國立臺灣大學數學系

九十七學年度博士班入學考試試題

科目：迴歸分析

2008.04.25

1. Suppose that $Y_{i1}, \cdots, Y_{in_i} \overset{\text{i.i.d.}}{\sim} N(\mu_i, \sigma^2)$, $i = 1, \cdots, k$ $(k > 2)$, and Y_{ij}'s are mutually independent.

(1a) (4%) (3%) Characterize this problem via an appropriate linear model and explain the meaning of parameters in the considered model.

(1b) (8%) Find the maximum likelihood estimators of $(\mu_1, \cdots, \mu_k)$ and σ^2.'

(1c) (15%) Derive the likelihood ratio test statistic for the null hypothesis $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ versus the alternative hypothesis $H_A : \mu_i \neq \mu_j$ for some $i \neq j$.

2. (15%) Let $Y_i = m_i/n_i$ be the proportion of m_i subjects out of fixed n_i subjects that possess a certain property, $i = 1, \cdots, n$. Assume that

$$E[Y_i | x_{i1}, \cdots, x_{ip}] = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})}.$$

Write the weighted least squares estimation criterion for the parameters $\beta_0, \beta_1, \cdots, \beta_p$.

3. Consider a linear regression model $Y_t = \beta_0 + \beta_1 x_{t1} + \cdots + \beta_p x_{tp} + \varepsilon_t$, $t = 1, \cdots, n$, where $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ with $|\rho| < 1$ and u_t's $\overset{i.i.d.}{\sim} N(0, \sigma_u^2)$.

(3a) (10%) (7%) Let $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^T$. Write the probability density function of $\boldsymbol{Y}$ and find a consistent estimator of $Var(\boldsymbol{\varepsilon})$.

(3b) (4%) (4%) Write the Durbin-Watson test statistic and the testing procedure for the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_A : \rho \neq 0$.

4. (15%) Consider a random sample $\{(x_i^T, Y_{ij}) : i = 1, \cdots n, j = 1 \cdots, n_i\}$ with $Y_{ij} = f(\mathbf{x}_i) + \varepsilon_{ij}$, where $\mathbf{x}_i = (1, x_{i1}, \cdots, x_{ip})^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$, and ε_{ij}'s $\overset{i.i.d.}{\sim} N(0, \sigma^2)$. State the testing procedure for the null hypothesis $H_0 : Y = \mathbf{x}^T\boldsymbol{\beta} + \varepsilon$ versus the alternative hypothesis $H_A : Y = f(\mathbf{x}) + \varepsilon$ for $f(\mathbf{x}) \neq \mathbf{x}^T\boldsymbol{\beta}$.

5. Consider the segmented quadratic lines

$$\begin{cases} Y = \beta_{10} + \beta_{11}x + \beta_{12}x^2 + \varepsilon \text{ for } x \leq c \\ Y = \beta_{20} + \beta_{21}x + \beta_{22}x^2 + \varepsilon \text{ for } x > c \end{cases} \quad \text{with } \beta_{10} + \beta_{11}c + \beta_{12}c^2 = \beta_{20} + \beta_{21}c + \beta_{22}c^2.$$

(5a) (4%) (4%) Express the segmented quadratic lines via an appropriate linear regression model and explain the meaning of parameters in the considered model.

(5b) (7%) Based on the constructed linear regression model, state the corresponding hypothesis for $H_0 : \beta_{11} = \beta_{21}$.