



# SUJET D-CORRIGE PARTIEL DE STATISTIQUE

L1-ECO

Janvier 2012

C

## 1 EXERCICE-1

Salaires	ni	fi	ficc	xi	Ai	di	ni xi	ni xi <sup>2</sup>	qi	qicc	Si
[0;10[	40	0,3125	0,3125	5	10	4	200	1000	0,1242	0,1242	0,0194
[10;15[	50	0,3906	0,7031	12,5	5	10	625	7812,5	0,3882	0,5124	0,1243
[15;20[	25	0,1953	0,8984	17,5	5	5	437,5	7656,25	0,2717	0,7842	0,1266
[20;30[	10	0,0781	0,9766	25	10	1	250	6250	0,1553	0,9394	0,0673
[30;35[	3	0,0234	1,0000	32,5	5	0,6	97,5	3168,75	0,0606	1,0000	0,0227
	128	1					1610	25887,5	1,0000		0,3604
Moyenn	12,58										
V	44,04										
Ecart-ty	6,64										
Indice de	0,2791										

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 :

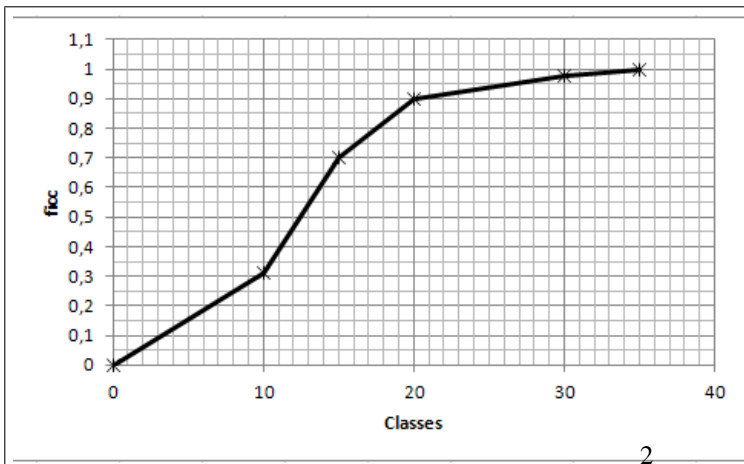
$$\frac{0,7031 - 0,3125}{15 - 10} = \frac{0,50 - 0,3125}{Q_2 - 10} \text{ soit } Q_2 - 10 = 5 * \frac{0,50 - 0,3125}{0,7031 - 0,3125} \text{ soit } Q_2 = 10 + 5 * \frac{0,50 - 0,3125}{0,7031 - 0,3125} = 12,40 \text{ Il y a donc 50\% de la population qui ont un salaire inférieur ou égal à 12,4 K€}$$

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

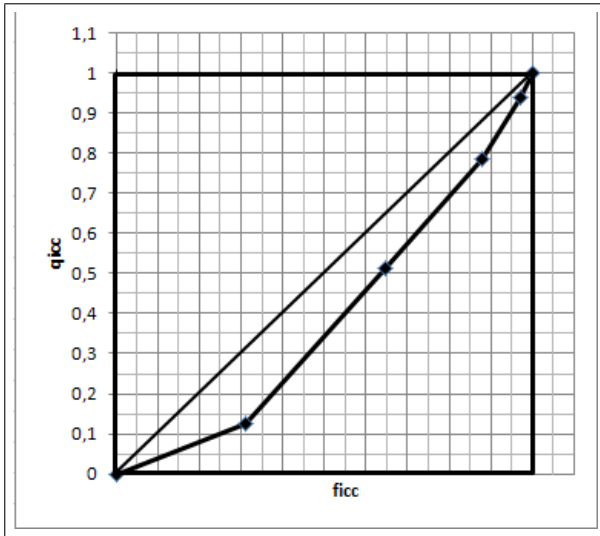
La classe modale est celle de plus grande densité, c'est-à-dire la classe [10; 15[ et le mode est calculé en considérant les

classes encadrant classe modale, ce qui donne avec les notations du cours :  $\begin{cases} x_1 = 10 \\ x_2 = 15 \end{cases}, \begin{cases} h = 10 \\ h_1 = 4 \text{ et } h_2 = 5 \end{cases}$

$\begin{cases} k_1 = h - h_1 = 6 \\ k_2 = h - h_2 = 5 \end{cases}$  . et pour conclure :  $M_o = \frac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \frac{5 * 10 + 6 * 15}{9} = 15,56$  On trouve :  $M_o \geq \bar{x} \geq M_e$ , ce qui indique une série asymétrique étalée à gauche.



4.



5. On calcule l'aire de concentration :  $A_C = 0.5 - \sum S_i = 0.5 - 0.3604 = 0.1396$ . On obtient alors l'indice de Gini :  $2 * \text{Aire de concentration}$  soit  $I_G = 2 * 0.1396 = 0.2792$

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

## 2 EXERCICE-2

1.

MOY	291,00	193,89
V	25605,00	45822,87
ET	160,02	214,06

2. cf tableau.

3. La covariance peut se calculer avec la formule : on obtient avec la calculatrice :  $\sum x_i y_i = 1192264.8$  soit en remplaçant :  $Cov(x; y) = \frac{1}{n} \sum x_i y_i - \bar{x} * \bar{y} = \frac{1197704.8}{14} - 291 * 193.89 = 29128.35$

4.  $r = \frac{Cov(x; y)}{\sigma(x) \sigma(y)} = \frac{29128.35}{160.02 * 214.06} = 0,8504$  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 :  $\hat{y} = \hat{a}x + \hat{b} = 1,1376x - 137,1405$  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 1138. 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.  $\hat{y}(750) = 1.1376 * 750 - 137.1405 = 716.06$

9.  $R^2 = 0.8504^2 = 0.7232$ ;  $R^2 = \frac{SCE}{SCT}$  ce coefficient donne la part de la variation totale expliquée par le modèle, ici 72.32 %

SCT	641520,21
sce	463562,5
SCR	177957,71

10. L'équation de l'analyse de la variance est :  $\mathbf{SCT} = \mathbf{SCE} + \mathbf{SCR}$

$\mathbf{SCT} = \mathbf{nV}(y) = 14 * 45822.87 = 641520.2$ ,  $\mathbf{SCE} = R^2 * \mathbf{SCT} = 0.7232 * 641520.2 = 463947.4$  et donc  
 $\mathbf{SCR} = \mathbf{SCT} - \mathbf{SCE} = 177957.71$

11.  $e_7 = y_7 - \hat{y}_7 = 340.5 - (1.1376 * 494 - 137.1405) = -84.33$  ; 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.