Linear Models Prelim Exam

12-4pm, Tuesday, August 20, 2013

1. Let $x = (X_1, X_2)^T \sim N_2(\mu 1_2, \Sigma)$, where $\Sigma = (1-\rho)I_2 + \rho J_2$. Let $Q_1 = (X_1 - X_2)^2$ and $Q_2 = (X_1 + X_2)^2$.
   (a) Show that $Q_1 / 2(1-\rho)$ has a χ^2 distribution.
   (b) Prove that Q_1 and Q_2 are distributed independently.

2. Consider a linear model, $\mathbf{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathbf{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\sigma}^2\mathbf{I_n}$, where r(X)=p. Show that $\sum_{i=1}^{n} \mathbf{Var}(\hat{y}_i) = \mathbf{p}\,\boldsymbol{\sigma}^2$ where $\hat{y}_i$ is the predicted value of $\mathbf{y_i}$ for i=1,…, n.

3. Let $\boldsymbol{Y_{ij}} = \boldsymbol{\mu} + \boldsymbol{\tau_i} + \boldsymbol{\epsilon_{ij}}$, and $\boldsymbol{\epsilon_{ij}} \sim i.i.d.\,\boldsymbol{N}(\mathbf{0}, \boldsymbol{\sigma}^2)$, i=1, …, a, j=1, …, n. For $\boldsymbol{\beta} = (\boldsymbol{\mu}, \boldsymbol{\tau_1}, \ldots, \boldsymbol{\tau_a})^T$, define
   $$\boldsymbol{c^T\beta} = \left[\sum_{i=1}^{l} \boldsymbol{\tau_i} - \boldsymbol{l} \cdot \boldsymbol{\tau_{l+1}}\right] / \sqrt{\boldsymbol{l}(\boldsymbol{l}+\mathbf{1})}.$$
   (a) Show that $\boldsymbol{\mu}$ is not estimable function.
   (b) Verify $\boldsymbol{c^T\beta}$ is estimable.
   (c) Construct a 95% confidence interval for $\boldsymbol{c^T\beta}$.

4. The multivariate linear regression model is $\underset{(n\times m)}{\mathbf{Y}} = \underset{(n\times(r+1))}{\mathbf{Z}}\ \underset{((r+1)\times m)}{\boldsymbol{\beta}} + \underset{(n\times m)}{\boldsymbol{\varepsilon}}$ with $E(\boldsymbol{\varepsilon}_{(i)}) = \mathbf{0}$ and $Cov(\boldsymbol{\varepsilon}_{(i)}, \boldsymbol{\varepsilon}_{(k)}) = \sigma_{ik}\mathbf{I}$, i, k = 1, 2, ..., m and the rank of the design matrix $\mathbf{Z}$, rank($\mathbf{Z}$)= r+1 < n. The m observations on the j^{th} trial have covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{ij}\}$, but observations from different trials are uncorrelated. Show that
   (a) The least square estimator $\hat{\boldsymbol{\beta}} = \left[\hat{\boldsymbol{\beta}}_{(1)} \ \vdots\ \hat{\boldsymbol{\beta}}_{(2)} \ \vdots\ \cdots \ \vdots\ \hat{\boldsymbol{\beta}}_{(m)}\right]$ satisfies $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $Cov(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(\mathbf{Z}'\mathbf{Z})^{-1}$, i, k = 1,2,..., m.
   (b) The residuals $\hat{\boldsymbol{\varepsilon}} = \left[\hat{\boldsymbol{\varepsilon}}_{(1)} \ \vdots\ \hat{\boldsymbol{\varepsilon}}_{(2)} \ \vdots\ \cdots \ \vdots\ \hat{\boldsymbol{\varepsilon}}_{(m)}\right] = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$ satisfy $E(\hat{\boldsymbol{\varepsilon}}) = \mathbf{0}$ and $E(\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}) = (n-r-1)\boldsymbol{\Sigma}$.
   (c) $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varepsilon}}$ are uncorrelated.

5. Let $\mathbf{X}_1$, $\mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4$ and $\mathbf{X}_5$ be independent and identically distributed random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

(i) Find the mean vector and covariance matrices for each of the two linear combinations of random vectors $\frac{1}{5}\mathbf{X}_1 + \frac{1}{5}\mathbf{X}_2 + \frac{1}{5}\mathbf{X}_3 + \frac{1}{5}\mathbf{X}_4 + \frac{1}{5}\mathbf{X}_5$ and $\mathbf{X}_1 - \mathbf{X}_2 + \mathbf{X}_3 - \mathbf{X}_4 + \mathbf{X}_5$ in terms of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

(ii) Obtain the covariance between the above two linear combinations of random vectors.

6. Suppose the observed data Y_i has a binomial distribution denoted as $Bin(n_i, \pi_i)$. Let $y_i = Y_i / n_i$ as a sample proportion of success for n_i trials and record a single predictor variable X_i along with the n_i trials, $\mathrm{i} = 1, 2, ..., \mathrm{N}$. A logistic regression model is fitted to the data as

$$\pi_i = \frac{\exp(\alpha + \beta X_i)}{1 + \exp(\alpha + \beta X_i)}$$

(i) Show that $\frac{\partial l}{\partial \alpha} = \sum_{i=1}^{N} n_i (y_i - \pi_i)$ and $\frac{\partial l}{\partial \beta} = \sum_{i=1}^{N} n_i (y_i - \pi_i) X_i$, where l is the logarithm of likelihood function with data $\{(Y_i, X_i, n_i), \mathrm{i} = 1, ..., \mathrm{N}\}$.

(ii) Show the maximum likelihood estimator of α and β using Fisher Scoring algorithm.