

SUJETC CORRIGE EXAMEN DE STATISTIQUE

LIECO

Jnavier 2012

1 EXERCICE-1

Salaires (en K€)	ni	fi	ficc	xi	Ai	di	ni xi	ni xi ²	qi	qicc	Si
[0;10[20	0,1852	0,1852	5	10	2	100	500	0,0662	0,0662	0,0061
[10;15[50	0,4630	0,6481	12,5	5	10	625	7812,5	0,4139	0,4801	0,1265
[15;20[25	0,2315	0,8796	17,5	5	5	437,5	7656,25	0,2897	0,7699	0,1447
[20;30[10	0,0926	0,9722	25	10	1	250	6250	0,1656	0,9354	0,0789
[30;35[3	0,0278	1,0000	32,5	5	0,6	97,5	3168,75	0,0646	1,0000	0,0269
	108	1					1510	25387,5	1,0000		0,3831
Moyenne	13,98										
V	39,59										
Ecart-type	6,29										
Indice de Gini	0,2338										

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 :

$$\frac{0,6481 - 0,1852}{15 - 10} = \frac{0,50 - 0,1852}{Q_2 - 10} \text{ soit } Q_2 - 10 = 5 * \frac{0,50 - 0,1852}{0,6481 - 0,1852} \text{ soit } Q_2 = 10 + 5 * \frac{0,50 - 0,1852}{0,6481 - 0,1852} = 13,40 \text{ Il y a donc 50\% de la population qui ont un salaire inférieur ou égal à 13,4 K€}$$

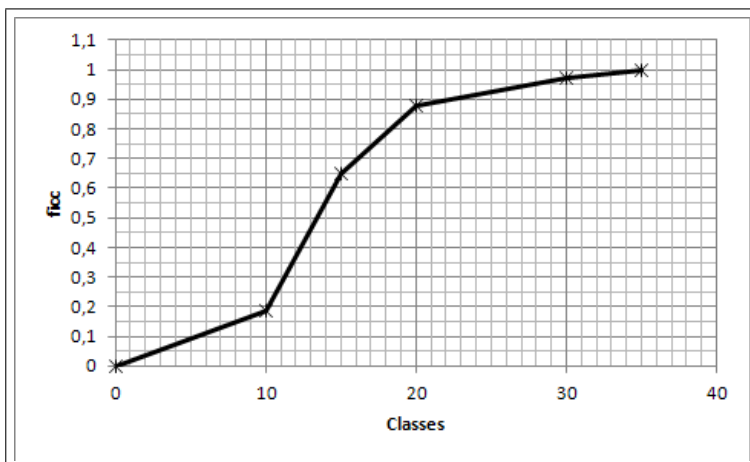
3. Pour déterminer le mode on peut utiliser indifféremment la densité ou les effectifs corrigés (ils sont proportionnels) ; on a utilisé ici les densités.

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

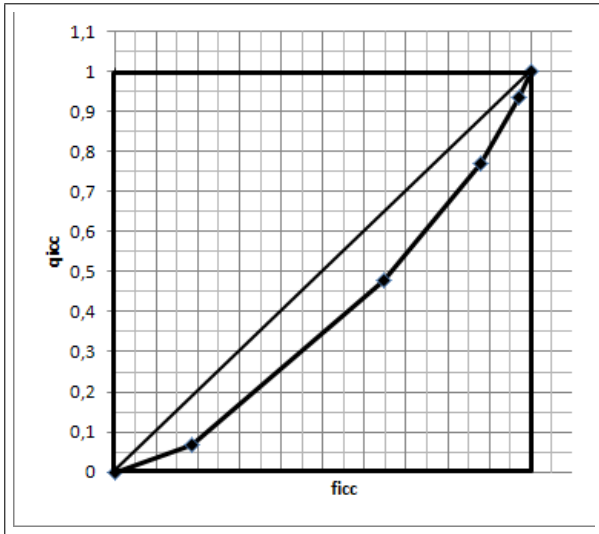
classes encadrant classe modale, ce qui donne avec les notations du cours : $\begin{cases} x_1 = 10 \\ x_2 = 15 \end{cases}$, $\begin{cases} h = 10 \\ h_1 = 2 \text{ et } h_2 = 5 \end{cases}$

$$\begin{cases} k_1 = h - h_1 = 8 \\ k_2 = h - h_2 = 5 \end{cases} \text{ et pour conclure : } M_o = \frac{k_2 x_1 + k_1 x_2}{k_2 + k_1} = \frac{5 * 10 + 8 * 15}{13} = 13,07 \text{ On trouve : } M_o \leq M_e \leq \bar{x}, \text{ ce}$$

qui indique une série asymétrique étalée à droite.



4.



5. On calcule l'aire de concentration : $A_C = 0.5 - \sum S_i = 0.5 - 0.3831 = 0.1169$. On obtient alors l'indice de Gini : $2 * \text{Aire de concentration}$ soit $I_G = 2 * 0.1169 = 0.2338$
- Le coefficient de Gini est plus proche de 0 que de 1, la concentration est faible.

2 EXERCICE-2

1.

MOY	289,57	193,18
V	25760,10	45879,35
ET	160,50	214,19

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 : $Cov(x; y) = \frac{1}{n} \sum x_i y_i - \bar{x} * \bar{y} = \frac{1192264.8}{14} - 289.57 * 193.18 = 29222.64$
4. $r = \frac{Cov(x; y)}{\sigma(x) \sigma(y)} = \frac{29222.64}{160.5 * 214.19} : 0.8501$ 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 : $\hat{y} = \hat{a}x + \hat{b} = 1.1344x - 135.32$ 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 1134. 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. $\hat{y}(750) = 1.1344 * 750 - 135.32 = 715.48$

9. $R^2 = 0.8501^2 = 0.7226$; $R^2 = \frac{SCE}{SCT}$ ce coefficient donne la part de la variation totale expliquée par le modèle, ici 72.26 %

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

$SCT = nV(y) = 14 * 45879.35 = 642310.924$, $SCE = R^2 * SCT = 0.7226 * 642310.924 = 464133.873$ et donc $SCR = SCT - SCE = 178177.05$

11. $e_7 = y_7 - \hat{y}_7 = 340.5 - (1.1344 * 494 - 135.32) = -84.57$; 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.