國立臺灣大學數學系

九十六學年度下學期博士班資格考試題

科目：迴歸分析

2008.02

1. (20 pts) Consider $\{(x_1, Y_1), \ldots, (x_n, Y_n)\}$. Suppose that the postulated regression model is

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

where

$$\epsilon_1, \cdots, \epsilon_n \overset{i.i.d.}{\sim} N(0,1).$$

In fact, the true model is

$$Y_i = x_i^2 + e_i, \quad 1 \le i \le n,$$

where $e_1, \cdots, e_n \overset{i.i.d.}{\sim} N(0,1)$. Let $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the least squares estimators of β_0 and β_1, respectively.

(a) (10 pts) Derive $E(r_i)$ and $Var(r_i)$ where r_i is the ith residual $Y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$.

(b) (10 pts) Suppose that $x_i = -1 + 2i/n$. Please describe what pattern you expect to see on the scatter plot of (x_i, r_i).

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

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

where $1 + p \le n$ and

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

Write $\mathbf{X} = (x_{ij})_{n\times p}$, and $\mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^T$. Let $\mathcal{L}$ denote the linear space spanned by $\mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_p$ where $\mathbf{1}_{n\times 1} = (1, \ldots, 1)^T$. It is known that the dimension of $\mathcal{L}$ is $p+1$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be an orthonormal basis for R^n such that $\mathbf{v}_1, \ldots, \mathbf{v}_{p+1}$ span $\mathcal{L}$. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$. Write $\mathbf{Y} = \sum_{i=1}^n a_i \mathbf{v}_i$, $\beta_0 \mathbf{1} + \sum_{j=1}^p \beta_j \mathbf{x}_j = \sum_{i=1}^n b_i \mathbf{v}_i$, and $(\epsilon_1, \ldots, \epsilon_n)^T = \sum_{i=1}^n c_i \mathbf{v}_i$.

(a) (10 pts) Determine a_i's, b_i's, and c_i's.

(b) (8 pts) Let e denote the residual vector after least-squares fit. Denote $S^2 = \mathbf{e}^T\mathbf{e}/(n - p - 1)$. Derive the distribution of S^2.

(c) (7 pts) Let $\hat{\beta}_k$ denote the least squares estimate of β_k. Derive the distribution of $\sqrt{n}(\hat{\beta}_k - \beta_k)/\sqrt{S^2}$.

3. (25 pts) Assume

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

where $\epsilon_1, \ldots, \epsilon_n$ are normally distributed independent random variables with mean 0 and variance σ^2. Propose a confidence interval for the ratio $\phi = -\beta_0/\beta_1$ which is valid for small sample size and give a justification or answer this question by the following suggestion proposed by Fieller (1940).

(a) (7 pts) Define $\delta = E[\bar{Y}]/E[\hat{\beta}_1]$ where $\bar{Y}$ is the average of $Y_1, \ldots, Y_n$ and $\hat{\beta}_1$ is the usual leat squares estimate.

(b) (9 pts) Determine $cov(\bar{Y}, \hat{\beta}_1)$ and $var(\bar{Y} - \delta\hat{\beta}_1)$.

(c) (9 pts) Determine the distribution of $(\bar{Y} - \delta\hat{\beta}_1)/S$ where S^2 is the usual estimate of σ^2 based on the residual. Then use it to give a confidence interval of ϕ.

4. (25 pts) Consider a simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where ϵ_i's, $1 \leq i \leq n$, are iid random variables with mean zero and variance σ^2. Since x_i's cannot be observed, Z_i's are observed instead. It is known that

$$Z_i = x_i + \eta_i,$$

where η_i's, $1 \leq i \leq n$, are iid random variables with mean zero and variance τ^2. Moreover, ϵ_i's and η_i's are independent. Suppose that a simple linear regression model

$$Y_i = \alpha_0 + \alpha_1 Z_i + e_i$$

is used to fit (Z_i, Y_i) by the least-squares method. The resulting estimates are denoted by $\hat{\alpha}_0$ and $\hat{\alpha}_1$.

(a) (10 pts) When $x_i = i/n$, does $\hat{\alpha}_1$ converge to a constant as $n \to \infty$? If it is, determine that constant.

(b) (15 pts) For each x_i, we observe

$$Z_{ij} = x_i + \eta_{ij}, \quad j = 1, 2.$$

Is it possible to get a consistent estimator of β_1? If it does, please give such a consistent estimator.