,339	17,322	19,311	22,307	24,996	28,259	30,578	37,697
16	9,312	15,338	18,418	20,465	23,542	26,296	29,633	32,000	39,252
17	10,085	16,338	19,511	21,615	24,769	27,587	30,995	33,409	40,790
18	10,865	17,338	20,601	22,760	25,989	28,869	32,346	34,805	42,312
19	11,651	18,338	21,689	23,900	27,204	30,144	33,687	36,191	43,820
20	12,443	19,337	22,775	25,038	28,412	31,410	35,020	37,566	45,315
21	13,240	20,337	23,858	26,171	29,615	32,671	36,343	38,932	46,797
22	14,041	21,337	24,939	27,301	30,813	33,924	37,659	40,289	48,268
23	14,848	22,337	26,018	28,429	32,007	35,172	38,968	41,638	49,728
24	15,659	23,337	27,096	29,553	33,196	36,415	40,270	42,980	51,179
25	16,473	24,337	28,172	30,675	34,382	37,652	41,566	44,314	52,620
26	17,292	25,336	29,246	31,795	35,563	38,885	42,856	45,642	54,052
27	18,114	26,336	30,319	32,912	36,741	40,113	44,140	46,963	55,476