臺灣大學數學系

100 學年度上學期博士班資格考試題

科目：迴歸分析 2011.09.16

1. (15%) Consider the multiple linear regression model $Y = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\mathbb{X}$ being a $n \times (p+1)$ full rank matrix and $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 I_n)$. Let $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\sigma}^2$ be the maximum likelihood estimators of $\boldsymbol{\beta}$ and σ^2. Derive the distribution of $n\widetilde{\sigma}^2/\sigma^2$.

2. Consider the 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 u_t's $\overset{i.i.d.}{\sim} N(0, \sigma_u^2)$.

(2a) (10%) Find the estimated generalized least squares estimator of $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)$.

(2b) (10%) State the testing procedure for the null hypothesis $H_0 : \rho = 0$ against the alternative hypothesis $H_A : \rho \neq 0$.

3. (10%) State the stepwise variable selection procedure under the mechanism of a linear regression model.

4. (15%) Let $Y_i^{(\lambda)}$ be the Box and Cox transformation of the positive response Y_i, $i = 1, \cdots, n$. Moreover, assume that $Y_i^{(\lambda)} = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$, where ε_i's $\overset{\text{i.i.d.}}{\sim} N(0, \sigma^2)$. State the testing procedure for the null hypothesis $H_0 : \lambda = \lambda_0$ versus the alternative hypothesis $H_0 : \lambda \neq \lambda_0$.

5. Consider the linear regression model $Y = \mathbb{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbb{X}$ is a $n \times p$, full rank covariate matrix, $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector, and $\boldsymbol{\varepsilon} \sim (0, \sigma^2 I_n)$.

(5a) (10%) Find the minimizer, say, $\widetilde{\boldsymbol{\beta}}$ of the sum of squares $(Y - \mathbb{X}\boldsymbol{\beta})^T (Y - \mathbb{X}\boldsymbol{\beta})$ with constraint $\mathbb{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$, where $\mathbb{C}$ is a $J \times p$ full rank matrix.

(5b) (10%) Assume that $\mathbb{C}\boldsymbol{\beta} = \boldsymbol{\gamma}$. Show that $Var(\widehat{\boldsymbol{\beta}}) - Var(\widetilde{\boldsymbol{\beta}})$ is at least positive semidefinite, where $\widehat{\boldsymbol{\beta}}$ is the ordinary least squares estimator of $\boldsymbol{\beta}$.

6. Consider the linear regression model $Y = \tilde{\mathbb{Z}}_{(s)}\boldsymbol{\beta}_{(s)} + \boldsymbol{\varepsilon}$, where $\tilde{\mathbb{Z}}_{(s)} = (\mathbf{1}, \mathbb{Z}_{(s)})$ 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 $\mathbb{Z}_{(s)}$.

(6a) (10%) Show that $(R^{-1})_{jj} = VIF_j$, $j = 1, \cdots, p$, where $R = (\mathbb{Z}_{(s)}^T \mathbb{Z}_{(s)})/n$ and VIF is the variance inflation factor.

(6b) (10%) Let $\widehat{\boldsymbol{\beta}}_{(0)}$ and $\widehat{\boldsymbol{\beta}}_{(0)R}$ denote the least squares estimator and the ridge regression estimator of $\boldsymbol{\beta}_{(0)}$, respectively. Show that $Var(\widehat{\boldsymbol{\beta}}_{(0)}) - Var(\widehat{\boldsymbol{\beta}}_{0R})$ is at least positive semidefinite.