Université Abdelmalek Essaâdi
Faculté poly disciplinaire
Tétouan

Année universitaire 2004-2005
Module : Méthodes Quantitatives I
Enseignant : M. El Merouani

Contrôle final de Statistique I

Durée 2 heures

Exercice 1 : (12 points)

Les salaires annuels (en 1000 DH) de 100 employés d'une entreprise sont distribués de la façon suivante :

Salaires compris entre	Nombre des employés
25 et 32	20
32 et 40	30
40 et 45	22
45 et 55	28

1) Déterminer le pourcentage des employés qui gagnent moins que 42350 DH par an.
2) Calculer les quartiles Q_1, Q_2 et Q_3 de cette distribution et représenter le diagramme de la boîte de Tukey. Interpréter cette représentation graphique.
3) Déterminer le salaire le plus fréquent.
4) Donner l'interprétation et la valeur de la médiale. Calculer la différence $\Delta M = Ml - Mé$. Comparer-la à l'étendue et conclure.
5) Calculer l'indice de concentration de GINI.
6) Tracer la courbe de concentration de Lorenz et interpréter sa forme selon le résultat trouvé dans la question précédente.

Exercice 2 : (8 points)

Dans un pays, on a étudié les exploitations agricoles de la région du nord et de la région du sud en fonction de la superficie. Les données se présentent comme suit :

Superficies en hectare	Nombre des exploitations dans la région du nord	Nombre des exploitations dans la région du sud
0 – 10	3	5
10 – 30	15	19
30 –50	20	21
50 - 100	12	5

1) Calculer la superficie moyenne de la région du nord, de la région du sud et celle de tout le pays.
2) Calculer la variance de la région du nord, de la région du sud et celle de tout le pays.
3) Calculer la variance intra-région et la variance inter-région. Interpréter et conclure.
4) Comparer la dispersion des exploitations dans la région du nord et la région du sud, en utilisant le coefficient de variation.

C.F. de Statistique I

Exercice 1 :

$[e_{i-1}, e_i[$	n_i	c_i	$n_i c_i$	a_i	$h_i = \frac{n_i}{a_i}$	f_i	$f_i c\uparrow$	$n_i c\uparrow$	$(n_i c_i) c\uparrow$
[25, 32[	20	28,5	570	7	2,86	0,2	0,2	20	570
[32, 40[	30	36	1080	8	3,75	0,3	0,5	50	1650
[40, 45[	22	42,5	935	5	4,4	0,22	0,75	72	2585
[45, 55[	28	50	1400	10	2,8	0,28	1	100	3985
	N=100		3985						

1°/

0,5

$$\left.\begin{array}{l} 40\,000 \longrightarrow 50\% \\ 42350 \longrightarrow x \\ 45\,000 \longrightarrow 75\% \end{array}\right\} \Rightarrow \frac{45\,000 - 40\,000}{42350 - 40\,000} = \frac{75-50}{x-50}.$$

$$\Rightarrow x - 50 = \frac{75-50}{45000-40\,000} \times (42\,350 - 40\,000) = \frac{25}{5000} \cdot 2350$$

0,5

$$= 11,75 \Rightarrow \boxed{x = 61,75\ \%}$$

2°/ les quartiles :

Q_1 ?

$\frac{N}{4} = 25 \Rightarrow$ On cherche cette valeur parmi les $n_i c\uparrow$

$\Rightarrow$ 50 est la 1ère valeur qui le dépasse $\Rightarrow [32, 40[$

$\Rightarrow$ Formule $\Rightarrow \quad Q_1 = e_{i-1} + \dfrac{\frac{N}{4} - n_{i-1}c\uparrow}{n_i} a_i$

0,5 $\Rightarrow Q_1 = 32 + \dfrac{25 - 20}{30} \cdot 8 = \boxed{33{,}33}$

* $\frac{N}{4} \times 2 = 50 \Rightarrow$ cette valeur se trouve exactement parmi les $n_i c\uparrow \Rightarrow$ alors on prend $Q_2 = e_2$

0,5 $\Rightarrow \boxed{Q_2 = 40}$

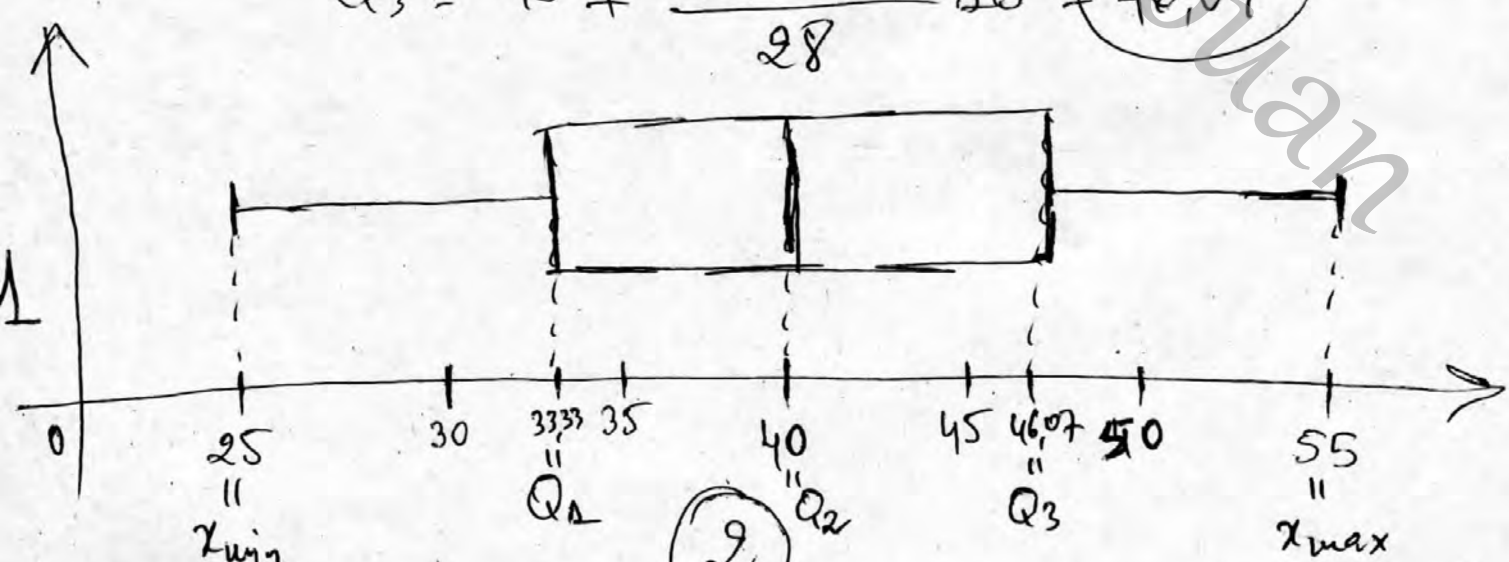
* $\frac{N}{4} \times 3 = 75 \Rightarrow$ cette valeur n'existe pas exactement parmi les $n_i c\uparrow$

$\Rightarrow$ 100 est la 1ère valeur qui la dépasse

$\Rightarrow [45, 55[$; la formule

$Q_3 = e_{i-1} + \dfrac{\frac{N}{4} \times 3 - n_{i-1}c\uparrow}{n_i} a_i$

0,5 $\quad Q_3 = 45 + \dfrac{75 - 72}{28} 10 = 46{,}07$



2

La boîte de Tuckey donne une vue rapide des caractéristiques élémentaires de la distribution statistique.

0,5 Q_2 indique le centre des donnés. La longueur de la boîte montre la <u>dispersion</u> de la moitié centrale des données et la longueur des segments latéraux extérieurs, la dispersion de la 4^{eme} partie inférieure et supérieur, respectivement.

3) <u>Le mode :</u>

Les amplitudes sont différentes.

On va appliquer la formule (I) : $Mo = l_{i-1} + \frac{l_{i+1}}{l_{i-1} + l_{i+1}} a_i$
0,5

Déterminons d'abord, la classe modale $\Rightarrow [40, 45[$

$$Mo = 40 + \frac{2,8}{3,75 + 2,8} \times 5$$

0,5 $Mo = 40 + \frac{14}{6,55} = 42,14$

Le salaire le plus fréquent est 42140 DH

4) <u>Interprétation :</u>

La médiale est la valeur de la variable qui diviserait la masse salariale totale en deux blocs égaux.

0,5 $\frac{\sum n_i c_i}{2} = \frac{3985}{2} = 1992,5$

On cherche cette valeur parmi les $(n_ic_i)c\uparrow$

la 1^{ere} valeur qui la dépasse est 2585

Donc la classe médiale est [40, 45[

On applique la formule $Ml = l_{i-1} + \dfrac{\frac{\sum n_ic_i}{2} - (n_{i-1}c_{i-1})c\uparrow}{n_ic_i} a_i$

1 $Ml = 40 + \dfrac{1992,5 - 1650}{935} \times 5 = 41,83$

* Calcul de ΔM :

0,5 On a $Q_2 = M\acute{e} = 40$

0,5 $\Delta M = 41,83 - 40 = 1,83$

0,5 $I = \dfrac{\Delta M}{E} = \dfrac{1,83}{x_{max} - x_{min}} = \dfrac{1,83}{55 - 25} = \dfrac{1,83}{30} = 0,061$

0,75 $\Rightarrow$ Faible concentration

5)

1

$p_i = \frac{n_ic\uparrow}{N} \times 100$	$q_i = \frac{(n_ic_i)c\uparrow}{\sum n_ic_i} \cdot 100$
20	14,30
50	41,40
72	64,87
100	100
242	220,57

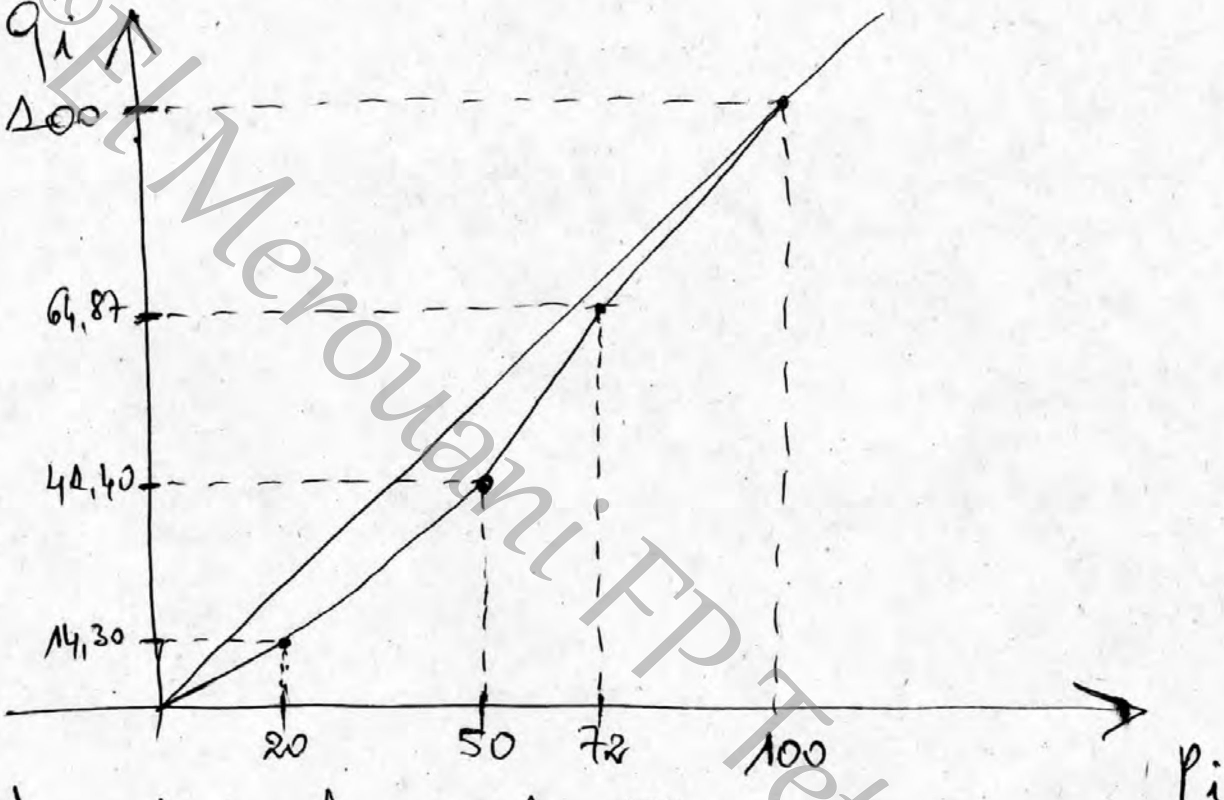
$I_G = 1 - \dfrac{\sum_{i=1}^{k-1} q_i}{\sum_{i=1}^{k-1} p_i}$

$I_G = 1 - \dfrac{120,57}{142}$

$I_G = 1 - 0,849$

$I_G = 0,151$

6) Courbe de Lorenz :



1

La courbe est proche de la 1ère bissectrice, ce qui indique bien une faible concentration comme on a
1 trouvé dans la question précédente $I_G = 0,151$
$\Longrightarrow$ Faible concentration.

Exercice 2 :

1) Nord :

$[e_{i-1}, e_i[$	n_i	c_i	$n_i c_i$	c_i^2	$n_i c_i^2$
$[0, 10[$	3	5	15	25	75
$[10, 30[$	15	20	300	400	6000
$[30, 50[$	20	40	800	1600	32000
$[50, 100[$	12	75	900	5625	67500
	$N=50$		2015		105575

0,5 $\bar{X} = \frac{1}{N} \sum_i n_i c_i = \frac{2015}{50} = 40{,}3$

Sud :

$[e_{i-1}, e_i[$	n_i	c_i	$n_i c_i$	c_i^2	$n_i c_i^2$
$[0, 10[$	5	5	25	25	125
$[10, 30[$	19	20	380	400	7600
$[30, 50[$	21	40	840	1600	33600
$[50, 100[$	5	75	375	5625	28125
	$N=50$		6		

$\Sigma = 1620$

69450

0,5 $\bar{X} = \frac{1}{N}\sum_i n_i c_i = \frac{1620}{50} = 32,4$

<u>Moyenne du pays :</u>

$$\bar{X} = \frac{1}{N}\left(N_1 \bar{X}_1 + N_2 \bar{X}_2\right)$$

0,5 $$= \frac{1}{100}\left(50 \times 40,3 + 50 \times 32,4\right) = \frac{2015 + 1620}{100}$$

$$= 36,35$$

2) <u>Nord :</u>

$$Var(X) = \frac{1}{N}\sum_i (c_i - \bar{x})^2 n_i = \left(\frac{1}{N}\sum_i n_i c_i^2\right) - \bar{x}^2$$

0,5 $$= \frac{105575}{50} - (40,3)^2 = 2111,5 - 1624,09$$

$$= 487,41$$

<u>Sud :</u>

$$Var(X) = \frac{69450}{50} - (32,4)^2 = 1389 - 1049,76$$

0,5 $$= 339,24$$

Variance globale :

$[x_{i-1}, x_i[$	n_i	c_i	c_i^2	$n_i c_i^2$
[0, 10[	8	5	25	200
[10, 30[	34	20	400	13600
[30, 50[	41	40	1600	65600
[50, 100[	17	75	5625	95625
	100			175025

$$Var(X) = \frac{175025}{100} - (36,35)^2 = 1750,25 - 1321,32$$

$$= 428,93$$

3) Variance Intra-région :

$$Var(X_i) = \frac{1}{N}\left[N_1 Var(X_1) + N_2 Var(X_2)\right]$$

$$= \frac{1}{100}\left[50 \times 487,41 + 50 \times 339,24\right]$$

$$= \frac{1}{100}\left[24370,5 + 16962\right] = 413,325$$

Variance Inter-région :

$$Var(\overline{X_i}) = \frac{1}{N}\left[N_1(\overline{X}_1 - \overline{X})^2 + N_2(\overline{X}_2 - \overline{X})^2\right]$$

$$= \frac{1}{100}\left[50(40,3 - 36,35)^2 + 50(32,4 - 36,35)^2\right]$$

1

$$= \frac{1}{100}\left[50(\times 15,602) + 50 \times 15,602\right] = 15,602.$$

Conclusion :

$$Var(X) = \overline{Var(X_i)} + Var(\overline{X_i})$$

0,5

$$428,93 = 413,328 + 15,602$$

Interprétation :

La dispersion des exploitations dans ce pays s'explique par une faible part (inter parmi $\rightarrow$ 3,64%)

0,5 par une dispersion des terres entre les régions, alors que la forte dispersion des terres, interne aux régions explique une grande partie (intra à l'intérieur $\rightarrow$ 96,36%) de la dispersion des terres.

9

4) 0,5 $$Cv(\text{Nord}) = \frac{\sigma(x)}{\overline{X}} = \frac{\sqrt{Var(x)}}{\overline{X}}$$

$$= \frac{\sqrt{487{,}41}}{40{,}3} = \frac{22{,}077}{40{,}3} = 0{,}55$$

0,5 $$Cv(\text{Sud}) = \frac{\sigma(x)}{\overline{X}} = \frac{\sqrt{339{,}24}}{32{,}4} = \frac{18{,}418}{32{,}4}$$

$$= 0{,}57$$

0,5 est presque la même.

$x \longrightarrow 100$

$413{,}325 \longrightarrow 428{,}93$

$15{,}602 \longrightarrow 428{,}93$

$$\text{Intra} \longrightarrow x = \frac{413{,}325 \times 100}{428{,}93} = 96{,}36\ \%$$

$$\text{Inter} \longrightarrow x = \frac{15{,}602 \times 100}{428{,}93} = 3{,}64\ \%$$

10