CORRIGE PARTIEL A

AES-STATISTIQUE

MAI 2012

1 EXERCICE 1

1. On note a_i et b_i les bornes inférieures et supérieures des classes et on calcule les amplitudes de classes : $b_i - a_i$. Les classes étant d'amplitudes inégales, on doit corriger les effectifs en utilisant la densité $d_i = \dfrac{n_i}{b_i - a_i}$. On a pris pour effectifs corrigés : $n_i cor = 20 d_i$, c'est-à dire le produit de la densité par l'amplitude minimale.

a_i	b_i	n_i	x_i	A_i	d_i	$n_i cor$	$n_i x_i$	$n_i x^2_i$
0	60	10	30	60	0,167	3,33	300	9000
60	70	48	65	10	4,8	96	3120	202800
70	90	52	80	20	2,6	52	4160	332800
90	110	34	100	20	1,7	34	3400	340000
110	130	28	120	20	1,4	28	3360	403200
130	150	22	140	20	1,1	22	3080	431200
150	200	6	175	50	0,12	2,4	1050	183750
		200					18470	1902750

La classe modale est celle qui a la plus grande densité, soit la classe $[60; 70[$; c'est le rectangle le plus haut de l'histogramme. Le mode est calculé en considérant les classes encadrant la classe modale, ce qui donne avec les notations du cours : $\begin{cases} x_1 = 60 \\ x_2 = 70 \end{cases}$, $\begin{cases} h = 96 \\ h_1 = 3.33 \text{ et } h_2 = 52 \end{cases}$ $\begin{cases} k_1 = h - h_1 = 96 - 3.33 = 92.67 \\ k_2 = h - h_2 = 44 \end{cases}$: et pour conclure :

$M_o = \dfrac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \dfrac{44 * 60 + 92.67 * 70}{44 + 92.67} \simeq \boxed{66.78}$; comme prévu, le mode est largement attiré à droite.

2. La calculatrice donne : $\overline{x} =$

3. $V(x) \simeq \frac{1902750}{200} - 92.35^2 = 985.23$ et $\sigma(x) = \sqrt{V(x)} = \sqrt{985.23} = 31.39$

4. $\overline{x} - 2\sigma(x) = 92.35 - 2 * 31.39 = 29.57$ et $\overline{x} + 2\sigma(x) = 92.35 + 2 * 31.39 = 155.13$. On doit donc ajouter les effectifs des classes de 60 à 150, puis estimer les effectifs des intervalles $[29.57; 60[$ et $[150; 155.13[$, ce qui se fait en utilisant leurs densités :

Intervalle	Amplitude	Densité	Effectif
$[29.57; 60[$	30.43	0.167	$30.43 * 0.167 = 5.08$
$[150; 155.13[$	5.13	0.12	$5.13 * 0.12 = 0.62$

, ce qui donne un total de :

$(48 + 52 + 34 + 28 + 22) + 5.08 + 0.62 \simeq 189.7$ soit une proportion de : $\dfrac{189.7}{200} \simeq 0.948\,5$ soit 94.85%.

2 EXERCICE-2

1.

	X	Y
Moyenne	**12**	**8.67**
Ecart-type	**3.74**	**2.84**
Variance	**14**	**8.09**

2. Covariance

i	x_i	y_i	x_iyi
1	6	4.46	26.8
2	7	5.77	40.4
3	8	5.98	47.8
4	9	7.33	66
5	10	7.32	73.2
6	11	6.58	72.4
7	12	7.81	93.7
8	13	7.84	102
9	14	11	154
10	15	10.7	160
11	16	10.8	173
12	17	13.6	232
13	18	13.5	244
			1485

, la covariance peut se calculer avec la formule : $Cov\,(x;y) = \frac{1}{n}\sum x_i y_i - \overline{x} * \overline{y} = \frac{1485}{13} - 12 * 8.67 \simeq 10.19$

3. On trouve : $\widehat{y} = \widehat{a}x + \widehat{b} = \boxed{0.7241x - 0.0148.}$
4. $\widehat{a} \simeq 0.7241$, représente en dollars, l'augmentation de salaire prévisible quand le nombre d'années d'étude augmente de 1 .
5. $r \simeq 0.952\,7$; ce coefficient mesure l'intensité de la relation linéaire liant x et y ; ce coefficient est toujours compris entre -1 et 1
6. $e_5 = y_5 - \widehat{y_5} = 7.32 - (0.7241 * 10 - 0.0148) = 0.093\,8$; ce résidu est positif et traduit le fait que le modèle surestime légèrement y_5 ; le point du nuage esr$t situé sous la droite de régréssion.
7. $\widehat{y}\,(20) = 0.7241 * 20 - 0.0148 \simeq 14.47\,\$$.

3 EXERCICE-3

1.

Effectifs et fréquences marginales				
Age	F	H	n_{i+}	f_{i+}
[16;25[	147	283	430	0,4304
[25;35[	61	275	336	0,3363
[35;45[	33	62	95	0,0951
[45;55[	20	55	75	0,0751
[55;60[	22	41	63	0,0631
n_{+j}	283	716	999	
f_{+j}	0,2833	0,7167		1

Age	x_i	fi+	fi+*xi
[16;25[	20,5	0,4304	8,8238
[25;35[	30	0,3363	10,0901
[35;45[	40	0,0951	3,8038
[45;55[	50	0,0751	3,7538
[55;60[	57,5	0,0631	3,6261
		1	30,10

2. Il s'agit de la fréquence partielle : $f_{11} = \frac{147}{999} = 0.147\,1$

3. Il s'agit d'une fréquence conditionnelle : $f_{(x=x_3/y=y_2)} = \dfrac{n_{32}}{n_{+2}} = \dfrac{62}{716} = 8.66 \times 10^{-2}$; 8.66% des hommes ont un âge compris entre 35 et 45 ans.

4. On extrait la série des âges , ce qui donne un âge moyen de 30.10 ans.

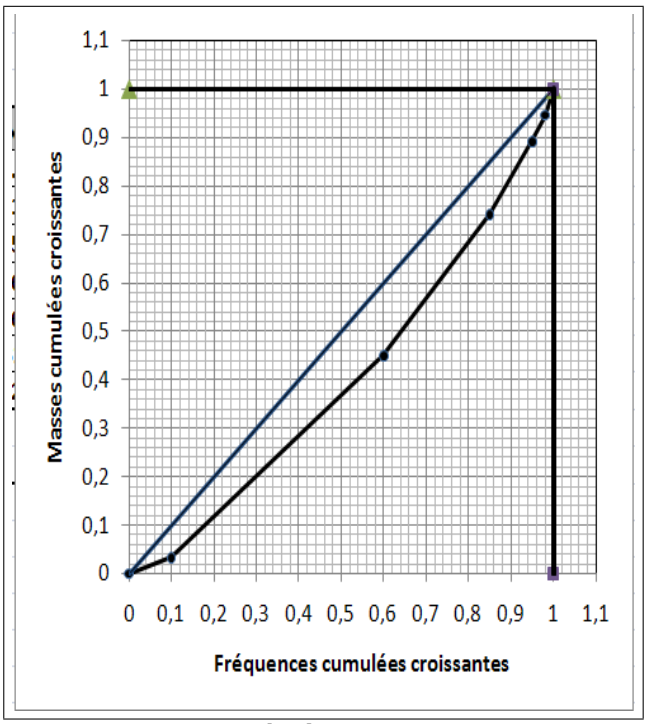
4 EXERCICE-4

1. La médiale se détermine comme la médiane à partir des masses cumulées croissantes. On la localise dans la classe $[15; 20[$, puis on la détermine par interpolation linéaire, entre les points : $A\,(15; 0.45)$ et $B\,(20; 0.7417)$: $\dfrac{0.7417 - 0.45}{20 - 15} = \dfrac{0.50 - 0.45}{Ml - 15}$ soit : $Ml = \dfrac{0.05 * 5}{0.2917} + 15 \simeq 15.86$. Les salariés ayant un salaire inférieur à 15.86 K€ se partagent la moitié de la masse salariale.

2. Indice de Gini

   Les calculs sont présentés dans le tableau, les S_i sont les aires des trapèzes situés sous la courbe de Gini, ce qui donne pour l'aire de concentration : $A = 0.5 - 0.4002 = 0.099\,8$ et pour l'indice de Gini : $I_G = 2 * A \simeq 2 * 0.0998 = 0.1996$; cet indice est toujours compris entre 0 et 1 ; ici il est plus proche de 0 que de 1, ce qui traduit une concentration assez faible.

Salaires	n_i	f_i	$f_i$cc	x_i	f_ix_i	q_i	$q_i$cc	S_i
[0;10[	10	0.1	0.10	5	0.5	0.0333	0.0333	0.0017
[10;15[	50	0.5	0.60	12.5	6.25	0.4167	0.4500	0.1208
[15;20[	25	0.25	0.85	17.5	4.375	0.2917	0.7417	0.1490
[20;25[	10	0.1	0.95	22.5	2.25	0.1500	0.8917	0.0817
[25;30[	3	0.03	0.98	27.5	0.825	0.0550	0.9467	0.0276
[30;50[	2	0.02	1.00	40	0.8	0.0533	1.0000	0.0195
	100				15	1.0000		0.4002
Indice de	0.1997							



Courbe de Lorentz

3. On localise la médiane grâce aux fréquences cumulées croissantes, dans la classe $[10; 15[$, puis on la détermine par

interpolation linéaire, entre les points : $A\,(10; 0.10)$ et $B\,(15; 0.60)$, ce qui donne : $\dfrac{0.60 - 0.10}{15 - 10} = \dfrac{0.50 - 0.10}{Me - 10}$ soit : $Me - 10 = \dfrac{0.40 * 5}{0.50}$ soit $Me = 14.$ 50% des salaires de cette entreprise sont inférieurs à 14 K€.

4. La médiale est supérieure à la médiane, ce qui est toujours le cas ; les salarés ayant un salaire inférieur à la médiane représentent 50% de l'effectif total et constituent les "bas salaires", ils se partagent donc une part de la masse salariale inférieure à 50%.