



Université de Poitiers Département de Mathématiques

Statistique descriptive, 1er semestre, année univ. 2009-2010

Fiche 3

Statistiques descriptives bivariées

Exercice 1

On désigne par $x = \{x_1, x_2, \dots, x_n\}$ et $y = \{y_1, y_2, \dots, y_n\}$ des échantillons pondérés par les poids $p = (p_1, \dots, p_n)$. On suppose que les p_i sont strictement positifs et de somme 1. On munit \mathbb{R}^n du produit scalaire $\langle x; y \rangle = \sum_{i=1}^n p_i x_i y_i$. On définit le vecteur $\underline{1} = (1; \dots; 1)$.

1) Montrer que la covariance empirique $s_{x,y}$ vérifie

$$s_{x,y} := \sum_{i=1}^n p_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n p_i x_i y_i - \bar{x}\bar{y}.$$

On remarque que $s_{x,x} = s_x$ (variance empirique). En déduire que

$$s_x^2 = \sum_{i=1}^n p_i x_i^2 - (\bar{x})^2.$$

2) Vérifier que $s_{ax+b, cy+d} = a c s_{x,y}$. En particulier, la covariance est invariante par changement d'origine.

3) Montrer que $s_{x+y}^2 = s_x^2 + s_y^2 + 2s_{x,y}$ et que $|s_{x,y}| \leq s_x s_y$. Quel est le cas d'égalité? En déduire que la corrélation $\rho_{x,y}$ est comprise entre -1 et 1 et retrouver les cas d'égalités $\rho_{x,y} = \pm 1$.

4) Que peut-on dire pour les corrélations $\rho_{ax+b, cy+d}$ et $\rho_{x,y}$? En particulier, la corrélation est invariante par changement d'origine et d'échelle, et donc par l'opération de centrage-réduction des données.

Exercice 2

Dans le cadre d'une enquête visant à comparer selon certains critères différents aliments vendus dans les "fast food" ont été retenues les informations regroupées dans le tableau ci-dessous.

Observation i	Aliment	Poids(g) x_i	Prix (Euro) y_i
1	Sandwich végétarien	150	3,9
2	Sandwich parisien au jambon	92	3,4
3	Big Mac	193	5,7
4	Hamburger	90	2,9
5	Hot dog	135	3,8
6	Fallafel	241	6,0
7	Quiche lorraine	169	3,3
8	Pizza tomates-jambon	165	3,3

1) Représenter sur un graphe, le poids (en abscisse) et le prix (en ordonnées) des aliments proposés. Existe-t-il une relation entre les deux variables?

2) Donner l'équation de régression. Dessiner cette droite sur le graphe précédent. Commentaire?

3) Calculer la somme des résidus, les sommes des carrés expliqués par la régression et le coefficient de détermination R .