

SUJET A-CORRIGE PARTIEL DE STATISTIQUE

L1-ECO

Janvier 2012

1 EXERCICE-1

Salaires (en	ni	fi	ficc	xi	Ai	di	ni xi	qi	qicc	Si
[0;10[10	0,1	0,10	5	10	1	50	0,0333	0,0333	0,0017
[10;15[50	0,5	0,60	12,5	5	10	625	0,4167	0,4500	0,1208
[15;20[25	0,25	0,85	17,5	5	5	437,5	0,2917	0,7417	0,1490
[20;25[10	0,1	0,95	22,5	5	2	225	0,1500	0,8917	0,0817
[25;30[3	0,03	0,98	27,5	5	0,6	82,5	0,0550	0,9467	0,0276
[30;50[2	0,02	1,00	40	20	0,1	80	0,0533	1,0000	0,0195
	100						1500	1,0000		0,4002

1. La moyenne est donnée par : $\bar{x} = \frac{1}{n} \sum n_i x_i$, $V(x) = \frac{1}{n} \sum n_i x_i^2 - \bar{x}^2 = \sigma^2(x)$; les résultats sont dans le tableau.

2. Calcul de Q_2 : Q_2 correspond à une fréquence cumulée croissante de 50% et on localise Q_2 dans la classe [10; 15[où la fréquence cumulée croissante dépasse 50%, puis on effectue une interpolation linéaire :

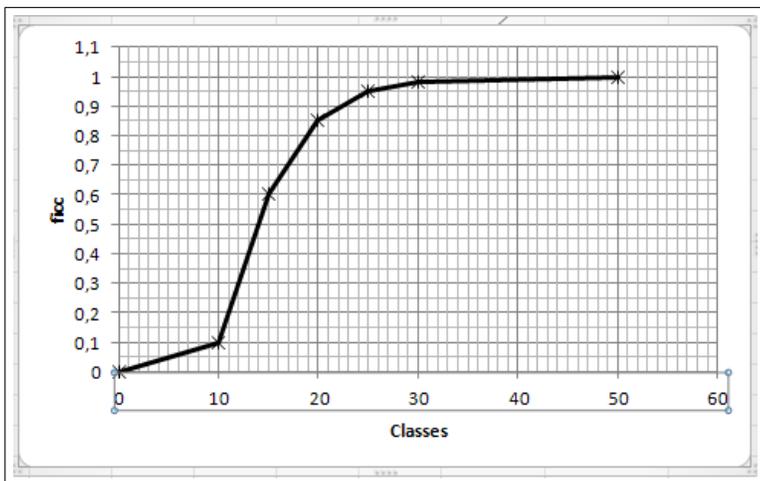
$$\frac{0,60-0,10}{15-10} = \frac{0,50-0,10}{Q_2-10} \text{ soit } Q_2 - 10 = \frac{0,40 \cdot 5}{0,50} \text{ soit } Q_2 = 10 + \frac{0,40 \cdot 5}{0,50} = 14. \text{ Il y a donc 50\% de la population qui ont un salaire inférieur ou égal à 14 K€}$$

3. Pour déterminer le mode on peut utiliser indifféremment la densité ou les effectifs corrigés (ils sont proportionnels) ; on a utilisé ici les densités.

La classe modale est celle de plus grande densité, c'est-à dire la classe [10; 15[et le mode est calculé en considérant les

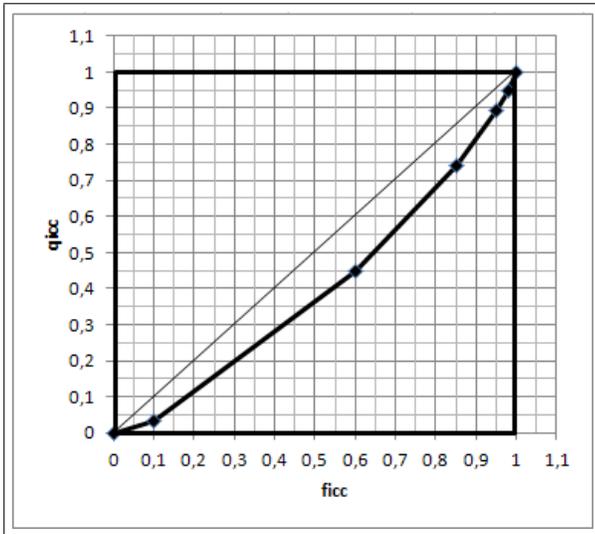
$$\text{classes encadrant classe modale, ce qui donne avec les notations du cours : } \begin{cases} x_1 = 10 \\ x_2 = 15 \end{cases}, \begin{cases} h = 10 \\ h_1 = 1 \text{ et } h_2 = 5 \end{cases}$$

$$\begin{cases} k_1 = h - h_1 = 9 \\ k_2 = h - h_2 = 5 \end{cases} \text{ et pour conclure : } M_o = \frac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \frac{5 \cdot 10 + 9 \cdot 15}{14} = 13,21. \text{ On trouve : } M_o \leq M_e \leq \bar{x}, \text{ ce qui indique une série asymétrique étalée à droite.}$$



4.

5.



6. On calcule l'aire de concentration : $A_C = 0.5 - \sum S_i = 0.5 - 0.4002 = 0.0998$. On obtient alors l'indice de Gini : $2 * \text{Aire de concentration}$ soit $I_G = 2 * 0.0998 = 0.1996$

Le coefficient de Gini est plus proche de 0 que de 1, la concentration est faible.

2 EXERCICE-2

1. On trouve : $\bar{X} = 286.85$ et $\bar{Y} = 183.42$
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 : $Cov(x; y) = \frac{1}{n} \sum x_i y_i - \bar{x} * \bar{y} = \frac{1090244.8}{13} - 286.85 * 183.42 = 31250.96$
4. $r = \frac{Cov(x; y)}{\sigma(x) \sigma(y)} = \frac{31250.96}{166.25 * 220.46} = 0.8527$ Ce coefficient est toujours compris entre -1 et 1 ; il est assez proche de 1 ce qui valide l'existence d'une corrélation linéaire entre les variables.

5. On trouve : $\hat{y} = \hat{a}x + \hat{b} = 1.1307x - 140.9229$ avec
$$\hat{a} = \frac{Cov(x; y)}{V(x)}$$
$$\hat{b} = \bar{y} - \hat{a}\bar{x}$$
6. \hat{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, \hat{b} n'a pas de sens, il donnerait le nombre de spectateurs pour un film projeté dans aucune salle...
7. cf graphique.
8. $\hat{y}(750) = 1.1307 * 750 - 140.9229 = 707.1$
9. $R^2 = 0.8527^2 = 0.7271$; $R^2 = \frac{SCE}{SCT}$ ce coefficient donne la part de la variation totale expliquée par le modèle, ici 72.71 %

10. L'équation de l'analyse de la variance est : **SCT** = **SCE** + **SCR** 27637,66864 48602,31101
SCT = $nV(y) = 13 * 48602.61 = 631833.93$, **SCE** = $R^2 * \text{SCT} = 0.7271 * 631833.93 = 459406.45$ et donc
SCR = **SCT** - **SCE** = $631833.93 - 459406.45 = 172427.48$
11. $e_7 = y_7 - \hat{y}_7 = 340.5 - (1.1307 * 494 - 140.9229) = -77.14$; ce résidu est négatif ; il mesure la différence entre la valeur observée y_7 et la valeur estimée par le modèle \hat{y}_7 ; ce résidu est négatif car le point M_7 du nuage est situé sous la droite de régression.

