Université Badji Mokhtar Annaba

02/06/2013

1ère année MI

Durée 1h30

Examen de Statistiques

Exercice 1 (6pts) : On demande à un échantillon de 30 consommateurs de déterminer le prix auquel ils achèteraient le nouveau produit qui leur est présenté. Les résultats sont les suivants:

Prix (DA)	100	120	130	140	160	180
Fréquences	0.10	0.20	0.30	0.23	0.13	0.03

1- Déterminer les effectifs et l'écart interquartile.

2- Donner le box-plot correspondant.

3- Déterminer la fonction de répartition de cette variable et tracer son graphe.

Exercice 2 (4pts): Ce tableau donne la durée des courses des clients d'un grand magasin:

Classes(minutes)	[15.5-20.5[	[20.5-25.5[	[25.5-30.5[	[30.5-35.5[	[35.5-40.5[
x_i	18	23	28	33	38
Fréquences	0.08	0.24	0.2	0.4	0.08

1-Définir la population, le caractère et préciser sa nature.

2-Tracer le polygone des fréquences.

3-Calculer le mode et les quartiles de cette distribution.

Exercice 3 (10pts) : La répartition des distances parcourues par 200 véhicules après un coup de frein selon leur vitesses est donnée dans le tableau suivant :

Y distances (m) / X Vitesse(Km/h)	[20,30[	[30,40[	[40,60[	[60,80[	[80,100[	Total
[70,80[	3	25	13	6	0	47
[80,90[	0	5	$\boxed{29}$	12	2	48
[90,100[	0	0	24	21	10	$\boxed{55}$
[100,110[	0	0	12	12	16	40
[110,120[	0	0	0	5	5	10
Total	3	30	78	56	33	200

1- Que représentent les nombres encadrés?

2-Déterminer les distributions et les moyennes marginales puis, déterminer la distribution conditionnelle de Y sachant X=75 et calculer sa moyenne et son écart type.

3- Construire le nuage de points. Est-il possible d'envisager une liaison linéaire entre X et Y? Justifier cette réponse à l'aide d'un calcul.

4-Déterminer la droite de régression de Y en X et en déduire la distance nécessaire à l'arrêt d'un véhicule lancé à 140 Km/h.

5- Tracer la courbe de régression de Y en X.

$\sum_{i=1}^{5} n_{i.}\,(x_i-\bar{X})^2 = 28238,\quad \sum_{i=1}^{5} n_{.j}\,(y_j-\bar{Y})^2 = 65488.875$

$\sum\sum n_{ij}(x_i-\bar{X})(y_j-\bar{Y}) = 26801.5$

Bon courage

Corrigé de l'examen de Statistiques

Ex 1 :

l'effectif : $n_i = f_i \times N$ où $N = 30$ 0,25

x_i	100	120	130	140	160	180	
f_i	0,10	0,20	0,30	0,23	0,13	0,03	
n_i	3	6	9	$6,9 \simeq 7$	$3,9 \simeq 4$	$0,9 = 1$	0,5
$f_{ic}^{\uparrow}$	0,10	$0,10+0,20 = 0,3$	0,6	0,83	0,96	$0,99 \simeq 1$	0,25

L'écart interquartile : $IQR = Q_3 - Q_1$ 0,25

$N = 30 = 2 \times 15 = 2k$ (pair) 0,25

$k = 15 = 2 \times 7 + 1 = 2p+1$ (impair) 0,25

$\Rightarrow Q_1 = x_{p+1} = x_{7+1} = x_8 = 120$ DA

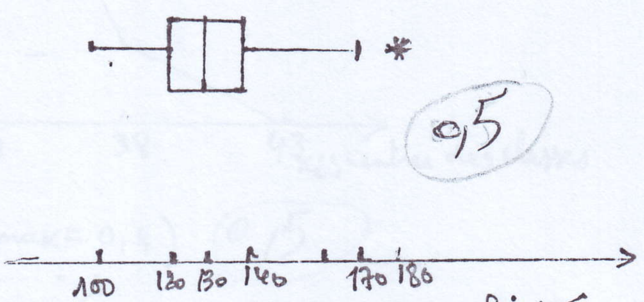
$Q_3 = x_{k+p+1} = x_{15+8} = x_{23} = 140$ DA

$\Rightarrow IQR = 140 - 120 = 20$ DA 0,25

La médiane :

$N = 2 \times 15$ (pair) $\Rightarrow Me = \dfrac{x_{15} + x_{16}}{2}$

0,5 $= \dfrac{130 + 130}{2} = 130$ DA



0,5

Boxplot du prix souhaité par les consommateurs

Les valeurs pivots :

0,25 $P_g = Q_1 - 1,5\, IQR = 120 - 1,5 \times 20 = 120 - 30 = 90$

0,25 $P_d = Q_3 + 1,5\, IQR = 140 + 30 = 170$

Les valeurs adjacentes :

x_g : la + petite valeur $\geq P_g \Rightarrow x_g = 100$ DA

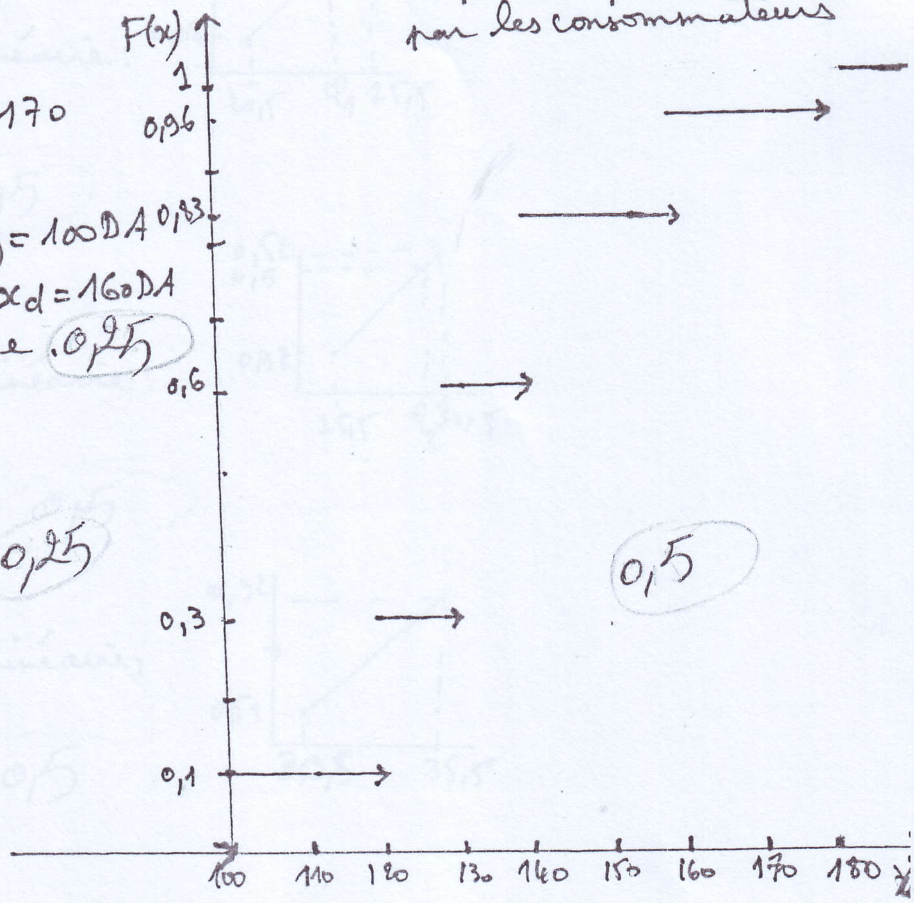
x_d : la + grande " $\leq P_d \Rightarrow x_d = 160$ DA

180 est une valeur extérieure 0,25

La fct de répartition :

$$F(x) = \begin{cases} 0 & \text{si } x < 100 \\ 0,10 & \text{si } 100 \leq x < 120 \\ 0,30 & \text{si } 120 \leq x < 130 \\ 0,60 & \text{si } 130 \leq x < 140 \\ 0,83 & \text{si } 140 \leq x < 160 \\ 0,96 & \text{si } 160 \leq x < 180 \\ 1 & \text{si } x \geq 180 \end{cases}$$

0,25



0,5

(1)

Classes	[15,5 – 20,5[	[20,5 – 25,5[	[25,5 – 30,5[	[30,5 – 35,5[	[35,5 – 40,5[
x_i	18	23	28	33	38
f_i	0,08	0,24	0,2	0,4	0,08
$f_{ic}\uparrow$	0,08	0,32	0,52	0,92	1

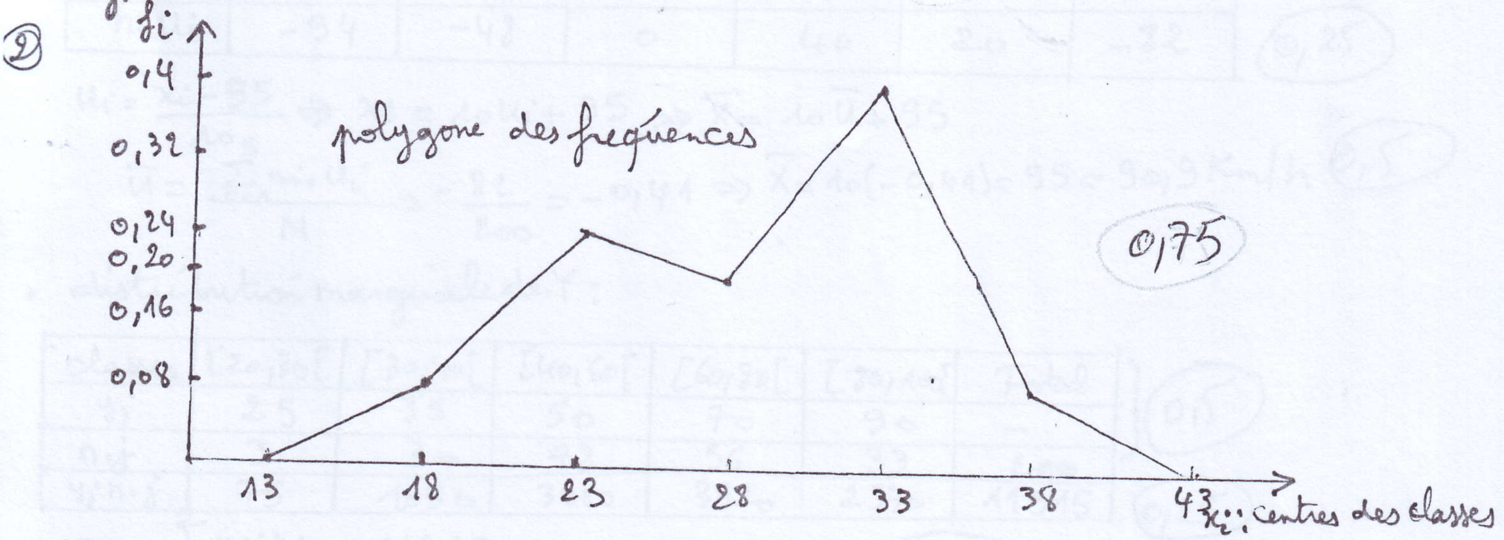
0,25

1) pop : les clients d'un grand magasin. 0,25

caractère : durée des courses des clients 0,25

type du caractère : quantitatif continu 0,5

2)

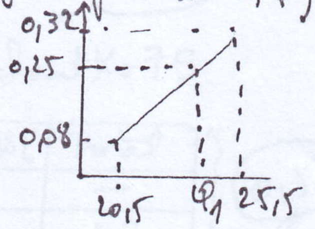


0,75

3) Mode : $M_o = 33$ mn (centre de la classe ayant $f_i max = 0,4$) 0,5

Quartiles :

$Q_1 = ?$

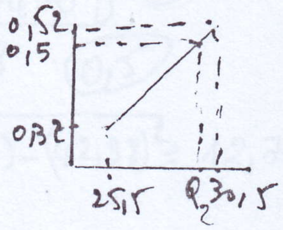


Calculons Q_1 par interpolation linéaire :

$$\frac{Q_1 - 20,5}{25,5 - 20,5} = \frac{0,25 - 0,08}{0,32 - 0,08} = \frac{0,17}{0,24}$$

$$Q_1 = 20,5 + \frac{5 \times 0,17}{0,24} = 24,04 \text{ mn}$$ 0,5

$Q_2 = ?$

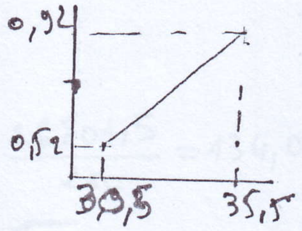


Calculons Q_2 par interpolation linéaire :

$$\frac{Q_2 - 25,5}{30,5 - 25,5} = \frac{0,5 - 0,32}{0,52 - 0,32} = \frac{0,18}{0,2}$$

$$Q_2 = 25,5 + \frac{5 \times 18}{2} = 30 \text{ mn}$$ 0,5

$Q_3 = ?$



Calculons Q_1 par interpolation linéaire,

$$\frac{Q_3 - 30,5}{35,5 - 30,5} = \frac{0,75 - 0,52}{0,92 - 0,52} = \frac{0,23}{0,4}$$ 0,5

$$Q_3 = 30,5 + 5 \times \frac{0,23}{0,4} \simeq 33,38 \text{ min}$$

EX03:

1- $n_{23}=29$. On a 29 véhicules ayant une vitesse entre 80 et 90 Km/h et nécessitant une distance entre 40 et 60 m pour s'arrêter après un coup de frein. (0,5)

$n_{3.}=55$. On a 55 véhicules ayant une vitesse entre 90 et 100 Km/h (0,5)

2. • distribution marginale de X:

classes	[70, 80[	[80,90[	[90,100[	[100,110[	[110,120[	Total
x_i	$\frac{70+80}{2}=75$	85	95	105	115	—
$n_{i.}$	47	48	55	40	10	200
u_i	-2	-1	0	1	2	
$n_{i.}u_i$	-94	-48	0	40	20	-82

(0,5) (0,25)

$u_i=\frac{x_i-95}{10} \Rightarrow x_i = 10u_i+95 \Rightarrow \overline{X}=10\overline{u}+95$

$\overline{u}=\frac{\sum_{i=1}^{5} n_{i.}u_i}{N} = -\frac{82}{200}=-0,41 \Rightarrow \overline{X}=10(-0,41)+95=90,9$ Km/h (0,5)

• distribution marginale de Y:

classes	[20,30[	[30,40[	[40,60[	[60,80[	[80,100[	Total
y_j	25	35	50	70	90	—
$n_{.j}$	3	30	78	56	33	200
$y_j n_{.j}$	75	1050	3900	3920	2970	11915

(0,5) (0,25)

$\overline{Y}=\frac{\sum_{j=1}^{5} n_{.j}y_j}{N}=\frac{11915}{200}=59,575$ m. (0,5)

• distribution conditionnelle de Y sachant X=75

classes	[20,30[	[30,40[	[40,60[	[60,80[	Total
y_j	25	35	50	70	—
n_{1j}	3	25	13	6	47.
$y_j n_{1j}$	75	875	650	420	2020
$y_j^2 n_{1j}$	1875	30625	32500	29400	94400

(0,5) (1)

$\overline{Y_1}=\overline{Y}/_{X=75}=\frac{1}{n_{1.}}\sum_{j=1}^{4} n_{1j}y_j=\frac{2020}{47}=42,98$ (0,5)

$\sigma_1=\sqrt{\frac{1}{n_{1.}}\sum_{j=1}^{4} n_{1j}y_j^2-\overline{Y_1}^2}=\sqrt{\frac{1}{47}(94400)-(42,98)^2}=12,70$ (0,5)

3. Le calcul qui justifierait un ajustement linéaire serait celui du coefficient de corrélation linéaire: (0,25)

$r(X,Y)=\frac{cov(X,Y)}{\sigma_X\,\sigma_Y}=0,62.$ (0,5)

$cov(X,Y)=\frac{1}{N}\sum\sum n_{ij}(x_i-\overline{X})(y_j-\overline{Y})=\frac{26801,5}{200}=134,0075$ (0,5)

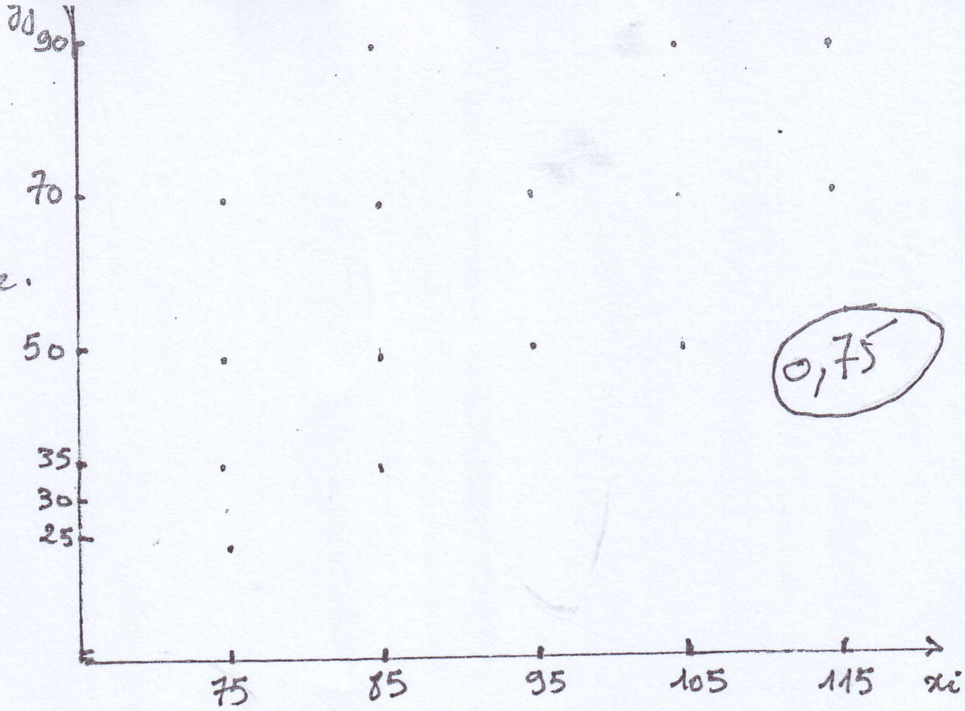
(0,25)

$\sigma_X=\sqrt{\frac{1}{N}\sum_{i=1}^{5} n_{i.}(x_i-\overline{X})^2}=\sqrt{\frac{28238}{200}}=\sqrt{141,19}=11,8823$ (0,25)

$\sigma_Y=\sqrt{\frac{1}{N}\sum_{j=1}^{5} n_{.j}(y_j-\overline{Y})^2}=\sqrt{\frac{65488,875}{200}}=18,0954$ (0,25)

Comme $0,6 \leq r < 0,8$ est acceptable donc une liaison linéaire entre X et Y est envisagable.

0,25



0,75

4. la droite de régression de Y en X s'écrit $y = ax + b$

avec $a = \dfrac{cov(X,Y)}{V_X}$, $b = \overline{Y} - a\overline{X}$

$a = \dfrac{134,0075}{141,19} \simeq 0,95$ 0,25 ; $b = 59,575 - (0,95)(90,9) = -26,78$ 0,25

$\Rightarrow$ l'eq de la droite de régression s'écrit $y = 0,95x - 26,78$ 0,25

Si $x = 140$ Km/h alors $y = 0,95(140) - 26,78 = 106,22$ m

Alors la distance nécessaire à l'arrêt d'un véhicule lancé à 140 Km/h est 106,22 m

0,25