CORRIGE PARTIEL BLANC

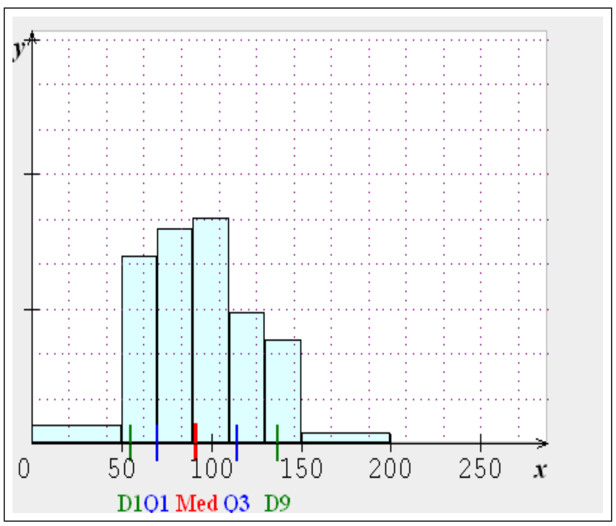
Janvier 2012

L1EcoStat.

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 : $b_i - a_i$. On représente ce caractère continu par un histogramme ; 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	Ai	di	nicor	nicc
0	50	10	25	50,00	0,20	4	10
50	70	40	60	20,00	2,00	40	50
70	90	46	80	20,00	2,30	46	96
90	110	48	100	20,00	2,40	48	144
110	130	28	120	20,00	1,40	28	172
130	150	22	140	20,00	1,10	22	194
150	200	6	175	50,00	0,12	2,4	200
		200					



2. La classe modale est celle qui a la plus grande densité, soit la classe $[90; 110[$; 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 : $\begin{cases} x_1 = 90 \\ x_2 = 110 \end{cases}$, $\begin{cases} h = 48 \\ h_1 = 46 \text{ et } h_2 = 28 \end{cases}$ $\begin{cases} k_1 = h - h_1 = 2 \\ k_2 = h - h_2 = 20 \end{cases}$ et pour conclure : $M_o = \dfrac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \dfrac{20 * 90 + 2 * 110}{22} \simeq \boxed{91.82}$; comme prévu, le mode est largement attiré à gauche.

3. La calculatrice donne : $\overline{x} = 93.1$, $\sigma(x) \simeq 32.35$ et $V(x) = \sigma^2(x) \simeq 32.35^2 \simeq 1046.52$.

4. $\overline{x} - 1.5\sigma(x) = 93.1 - 1.5 * 32.35 = 44.58$ et $\overline{x} + 1.5\sigma(x) = 93.1 + 1.5 * 32.35 = 141.63$. On doit donc ajouter les effectifs des classes de 50 à 130, puis estimer les effectifs des intervalles $[44.58; 50[$ et $[130; 141.63[$, ce qui se fait

en utilisant leurs densités :

Intervalle	Amplitude	Densité	Effectif
$[44.58; 50[$	5.42	0.2	$5.42 * 0.2 = 1.08$
$[130; 141.63[$	11.63	1.1	$11.63 * 1.1 = 12.79$

, ce qui donne un total de : $(40 + 46 + 48 + 28) + 1.08 + 12.79 = 175.87$, soit une proportion de : $\dfrac{175.87}{200} = 0.8794$, soit 87.94%.

5. On doit calculer les quartiles pour déterminer le coefficient de Yule : $C_Y = \dfrac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1}$;pour calculer les quartiles, on utilise les effectifs cumulés croissants et on effectue une interpolation linéaire.

   On trouve directement $Q_1 = 70$; pour déterminer Q_2,on calcule $\dfrac{n}{2} = 100$; les effectifs cumulés croissants permettent de localiser Q_2 dans la classe $[90; 110[$; on considère alors $A(90, 96)$, $B(110, 144)$ et $M\,(Q_2; 100)$, ce qui donne : $\dfrac{144 - 96}{110 - 90} = \dfrac{100 - 96}{Q_2 - 90}$ soit : $Q_2 - 90 = \dfrac{100 - 96}{144 - 96} * 20$,soit $Q_2 = 90 + \dfrac{100 - 96}{144 - 96} * 20 \simeq 91.67$; et enfin $\dfrac{172 - 144}{130 - 110} = \dfrac{150 - 144}{Q_3 - 110}$soit $Q_3 - 110 = \dfrac{150 - 144}{172 - 144} * 20$ et $Q_3 = 110 + \dfrac{150 - 144}{172 - 144} * 20 = 114.29$; on trouve alors : $C_Y = \dfrac{Q_1 + Q_3 - 2Q_2}{Q_3 - Q_1} = \dfrac{70 + 114.29 - 2 * 91.67}{114.29 - 70} = 2.1450 \times 10^{-2}$; ce coefficient est positif, ce qui traduit un étalement à droite ; cependant ce coefficient est très proche de 0,ce qui est le signe d'une distribution proche de la symétrie.

2 EXERCICE-2

1. On trouve : $\overline{X} = 582.8$ et $\overline{Y} = 365.5$.

	X	Y
Moyenne	**582.8**	**365.5**
Ecart-type	**26.06**	**34.40**
Variance	**678.96**	**1183.05**

2.

x_i	y_i	x_iyi
525	325	170625
554	362	200548
575	315	181125
579	355	205545
585	325	190125
586	370	216820
590	390	230100
608	420	255360
610	410	250100
616	383	235928
5828	3655	2136276

3. La covariance peut se calculer avec la formule : $Cov\,(x; y) = \dfrac{1}{n} \sum x_i y_i - \overline{x} * \overline{y} = \dfrac{2136276}{10} - 582.8 * 365.5 \simeq 614.2$.

4. On trouve : $\widehat{y} = \widehat{a}x + \widehat{b} = 0.9046x - 161.71$

5. $\widehat{a} \simeq 0.9046$, représente en roupies, la variation(augmentation) de la dépense alimentaire quand la dépense totale augmente de 1 roupie.

6. $r = \dfrac{Cov\,(x;y)}{\sigma\,(x)\,\sigma\,(y)} = \frac{614.2}{26.06*34.40} = 0.685\,1$. 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.
   $R^2 = 0.685\,1^2 \simeq 0.469\,4$; ce coefficient donne la part de la variation totale expliquée par le modèle, ici 46.94%.
7. $\widehat{y}\,(650) = 0.9046 * 650 - 161.71 \simeq 426.28$ roupies
8. **Complément** : l'équation de l'analyse de la variance est : $\mathbf{SCT} = \mathbf{SCE} + \mathbf{SCR}$
   $\mathbf{SCT} = \mathbf{n}V(y) = 11830.5$. $\mathbf{SCE}$ $=R^2*\mathbf{SCT}$ $=0.469\,4*11830.5 = 5553.\,2$ et donc $\mathbf{SCR} = \mathbf{SCT} - \mathbf{SCE}$ $=11830.5-5553.\,2 = 6277.\,3$; on rappelle que $R^2 = \dfrac{\mathbf{SCE}}{\mathbf{SCT}} = \dfrac{5553.\,2}{11830.5} = 0.469\,4$.

3 EXERCICE-3

1.

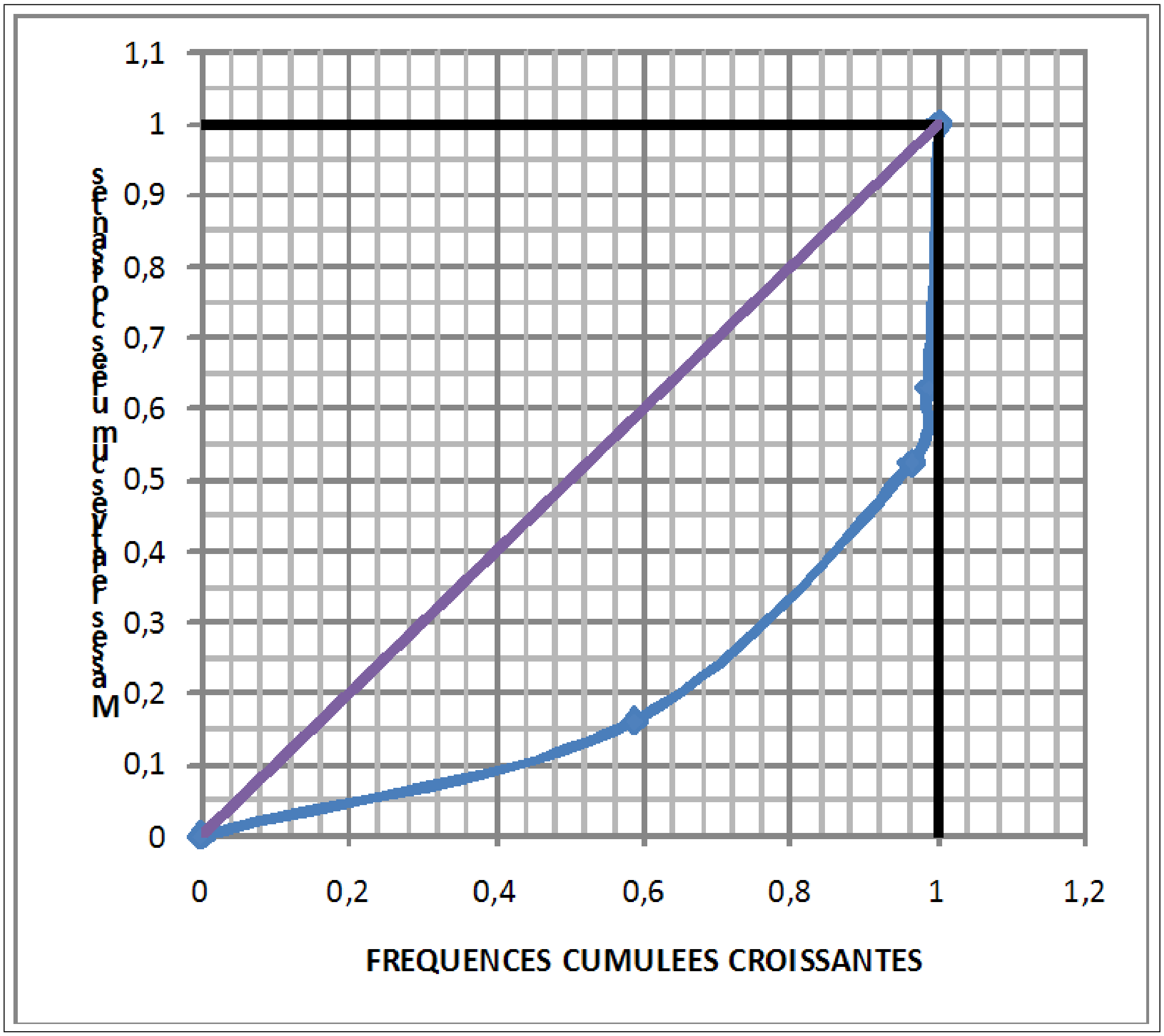
	Sportive	Sûre	n_{i+}	f_{i+}
BMW	256	74	330	**0,6433**
MERCEDES	41	42	83	**0,1618**
LEXUS	66	34	100	**0,1949**
n_{j+}	363	150	513	
f_{j+}	**0,7076**	**0,2924**		**1**

   Les fréquences marginales sont en gras dans le tableau ci-dessus.
2. Cette fréquence partielle est donnée par : $f_{11} = \dfrac{256}{513} = 0.499\,0$. Il y a 49.90 % de voitures qui sont de marque BMW et perçues comme sportives
3. Il s'agit d'une fréquence conditionnelle : $f_{(j=2/i=3)} = f_{y=Sure/x=Lexus} = \dfrac{34}{100} = 0.34$. Il y a 34% de véhicules perçus comme sûrs parmi les Lexus.

4 EXERCICE-4

a_i	b_i	n_i	x_i	f_i	f_icc	n_ixi	q_i	q_icc	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 : $A_c = 0.5 - \sum S_i = 0.50 - 0.2009 = 0.299\,1$ et l'indice de Gini est défini par : $I_G = \dfrac{A_c}{Aire(OAB)} = \dfrac{0.2991}{0.5} = 0.598\,2$; 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.