Université de Rennes 1
IFSIC - IUP1

Année 2005-2006

TD PS
Statistiques - Caractéristiques de 2 variables statistiques quantitatives

1. On fait subir à une variable statistique quantitative X pour laquelle on a n observations $X_1, X_2, \cdots, X_n$ une transformation linéaire ,i.e. $Y = aX + b$ et à une variable statistique quantitative V la transformation $W = cV + d$. Calculer la covariance entre Y et W $cov(Y, W)$ en fonction de la covariance entre X et V : $cov(X, V)$. Faites la même chose avec la corrélation $r_{Y,W}$ en fonction de $r_{X,V}$.

2. **Etude conjointe de la consommation et de la température**
   Nous reprenons les données du TD1 concernant la consommation d'électricité en fonction de la température.

   A. Consommation en fonction de la température en Farenheit

   Le nuage de points de la consommation en fonction de la température se trouve figure 1. La covariance entre la consommation d'électricité kwh et la température $tempf$ est égale à -323.69 .
   (a) Calculez le coefficient de corrélation entre kwh et $tempf$.
   (b) Donnez l'expression de la droite de régression de kwh sur $tempf$.
   (c) Quel est le pourcentage de la variance de la consommation expliquée par la régression précédente ?
   (d) Le graphe des résidus se trouve figure 2. Commentez-le rapidement. L'ajustement précédent vous parait-il acceptable ?

   B. Consommation et température en Celsius
   (a) Quelle est la covariance entre kwh et $tempc$ la température en Celsius ?
   (b) Quelle est valeur de la corrélation entre la consommation kwh et $tempc$.
   (c) Donnez l'équation de la droite de régression de la consommation kwh sur $tempc$ la température en Celsius.

   C. Transformation des variables
   Dans cette question, on transforme les deux variables kwh et $tempf$ en prenant leurs logarithmes. On a deux nouvelles variables $x = log(tempf)$ et $y = log(kwh)$ (tous les logarithmes sont en base 10). Vous trouverez figure 3 le nuage de points de y en fonction de x. Par ailleurs les données sur x et y figurent dans le tableau de données colonnes 6 et 7.
   (a) Pensez-vous qu'un ajustement linéaire entre x et y pourrait convenir ?
   (b) Le coefficient de corrélation entre x et y est égal à -0.901. Par ailleurs, on a les statistiques suivantes sur x et y :

variable	moyenne	écart-type
$x = log(tempf)$	1.71	0.134
$y = log(kwh)$	1.57	0.238

   Quelle est l'équation de la droite de régression de y sur x, c'est-à-dire de $log(kwh)$ sur $log(tempf)$.
   (c) En déduire la relation entre kwh et $tempf$.
   (d) Les résidus normés de la régression sont représentés figure 4. Examinez-les. Quel modèle choisirez vous entre le premier sur les variables kwh et $tempf$ ajusté à la question 1-3-1 ou le second (question 1-3-3) sur les variables transformées ?

3. On a mesuré les valeurs de deux variables X_1 et X_2 sur 10 individus et on a obtenu les résultats suivants :

	X_1	X_2
1	0.5	0
2	-0.1	1.2
3	-0.5	0.5
4	-0.3	0.1
5	0	2.5
6	1.6	-0.7
7	2.0	2.0
8	2.4	1.2
9	0.5	3.5
10	2.7	-0.9

(a) Calculez la matrice de covariance C entre X_1 et X_2

(b) Quelle est la combinaison linéaire $F = a_1X_1 + a_2X_2$ qui a la variance maximale, les coefficients de a_1 et a_2 vérifiant $a_1^2 + a_2^2 = 1$. Pour cela calculez la variance de F et exprimez la sous la forme $a'Ca$ où a' est le vecteur transposé de a et a a pour composantes a_1 et a_2.

Puis pour déterminer a_1 et a_2, dérivez la fonction appelée le lagrangien $L(a, \lambda)$ où $L(a, \lambda) = a'Ca - \lambda(a'a - 1)$ par rapport à a et à λ. Sachant que $\lambda_1 = 1.99$ est valeur propre de C, déterminer la seconde valeur propre.

(c) Sachant que le vecteur propre normé u correspondant à la plus grande valeur propre a pour coordonnées respectives $-0.5/\sqrt{1.25}$ et $1/\sqrt{1.25}$, déterminer F Comment pourrait-on représenter les points sur F. Calculer les coordonnées du premier individu sur F.

Cette méthode s'appelle l'analyse en composantes principales sur matrice de covariance (au programme de IUP2)

Fig. 1 – Consommation d'électricité en fonction de la température

Fig. 2 – Résidus de la régression consommation/température

Fig. 3 – log Consommation d'électricité en fonction du log de la température

Fig. 4 – Résidus de la régression log consommation/ log température