臺灣大學數學系
101 學年度上學期博士班資格考試題
科目：迴歸分析

2012.09.14

1. Consider the linear regression model $Y_i = \beta x_i + \varepsilon_i$, $i = 1, \ldots, n$, where ε_i's are independently distributed with mean 0 and variance $\sigma^2 x_i^2$.

(1a) (6%) (4%) Find the weighted least squares estimator, say, $\widehat{\beta}_w$ of β and the natural unbiased estimator of σ^2.

(1b) (10%) Show that $\widehat{\beta}_w$ is the best linear unbiased estimator of β.

2. Consider the linear regression model $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \delta z_i + \varepsilon_i$, $i = 1, \ldots, n$, where $z_i = 1_{\{i=1\}}$ and $\varepsilon_1, \ldots, \varepsilon_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$.

(2a) (8%) Derive the maximum likelihood estimator of $(\beta_0, \ldots, \beta_p, \delta)$.

(2b) (12%) State the testing procedure for the null hypothesis $H_0 : \delta = 0$ versus the alternative hypothesis $H_A : \delta \neq 0$.

3. (7%)(8%) Let $\widehat{\beta}_j$'s be the least squares estimators of β_j's based on the regression model $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i$, where ϵ_i's are uncorrelated with zero mean and constant variance σ^2, $i = 1, \ldots, n$. Moreover, let $e_{i(j)}$'s and d_{ij}'s denote the residuals computed based on the regression models $Y_i = \beta_{0j} + \sum_{l \neq j} \beta_{lj} x_{il} + \eta_i$ and $x_{ij} = \gamma_0 + \sum_{l \neq j} \gamma_l x_{il} + \xi_i$, respectively. Show that $\widehat{\beta}_j = r_{Y,x_j} \sqrt{\sum_{i=1}^n e_{i(j)}^2 / \sum_{i=1}^n d_{ij}^2}$ and $Var(\widehat{\beta}_j) = \sigma^2 / \sum_{i=1}^n d_{ij}^2$, where r_{Y,x_j} is the partial correlation between Y and x_j.

4. (7%) (8%) Consider a linear regression model $\boldsymbol{Y} = \tilde{\mathbb{Z}}_{(s)} \boldsymbol{\beta}_{(s)} + \boldsymbol{\varepsilon}$, where $\tilde{\mathbb{Z}}_{(s)} = (\boldsymbol{1}, \mathbb{Z}_{(s)})$ with $\mathbb{Z}_{(s)}$ being a $n \times p$ standardized covariate matrix, $\boldsymbol{\beta}_{(s)} = \begin{pmatrix} \gamma_0 \\ \boldsymbol{\beta}_{(0)} \end{pmatrix}$, and $\boldsymbol{\varepsilon} \sim (0, \sigma^2 I_n)$. Assume that an approximate linear relationship exists among the column vectors of $\mathbb{Z}_{(s)}$. Write the ridge regression estimator, say $\hat{\boldsymbol{\beta}}_{(0)R}$, for $\boldsymbol{\beta}_{(0)}$, and show that $Var(\hat{\boldsymbol{\beta}}_{(0)}) - Var(\hat{\boldsymbol{\beta}}_{0R})$ is at least positive semi-definite.

5. (15%) Let $Y_{ij} = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_{ij}$, where $\mathbf{x}_i = (1, x_{i1}, \cdots, x_{ip})^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)^T$, and $\varepsilon_{ij} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$. 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 $f(\mathbf{x}) \neq \mathbf{x}^T \boldsymbol{\beta}$.

6. (15%) Assume that

$$E[Y_{ij} | x_{i1}, \ldots, x_{ip}] = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip})} \text{ and } corr(Y_{ij_1}, Y_{ij_2} | x_{i1}, \ldots, x_{ip}) = \rho$$

for $j_1 \neq j_2$. Let $Y_i = \sum_{j=1}^m Y_{ij}/m$ be the proportion of fixed m subjects that possess a certain property, $i = 1, \ldots, n$. Write the weighted least squares estimation criterion for the parameters $\beta_0, \beta_1, \ldots, \beta_p$.