Université Abdelmalek Essaâdi
Faculté Polydisciplinaire de Tétouan
Département de Statistique et Informatique

Année Universitaire 2014/2015
Filière : Sciences Éco. & Gestion
Semestre : Premier (S1)
Session d'automne

Contrôle final de Statistique Descriptive
(Durée : 1 heure 30 min.)

Exercice 1 : (10 points)
Dans un magasin de pièces détachées, sur un lot de 100 pièces vendues en une année ; les prix s'échelonnent entre 200 DH et 800 DH selon la répartition suivante :

Prix en DH	[200 ; 300[	[300; 450[	[450 ; 550[	[550 ; 600[	[600 ; 800[
Nombre de pièces vendues	15	35	25	10	15

1. Calculer les effectifs cumulés croissants et les fréquences cumulées croissantes de cette série statistique.
2. Déterminer le mode *Mo* de cette série, graphiquement et par le calcul.
3. Calculer la médiane *Mé* de cette série statistique en explicitant vos calculs. Donner son interprétation.
4. Déterminer les quartiles Q_1 et Q_3. Représenter le diagramme de Box & Wiskers.
5. Calculer l'écart interquartile relatif. Conclure.

Exercice 2 : (10 points)
Cinq personnes soufrant d'obésité suivent un régime d'amincissement.
Le tableau suivant donne le nombre de Kgs perdus par chacune d'elle pendant la période de cure suivie.

N° de l'individu	1	2	3	4	5
Durée X (en mois)	3	1	2	4	5
Nombre Y de Kgs perdus	6	4	5	9	11

1. Calculer la moyenne arithmétique de la variable X et celle de la variable Y.
2. Calculer la variance de la variable statistique X et celle de la variable Y.
3. Calculer la covariance des variables statistiques X et Y.
4. Donner la droite d'ajustement linéaire de Y en fonction de X, par les formules de la méthode des moindres carrées.
5. Calculer le coefficient de corrélation linéaire. Conclure.

Bonne Chance !

Corrections du Contrôle de Statistique Descriptive

Exercice 1 :

$[e_{i-1}, e_i[$	n_i	n_icc	f_i	f_icc	a_i	$h_i = \frac{n_i}{a_i}$
[200 ; 300[	15	15	0,15	0,15	100	0,15
[300 ; 450[	35	50	0,35	0,5	150	0,23
[450 ; 550[	25	75	0,25	0,75	100	0,25
[550 ; 600[	10	85	0,1	0,85	50	0,2
[600 ; 800[	15	100	0,15	1	200	0,075

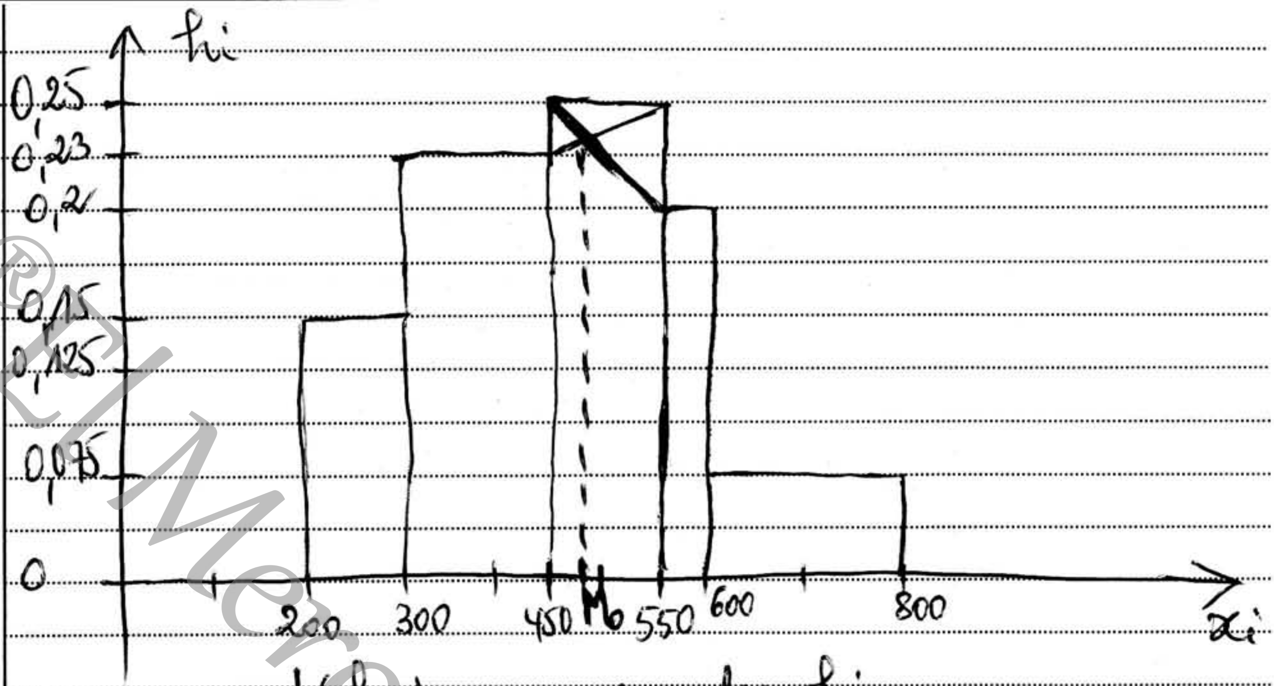
N = 100

1°) voir tableau : $n_icc = \sum_{j \leq i} n_j$ et $f_icc = \sum_{j \leq i} f_j$

2°) * le mode Mo par la méthode graphique :
Comme les amplitudes sont différentes (voir tableau, colonne des a_i), on représente l'histogramme des $h_i = \frac{n_i}{a_i}$ les fréquences moyennes par unité d'amplitude

– voir la représentation dans la page suivante –

On utilise la méthode des diagonales dans le rectangle du plus grande hauteur $h_i = \frac{n_i}{a_i}$



l'histogramme des hi

* le mode Mo par le calcul :

Comme les amplitudes sont différentes, on définie la classe modale comme étant celle du plus grande hauteur dans l'histogramme des $hi = \frac{ni}{ai}$

hi la plus grande est 0,25, elle correspond à la classe [450 ; 550[ou on applique l'une des formules suivante :

$$Mo = e_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} a_i \qquad (I)$$

ou

$$Mo = e_{i-1} + \frac{h_i - h_{i-1}}{(h_i - h_{i+1}) + (h_i - h_{i-1})} a_i \qquad (II)$$

On a : $e_{i-1} = 450$ $\quad a_i = 100$

$h_i = 0,25$; $h_{i-1} = 0,23$; $h_{i+1} = 0,2$

par (I) on aura $M_0 = 450 + \frac{0,2}{0,23 + 0,2} \times 100$

$M_0 = 450 + 46,51$

$\Rightarrow \boxed{M_0 = 496,51}$

par (II) on aura $M_0 = 450 + \frac{0,25 - 0,23}{(0,25 - 0,2) + (0,25 - 0,23)} \times 100$

$M_0 = 450 + 28,57$

$\Rightarrow \boxed{M_0 = 478,57}$

3°) la médiane par le calcul :

$\frac{N}{2} = \frac{100}{2} = 50$

Cette valeur existe exactement parmi les nicc

$\Rightarrow$ la classe médiane est $[300 ; 450]$ et on prend $\boxed{Mé = e_2 = 450}$

Interprétation :

Il y a 50 % des pièces détachées qui ont un prix inférieur à 450 DH et 50 % autres qui ont un prix supérieur à 450 DH.

4°) Q_1 ?

$\frac{N}{4} = \frac{100}{4} = 25$

Ce 25 n'existe pas exactement parmi les nicc mais 50 est la 1ère valeur qui la dépasse alors, on applique la formule :

$$Q_1 = e_{i-1} + \frac{\frac{N}{4} - n_{i-1}cc}{n_i} a_i$$

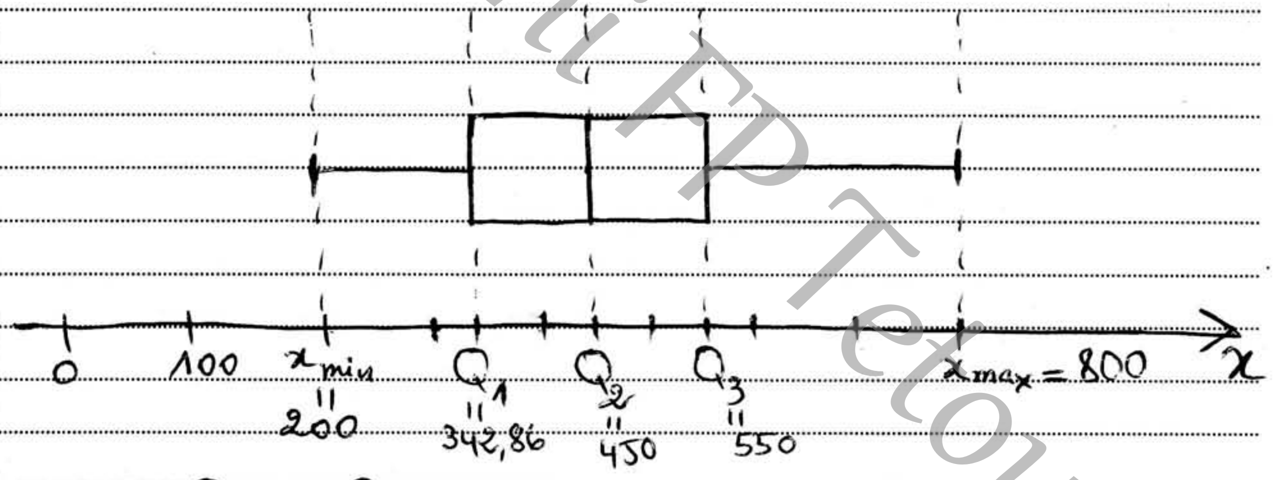
$$Q_1 = 300 + \frac{25 - 15}{35} \times 150$$

$$Q_1 = 300 + 42,86 = \boxed{342,86}$$

$$Q_2 = Mé = \boxed{450}$$

Q_3 ?

$\frac{N}{4} \times 3 = 25 \times 3 = 75$ existe exactement parmi les nicc, alors on prend $\boxed{Q_3 = e_3 = 550}$



5°) $\frac{Q_3 - Q_1}{Q_2} = \frac{550 - 342,86}{450} = 0,46 < 1,5$

<u>Conclusion</u> : Comme $\frac{Q_3 - Q_1}{Q_2} < 1,5$, alors la série est homogène.

Exercice 2 :

x_i	y_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$	$y_i - \bar{Y}$	$(y_i - \bar{Y})^2$	$(x_i - \bar{X})(y_i - \bar{Y})$
3	6	0	0	-1	1	0
1	4	-2	4	-3	9	6
2	5	-1	1	-2	4	2
4	9	1	1	2	4	2
5	11	2	4	4	16	8
15	35		10		34	18

1°) $\boxed{\bar{X} = \frac{15}{5} = 3}$ $\boxed{\bar{Y} = \frac{35}{5} = 7}$

2°) $\boxed{Var(X) = \frac{1}{N} \sum_i (x_i - \bar{X})^2 = \frac{10}{5} = 2}$

$\sigma(X) = \sqrt{2}$

$\boxed{Var(Y) = \frac{1}{N} \sum_i (y_i - \bar{Y})^2 = \frac{34}{5} = 6{,}8}$

$\sigma(Y) = \sqrt{6{,}8}$

3°) $\boxed{Cov(X, Y) = \frac{1}{N} \sum_i (x_i - \bar{X})(y_i - \bar{Y})}$

$\boxed{= \frac{18}{5} = 3{,}6}$

4°) Droite de moindres carrées : $Y = aX + b$

avec $\boxed{a = \frac{Cov(X,Y)}{Var(X)} = \frac{3{,}6}{2} = 1{,}8}$

$$b = \bar{y} - a\bar{x} = 7 - 1{,}8 \times 3$$
$$= 7 - 5{,}4 = \boxed{1{,}6}$$

D'où $\boxed{y = 1{,}8\,x + 1{,}6}$

5°) $$r = \frac{Cov(x,y)}{\sqrt{Var(x)\,Var(y)}} = \frac{3{,}6}{\sqrt{2 \times 6{,}8}}$$

$$= \frac{3{,}6}{\sqrt{13{,}6}} = \frac{3{,}6}{3{,}688} = \boxed{0{,}98 \simeq 1}$$

Conclusion : Il y a une corrélation positive forte entre la durée X de cure et le nbre Y de Kgs perdus par le patient.