CORRIGE EXAMEN BLANC

L1Eco. Statistique

Mai 2012

1 EXERCICE-1

1. La moyenne est donnée par : $\overline{x} = \sum f_i x_i$; la calculatrice donne : $\overline{x} = 43.94$
2. Cette série représentant les données relatives à un caractère quantitatif continu ; les classes étant d'amplitudes inégales, on utilise la densité, $d_i = \dfrac{n_i}{A_i}$, la classe modale étant celle de plus grande densité ; les effectifs corrigés $n_{icor} = 5d_i$, 5 étant l'amplitude minimale de classe.

a_i	b_i	f_i	x_i	Ai	di	ficor	ficc
0	20	23.03	10	20	1.15	5.757	23.026
20	60	47.53	40	40	1.19	5.941	70.557
60	65	5.99	62.5	5	1.20	5.990	76.547
65	75	11.11	70	10	1.11	5.555	87.657
75	105	12.34	90	30	0.41	2.057	100.000
		100.000					

La classe modale est celle de plus grande densité, c'est-à dire la classe $[60; 65[$ et le mode est calculé en considérant les classes encadrant la classe modale, ce qui donne avec les notations du cours : $\begin{cases} x_1 = 60 \\ x_2 = 65 \end{cases}$, $\begin{cases} h = 5.99 \\ h_1 = 5.941 \text{ et } h_2 = 5.555 \end{cases}$

$\begin{cases} k_1 = h - h_1 = 5.99 - 5.941 = 0.049 \\ k_2 = h - h_2 = 5.99 - 5.555 = 0.435 \end{cases}$: et pour conclure : $M_o = \dfrac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \dfrac{0.435 * 60 + 0.049 * 65}{0.435 + 0.049} = 60.51$ ans

3. $V(x) = \sum f_i x_i^2 - \overline{x}^2 = \sigma^2(x) = 25.11^2 = 630.51$ et $\sigma(x) = 25.11$
4. On donne l'écart inter-quartile $EIQ = 42.05 = Q_3 - Q_1$ donc $Q_3 = 42.05 + 21.66. = 63.71$.
5. Calcul de Q_2 : Q_2 correspond à une fréquence cumulée croissante de 50% et on localise Q_2 dans la classe $[20; 60[$ où la féquence cumulée croissante dépasse 50%, puis on effectue une interpolation linéaire :
$\frac{70.557-23.026}{60-20} = \frac{50-23.026}{Q_2-20}$ soit $Q_2 - 20 = 40 * \frac{50-23.026}{70.557-23.026}$soit $Q_2 = 20 + 40 * \frac{50-23.026}{70.557-23.026} = 42.7$ Il y aura donc en 2015,selon les
projections de l'Insee, 50% de la population qui aura un âge inférieur ou égal à 42.70 ans.
6. $C_Y = \dfrac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1} = \dfrac{21.66 + 63.71 - 2 * 42.70}{42.05} = -7.134\,4 \times 10^{-4}$; ce coefficient est négatif, ce qui traduit un étalement à gauche, mais en fait il est quasiment nul, ce qui est le signe d'une distribution proche de la symétrie.

2 EXERCICE-2

1. On trouve : $\overline{X} = 286.85$ et $\overline{Y} = 183.42$
2. $\sigma(X) \simeq 166.25$ et $\sigma(Y) \simeq 220.46$,donc $V(X) \simeq 166.25^2 = 27639.06$ et $V(Y) \simeq 220.46^2 = 48602.61$
3. La covariance peut se calculer avec la formule : on obtient avec la calculatrice : $\sum x_i y_i = 1090244.8$, soit en remplaçant : $Cov(x; y) = \dfrac{1}{n} \sum x_i y_i - \overline{x} * \overline{y} = \dfrac{1090244.8}{13} - 286.85 * 183.42 = 31250.96$
4. $r = \dfrac{Cov(x; y)}{\sigma(x)\,\sigma(y)} = \dfrac{31250.96}{166.25 * 220.46} = 0.852\,7$ Ce coefficient est toujours compris entre -1et 1; il est assez proche de 1ce

qui valide l'existence d'une correlation linéaire entre les variables.

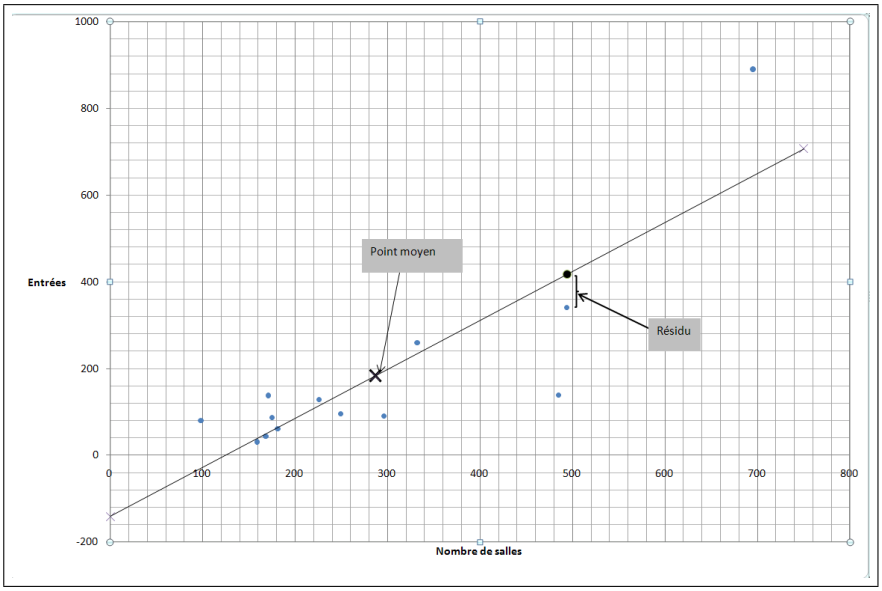
5. On trouve : $\widehat{y} = \widehat{a}x + \widehat{b} = 1.1307x - 140.9229$ avec $\boxed{\begin{array}{c} \widehat{a} = \dfrac{Cov\,(x;y)}{V(x)} \\ \widehat{b} = \overline{y} - \widehat{a}\overline{x} \end{array}}$

6. $\widehat{a}$ représente la variation de Y consécutive à une augmentation de X de une unité : si le nombre de salle augmente de 1, on peut estimer la variation du nombre de spectateurs à une augmentation de 1131. Ici, $\widehat{b}$ n'a pas de sens, il donnerait le nombre de spectateurs pour un film projeté dans aucune salle...

7. cf graphique.

8. $\widehat{y}\,(750) = 1.1307 * 750 - 140.9229 = 707.1$

9. $e_7 = y_7 - \widehat{y}_7 = 340.5 - (1.1307 * 494 - 140.9229) = -77.14$; ce résidu est négatif ; il mesure la différence entre la valeur observée y_7 et la valeur estimée par le modèle $\widehat{y}_7$; ce résidu est négatif car le point M_7 du nuage est situé sous la droite de régréssion.



3 EXERCICE-3

1.

Effectifs partiels et marginaux				
	Salaire			
Nombre d'enfants	2	3	5	n_{i+}
1	3	6	12	21
2	5	10	20	35
3	2	4	8	14
4	4	8	16	28
n+j	14	28	56	98

;

x_i	ni+	$n_{i+}*x_i$
1	21	21
2	35	70
3	14	42
4	28	112
n+j	98	245

on peut à partir de ce tableau extraire la série marginale des x :

, qui donne : $\overline{x} = \frac{245}{98} = 2.5$, soit une moyenne de 2.5 enfants.

2. Il s'agit de la fréquence partielle : $f_{22} = \dfrac{n_{22}}{n_{++}} = \frac{10}{98} = 0.1020$

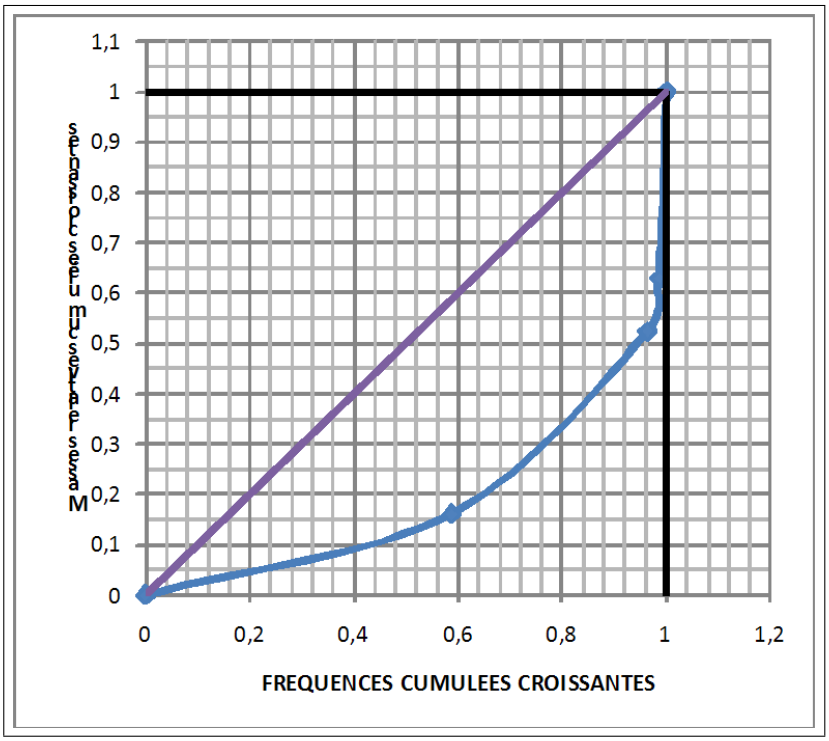
3. Il s'agit d'une fréquence conditionnelle : $f_{(y=5/x=4)} = \dfrac{n_{43}}{n_{4+}} = \dfrac{16}{28} = 0.5714$. $57.14\ \%$ des parents de 4 enfants ont un salaire de 5000 euros.

4 EXERCICE-4

a_i	b_i	n_i	x_i	f_i	$f_i cc$	$n_i xi$	q_i	$q_i cc$	S_i
0	2	222279	1	0,5870	0,5870	222279	0,1614	0,1614	0,0474
2	5	142380	3,5	0,3760	0,9630	498330	0,3617	0,5231	0,1287
5	30	8331	17,5	0,0220	0,9850	145792,5	0,1058	0,6289	0,0127
30	150	5680	90	0,0150	1,0000	511200	0,3711	1,0000	0,0122
		378670				1377601,5	1		0,2009

1. La médiane se localise à l'aide des fréquences cumulées croissantes dans la classe $[0;2[$, puis on effectue une interpolation linéaire :

   $A\,(0:0)$, $B\,(2;0.5870)$ et $M\,(Me;0.50)$, ce qui donne : $\frac{0.587}{2} = \frac{0.5}{Me}$, soit $Me = \dfrac{1}{0.587} \simeq 1.70$. 50% des entreprises de ce secteur ont un chiffre d'affaire inférieur ou égal à 1.70 millions d'euros.

2. 

3. On doit calculer l'aire de concentration : $A_c = 0.5 - \sum S_i = 0.50 - 0.2009 = 0.2991$ et l'indice de Gini est défini par : $I_G = \dfrac{A_c}{Aire(OAB)} = \dfrac{0.2991}{0.5} = 0.5982$; l'indice de Gini est toujours compris entre 0 et 1 ; quand il est proche de 1, la concentration est forte ; ici la concentration est assez forte.