Université de Carthage
Ecole Supérieure de la Statistique et de l'Analyse de l'Information à Tunis

Année Universitaire 2011-2012
Première Année

Examen de Statistique Descriptive

Session principale

Janvier 2012

Cours de Mme Ouaili-Mallek — Durée 1h30

(2 pages)

Exercice 1 *On dispose d'une série d'observations sur les variables x et y.*

x	3.5	5.82	4.92	6.16	6.29	7.14	9.21
y	1.5	2.2	3.17	5.26	5.97	7.06	8.33

1. On envisage un ajustement linéaire de y sur x. Calculer le coefficient de corrélation empirique.
2. Déterminer la droite des moindres carrés.
3. Evaluer l'ajustement obtenu.

Exercice 2 *Un importateur d'automobiles désire augmenter sa part de marché. Il étudie alors la répartition des* 500 *modèles commercialisés par ses concurrents en fonction du prix* y*, exprimé en milliers de dinars, et de la puissance fiscale* x*, exprimée en chevaux :*

x \ y	$[15, 30[$	$[30, 35[$	$[35, 40[$	$[40, 45[$	$[45, 50[$	$[50, 55[$	*total*
3-4	*22*	*8*					*30*
5-6	*36*	*41*	*20*	*13*			*110*
7-8	*12*	*36*	*68*	*50*	*22*	*22*	*210*
9-10			*12*	*26*	*32*	*30*	*100*
11-12				*6*	*14*	*20*	*40*
13-15					*2*	*8*	*10*
total	*70*	*85*	*100*	*95*	*70*	*80*	*500*

Les deux parties sont indépendantes.

1ère partie :

On sait préalablement que $\overline{y} = 39.3$, $\overline{x} = 7.67$ *et* $s_Y^2 = 87.26$.

1. *Calculer le moment simple d'ordre 2 de la puissance fiscale de l'ensemble des voitures étudiées. En déduire la variance de la variable* x.

2. *Calculer la covariance du prix et de la puissance fiscale.*

3. *Calculer la distance du khi 2. Interpréter.*

   (Indication : $\sum_{i=1}^{6}\sum_{j=1}^{6} n_{ij}x_iy_j = 157\ 992.5 \qquad \sum_{i=1}^{6}\sum_{j=1}^{6} \frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} = 2.319$).

2ème partie :

L'importateur décide d'axer son action sur les véhicules de 7 à 8 chevaux.

1. *Tracer l'histogramme de la distribution des prix des voitures de 7 à 8 chevaux.*

2. *Déterminer le mode et la médiane. Sachant que la distribution est unimodale, déduire la moyenne. Peut-on conclure la symétrie de la distribution?*

3. *On s'intéresse maintenant à l'inégalité de la répartition des prix. Représenter la courbe de Lorenz.*

4. *Calculer l'indice de Gini. Conclure.*

Corrigé de l'exercice 1 *:*

x	*3.5*	*5.82*	*4.92*	*6.16*	*6.29*	*7.14*	*9.21*
y	*1.5*	*2.2*	*3.17*	*5.26*	*5.97*	*7.06*	*8.33*

1. $r_{xy} = \dfrac{s_{xy}}{s_x s_y}$

   $s_x^2 = \frac{1}{7}\sum x_i^2 - \overline{x}^2 \qquad \sum x_i^2 = 283.642 \qquad \overline{x} = 6.149$

   $s_x^2 = \frac{1}{7} * 283.642 - 6.149^2 = 2.710 = (1.646\,2)^2$

   $s_y^2 = \frac{1}{7}\sum y_i^2 - \overline{y}^2 \qquad \sum y_i^2 = 199.68 \qquad \overline{y} = 4.784$

   $s_y^2 = \frac{1}{7} * 199.68 - 4.784^2 = 5.\,639 \ = (2.374\,7)^2$

   $s_{xy} = \frac{1}{7}\sum x_i y_i - \overline{xy} \qquad \sum x_i y_i = 230.731$

   $s_{xy} = \frac{1}{7} * 230.731 - 6.149 * 4.784 = 3.\,545$

   $r_{xy} = \dfrac{3.\,544\,8}{1.646\,2 * 2.374\,7} = 0.907$

2. $\widehat{y} = \widehat{\alpha}x + \widehat{\beta}$

   $\widehat{\alpha} = \dfrac{s_{xy}}{s_x^2} = \dfrac{3.\,545}{2.710} = 1.308$

   $\widehat{\beta} = \overline{y} - \widehat{\alpha}\overline{x} = 4.784 - 1.308 * 6.149 = -3.259$

3. $\dfrac{s_Y^2}{s_y^2} = \dfrac{\widehat{\alpha}^2 s_x^2}{s_y^2} = \dfrac{1.308^2 * 2.710}{5.\,639} = 0.822$

   La variance expliquée par la régression constitue une forte proportion de la variance totale.

   L'ajustement est donc de bonne qualité.

Corrigé de l'exercice 2 *:*

1ère partie

1. $m_{x\ 2} = \frac{1}{n}\sum n_{i.}c_i^2$

   $m_{x\ 2} = \frac{1}{500}\left(3.5^2 * 30 + 5.5^2 * 110 + 7.5^2 * 210 + 9.5^2 * 100 + 11.5^2 * 40 + 14^2 * 10\right)$

   $m_{x\ 2} = 63.565$

   $s_x^2 = m_{x\ 2} - \overline{x}^2 = 63.565 - 7.67^2 = 4.736\,1$

2. $s_{xy} = \frac{1}{n}\sum\limits_{i=1}^{6}\sum\limits_{j=1}^{6} n_{ij}x_i y_j - \overline{x}\ \overline{y} = \frac{1}{500} * 157992.5 - 7.67 * 39.3 = 14.554$

3. $D^2 = n\left(\sum\limits_{i=1}^{6}\sum\limits_{j=1}^{6}\frac{n_{ij}^2}{n_{i.}n_{.j}} - 1\right) = 500{*}1.319 = 659.5 >> 0$. *On peut donc conclure avec très peu de risques que les variables puissance fiscale et prix des voitures ne sont pas indépendantes. On a* $D^2 \leq \min\left(n\left(k-1\right), n\left(p-1\right)\right) = 400 * 5 = 2000 >> D^2$. *On ne peut donc pas envisager de liaison fonctionnelle*

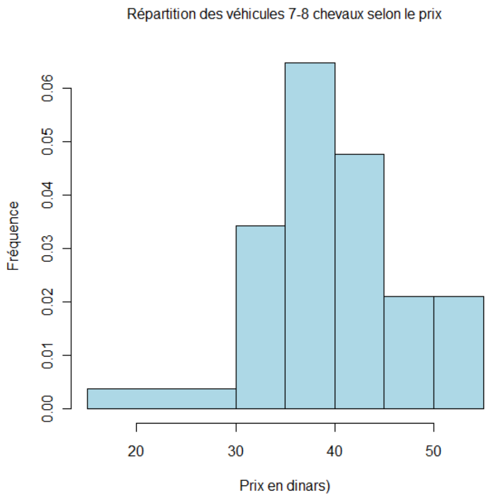

Figure 1:

2ème partie

1. *Distribution des prix des voitures de 7 à 8 chevaux.*

x	$[15,30[$	$[30,35[$	$[35,40[$	$[40,45[$	$[45,50[$	$[50,55[$	*total*
n	12	36	68	50	22	22	210
n_i^c	4	36	68	50	22	22	$-$
f_i	0.057	0.171	0.324	0.238	0.105	0.105	1
$F(e_i)$	0.057	0.228	0.552	0.79	0.895	1	
$f_i c_i$	1.283	5.558	12.15	10.115	4.988	5.513	39.607
$q(e_i)$	0.032	0.173	0.479	0.735	0.861	1	

Histogramme :

2. *Classe modale :* $[35,40[$

$$M_O = 35 + 5 * \frac{68-36}{(68-36)+(68-50)} = 38.2\ K\ DT$$

Classe médiane : $[35,40[$

$$M_e = 35 + 5 * \frac{0.5-0.228}{0.552-0.228} = 39.2\ K\ DT$$

$$\overline{x} \simeq \frac{3M_e - M_O}{2} = \frac{3*39.2-38.2}{2} = 39.7\ K\ DT$$

On a $M_O < M_e < \overline{x}$*. Il s'agit donc d'une distribution asymétrique conformément à l'histogramme.*

3. *Courbe de Lorenz:*

x	$[15,30[$	$[30,35[$	$[35,40[$	$[40,45[$	$[45,50[$	$[50,55[$	*total*
$F(e_i)$	0.057	0.228	0.552	0.79	0.895	1	
$q(e_i)$	0.032	0.173	0.479	0.735	0.861	1	

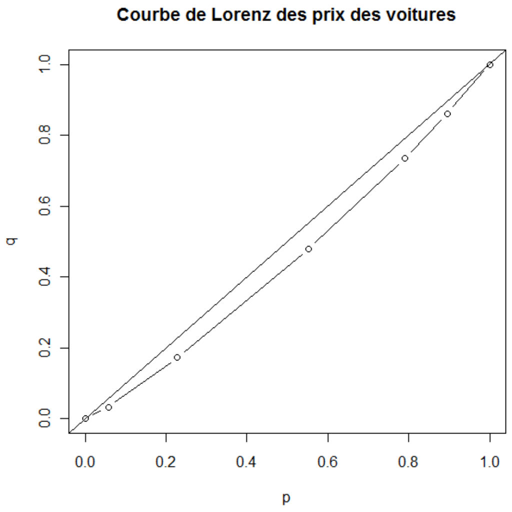

Figure 2:

4. $G = 1 - \sum_{i=1}^{6} \left(p\left(e_i\right) - p\left(e_i - 1\right)\right)\left(q\left(e_i\right) + q\left(e_{i-1}\right)\right)$

$$\begin{aligned} G = \; & 1 - [(0.057 * 0.032) + (0.228 - 0.057) * (0.173 + 0.032) \\ & + (0.552 - 0.228) * (0.479 + 0.173) + (0.79 - 0.552) * (0.735 + 0.479) \\ & + (0.895 - 0.79) * (0.861 + 0.735) + (1 - 0.895) * (1 + 0.861)] \\ G = \; & 0.1 < 0.2 \end{aligned}$$

La distribution est peu inégalitaire