國立臺灣大學數學系

九十六學年度上學期博士班資格考試題

科目：迴歸分析

2007.09

1. (20 pts) Consider data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$. Suppose that the postulated regression model is

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad 1 \le i \le n,$$

where

$$\epsilon_1, \cdots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

In fact, the true model is

$$y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 x_i^3 + e_i, \quad 1 \le i \le n,$$

where $e_1, \cdots, e_n \overset{i.i.d.}{\sim} N(0, \sigma_0^2)$. Let $\hat\beta_0$ and $\hat\beta_1$ denote the least squares estimators of β_0 and β_1, respectively.

   (a) (10 pts) Suppose that $n = 5m$ and $x_{5j-k} = -k + 2$ where $1 \le j \le m$ and $k = 1, 2, 3, 4, 5$. Determine $E(\hat\beta_0)$, $E(\hat\beta_1)$, and $Var(\hat\beta_1)$ as $m \to \infty$.

   (b) (10 pts) Suppose that $\alpha_0 = \alpha_1 = \alpha_2 = 0$, $n = 4m + 1$, and $x_i = (i-1)/m - 2$. Find $E(r_1)$, $E(r_{m+1})$, $E(r_{2m+1})$, and $Var(r_{2m+1})$ where $r_i = y_i - \hat\beta_0 - \hat\beta_1 x_i$.

2. (25 pts) Consider the multiple linear regression model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{i,p} + \epsilon_i, \quad 1 \le i \le n,$$

where

$$\epsilon_1, \cdots, \epsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

Let $\hat\beta_0, \ldots, \hat\beta_p$ denote the least squares estimators and $\hat Y_{x_1,\ldots,x_p} = \hat\beta_0 + \sum_{j=1}^p \hat\beta_j x_j$ which is the predicted value of Y at $(x_1, \ldots, x_p)$.

   (a) (10 pts) When $p = 1$, write

$$Y_i = \alpha_0 + \alpha_1 (x_{i1} - \bar x_{.1}) + \epsilon_i, \quad 1 \le i \le n,$$

   where $\bar x_{.1} = \sum_{i=1}^n x_{i1}/n$. Write $\hat\alpha_0$ and $\hat\alpha_1$ in terms of $\hat\beta_0$, $\hat\beta_1$, and $\bar x_{.1}$ and determine $Var(\hat Y_{\bar x_{.1}})$. Moreover, show that $Var(\hat Y_x) \ge \sigma^2/n$.

   (b) (15 pts) For general p, show that $Var(\hat Y_{x_1,\ldots,x_p}) \ge \sigma^2/n$ and determine whether $Var(\hat Y_{x_1,\ldots,x_p}) = \sigma^2/n$ for some $(x_1, \ldots, x_p)$.

3. (20 pts) Suppose

$$Y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i, \quad 1 \le i \le n,$$

where $\epsilon_1, \ldots, \epsilon_n \overset{iid}{\sim} N(0, \sigma^2)$. Write $\mathbf{X}$ as the $n \times p$ design matrix with the ith row $(x_{i0}, x_{i1}, \ldots, x_{i,p-1})$. Assume $\mathbf{X}$ is full rank. Let r be a nonzero positive constant.

   (a) (8 pts) Let $\hat\beta_{r0}, \hat\beta_{r1}, \hat\beta_{r2}, \ldots, \hat\beta_{r,p-1}$ denote the maximum likelihood estimates of $\beta_0, \beta_1, \beta_2, \ldots, \beta_{p-1}$ under the constraint $\beta_0^2 + \beta_1^2 + \cdots + \beta_{p-1}^2 \le \tau^2$. Determine $\hat\beta_{r0}, \hat\beta_{r1}, \hat\beta_{r2}, \ldots, \hat\beta_{r,p-1}$.

(b) (12 pts) When $p = 2$ and

$$\mathbf{X}^T\mathbf{X} = n \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

determine the maximum likelihood estimates of β_0 and β_1 under the constraint $|\beta_0| + |\beta_1| \leq \tau$.

4. (15 pts) Consider

$$y_i = \beta_0 + \beta_1 x_i + \beta_2(x_i - x)\delta_i + \epsilon_u, \quad i = 1, \ldots, n,$$

where $\delta_i = I(x_i > x)$, $I(\cdot)$ is the indicator function, and ϵ_i, $i = 1, \ldots, n$, are uncorrelated random errors. x is unknown but $x_{(1)} < x < x_{(n)}$. Devise a method to estimate β_2 and explain why it works.

5. (20 pts) Suppose that

$$Y_i = \beta_0 + \beta_1(x_{i1} - \bar{x}_{\cdot 1}) + \beta_2(x_{i2} - \bar{x}_{\cdot 2}) + \epsilon_i, \quad i = 1, \ldots, n.$$

Here $\bar{x}_{\cdot j}$ denotes the average of $x_{1j}, \ldots, x_{nj}$. Someone suggests to find the least squares estimates of β_0, β_1, and β_2 by the following two-stage procedure:

(a) Determine the least squares estimates of

$$Y_i = \beta_0 + \beta_1(x_{i1} - \bar{x}_{\cdot 1}) + \epsilon_i, \quad i = 1, \ldots, n.$$

(b) Regress the residual from (1) on $(x_{i1} - \bar{x}_{\cdot 2})$.

Does this two-stage procedure leads to the usual least squares estimates of β_0, β_1, and β_2? Justify your answer.