CORRIGE EXAMEN DE STATISTIQUE sujet B

I EXERCICE-1 $(8pts)$

1. La moyenne est donnée par : $\overline{x} = \sum f_i x_i$; la calculatrice donne : $\overline{x} = 40.12$
2. Cette série représentant les données relatives à un caractère quantitatif continu ; les classes étant d'amplitudes inégales, on utilise la densité, $d_i = \dfrac{f_i}{A_i}$, la classe modale étant celle de plus grande densité ; les effectifs corrigés $f_{icor} = 5d_i$, 5 étant l'amplitude minimale de classe.

a_i	b_i	f_i	x_i	Ai	di	ficor	ficc
0	20	24.78	10	20	1.24	6.196	24.784
20	60	53.76	40	40	1.34	6.720	78.542
60	65	4.94	62.5	5	0.99	4.939	83.480
65	75	8.06	70	10	0.81	4.031	91.543
75	100	8.46	87.5	25	0.34	1.692	100.000
		100.000					

La classe modale est celle de plus grande densité, c'est-à dire la classe $[20;60[$ et le mode est calculé en considérant les classes encadrant classe modale, ce qui donne avec les notations du cours : $\left\{\begin{array}{l} x_1 = 20 \\ x_2 = 60 \end{array}\right.$, $\left\{\begin{array}{l} h = 6.72 \\ h_1 = 6.196 \text{ et } h_2 = 4.939 \end{array}\right.$

$\left\{\begin{array}{l} k_1 = h - h_1 = 6.72 - 6.196 = 0.524 \\ k_2 = h - h_2 = 6.72 - 4.939 = 1.781 \end{array}\right.$. et pour conclure : $M_o = \dfrac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \dfrac{1.781 * 20 + 0.524 * 60}{1.781 + 0.524} = 29.09$

3. $V(x) = \sum f_i x_i^2 - \overline{x}^2 = \sigma^2(x) = 22.62^2 = 511.66$ et $\sigma(x) = 22.62$
4. Calcul de Q_1 : Q_1 correspond à une fréquence cumulée croissante de 25% et on localise Q_1 dans la classe $[20;60[$ où la féquence cumulée croissante dépasse 25%, puis on effectue une interpolation linéaire :
$\frac{78.542-24.784}{60-20} = \frac{25-24.784}{Q_1-20}$ soit $Q_1 - 20 = 40 * \frac{25-24.784}{78.542-24.784}$soit $Q_1 = 20 + 40 * \frac{25-24.784}{78.542-24.784} = 20.161$ Il y aura donc en 2015,selon les
projections de l'Insee, 50% de la population qui aura un âge inférieur ou égal à 20.16 ans.
5. On a : On donne $Q_2 \simeq 38.76$ et $Q_3 \simeq 57.37$. Calculer le coefficient deYule et commenter le résultat.
$C_Y = \dfrac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1} = \dfrac{20.16 + 57.37 - 2 * 38.76}{57.37 - 20.16} = 2.6874 \times 10^{-4}$; ce coefficient est positif, ce qui traduit un étalement à droite, mais en fait il est quasiment nul, ce qui est le signe d'une distribution proche de la symétrie.

II EXERCICE-2$(12pts)$

1. On trouve : $\overline{X} = 0.2381$ et $\overline{Y} = 0.2003$
2.

	X	Y
Ecart-type	0.0449	0.0326
Variance	0.0020	0.0011

3. La covariance peut se calculer avec la formule : on obtient avec la calculatrice $\sum x_i y_i = 0.6330$ puis en remplaçant : $Cov\,(x;y) = \dfrac{1}{n}\sum x_i y_i - \overline{x} * \overline{y} = \dfrac{0.6330}{13} - 0.2381 * 0.2003 = 1.000\,9 \times 10^{-3}$
4. On trouve : $\widehat{y} = \widehat{a}x + \widehat{b} = 0.5051x + 0.0800$ avec $\boxed{\begin{array}{c} \widehat{a} = \dfrac{Cov\,(x;y)}{V\,(x)} \\ \widehat{b} = \overline{y} - \widehat{a}\overline{x} \end{array}}$
5. $r = \dfrac{Cov\,(x;y)}{\sigma\,(x)\,\sigma\,(y)} = \frac{1.000\,9\times10^{-3}}{0.0449*0.0326} = 0.683\,8$ 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.
6. $\widehat{y}\,(0.2068) = 0.5051 * 0.2068 + 0.0800 = 0.184\,5$
7. $R^2 = 0.6838^2 = 0.4676$; $R^2 = \frac{\mathbf{SCE}}{\mathbf{SCT}}$ ce coefficient donne la part de la variation totale expliquée par le modèle, ici $46.76\ \%$
8. L'équation de l'analyse de la variance est : $\mathbf{SCT} = \mathbf{SCE} + \mathbf{SCR}$
   $\mathbf{SCT} = \mathbf{n}V(y) = 13{*}0.0011 = 0.014\,3$, $\mathbf{SCE} = R^2{*}\mathbf{SCT} = 0.4676{*}0.014\,3 = 6.686\,7{\times}10^{-3}$ et donc $\mathbf{SCR} = \mathbf{SCT} - \mathbf{SCE} = 0.014\,3 - 6.686\,7 \times 10^{-3} = 7.\,613\ \times 10^{-3}$