MATH 906 : Modélisation et diagnostic

Partiel : durée 1h30.

Mercredi 3 novembre 2010.

Exercice 1 (Formule de Lorentz)**.** La formule de Lorentz donne le poids (en kg) « idéal » d'un adulte en fonction de sa taille (en cm) :

- pour une femme : poids $= 0,6 \times$ taille $- 40$;
- pour un homme : poids $= 0,75 \times$ taille $- 62,5$.

Le tableau suivant donne le poids en kg noté y et la taille en cm notée x sur un échantillon de dix femmes adultes :

Taille en cm notée x	169	166	157	181	158	165	167	156	168	165
Poids en kg notée y	67	58	56	64	53	52	56	49	55	65

1. Calculer
   (a) La moyenne et la variance du caractère x sur l'échantillon observé ;
   (b) La moyenne et la variance du caractère y sur l'échantillon observé ;
   (c) La covariance des variables x et y sur l'échantillon observé.

2. On considère le modèle $y_i = \beta_1 + \beta_2 x_i + u_i$, $i = 1, \ldots, 10$, où les u_i sont indépendantes et identiquement distribuées suivant la loi gaussienne centrée de variance σ^2.
   (a) Calculer les coefficients $\hat{\beta}_1$ et $\hat{\beta}_2$ de la régression linéaire de y par rapport à x.
   (b) Préciser la valeur de l'estimateur $\hat{\sigma}^2$ de σ^2.
   (c) Donner un intervalle de confiance à 5% pour β_2.

Rappel. Si X suit la loi gaussienne centrée réduite $\mathbb{P}(|X| > 1,96) \simeq 5\%$, si Y suit la loi de Student à 8 degrés de liberté $\mathbb{P}(|Y| > 2,31) \simeq 5\%$.

   (d) Au niveau 5%, le facteur $0,6$ de la formule de Lorentz vous semble-t-il correct ?
   (e) Donner un intervalle de confiance à 95% pour le poids d'une femme mesurant 160 centimètres.

Exercice 2. On utilise le modèle de régression linéaire multiple

$$Y = \beta_1 + \beta_2 X + \beta_3 Z + U, \quad U \rightsquigarrow \mathcal{N}(0, \sigma^2 I_n).$$

On observe : $\|Y - \bar{y}\,\mathbf{1}\|^2 = 1680,8$, $\|\hat{Y} - \bar{y}\,\mathbf{1}\|^2 = 1504,4$ et $\hat{\sigma}^2 = 19,6$.

1. Quel est le coefficient de détermination R^2 du modèle ?
2. Tester l'hypothèse $H_0 : \beta_2 = \beta_3 = 0$ contre l'hypothèse alternative.

 Tournez S.V.P.

Exercice 3. On considère la régression linéaire multiple

$$Y_i = \beta_1 + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + U_i, \quad i = 1, \ldots, n,$$

où les U_i sont i.i.d. suivant la loi $\mathcal{N}(0, \sigma^2)$.

Les résultats numériques sont les suivants : $\|\hat{u}\|^2 = 24$,

$$^tXX = 100 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & -2 & -2 \\ 0 & -2 & 4 & -2 \\ 0 & -2 & 4 & -2 \end{pmatrix}, \quad \left(^tXX\right)^{-1} = \frac{1}{600} \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 5 & 4 & 3 \\ 0 & 4 & 5 & 3 \\ 0 & 3 & 3 & 3 \end{pmatrix}, \quad {}^tXY = \begin{pmatrix} 250 \\ -200 \\ 400 \\ 200 \end{pmatrix}.$$

1. Quel est le nombre d'observations n ?
2. Donner les valeurs des estimateurs $\hat{\beta}$ et $\widehat{\sigma}^2$.
3. (a) Donner un intervalle de confiance à 95% pour chacun des β_j.
   (b) Tester l'hypothèse $\beta_2 = 1$ contre $\beta_2 \neq 1$.
4. (a) Donner une région de confiance à 95% pour le couple (β_1, β_3).
   (b) Tester l'hypothèse $4\beta_1 - 3\beta_2 = 3$ contre l'hypothèse alternative au niveau 5%.