第一卷

Qualifying Exam (Linear Models)

3/5/03

1. Let $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\boldsymbol{\Sigma} = 2\mathbf{I}_3 + \mathbf{J}_3 = 2\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

   (a) Find the distribution of $\mathbf{Y}'\mathbf{A}\mathbf{Y}$, where $\mathbf{A} = \mathbf{I}_3 - \frac{1}{3}\mathbf{J}_3$.

   (b) Let $U_1 = (Y_1 + Y_2 + Y_3 - 6)^2$. Find the distribution of U_1 and find a quadratic form U_2 such that U_1 and U_2 are independent.

   (c) Find $P(U_1/U_2 > 1)$.

2. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is $n \times p$ of rank p and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

   (a) Find two matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ such that both $\mathbf{H}_1\boldsymbol{\beta} = \mathbf{0}$ and $\mathbf{H}_2\boldsymbol{\beta} = \mathbf{0}$ can be used to test the null hypothesis $\beta_1 = \beta_2 = 2\beta_3$, and show that how they are related.

   (b) Prove that the test statistic using $\mathbf{H}_1$ equals the test statistic using $\mathbf{H}_2$. What is the test statistic distributed?

第二卷

1. **(60%)** Let R_k and R_{k-1} be the multiple correlation coefficient of $\boldsymbol{y}$ with respectively $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ and $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{k-1}$. Suppose $\boldsymbol{x}_1 = \boldsymbol{J}$.

(1) Show that the t-statistic for testing $H_0 : \beta_k = 0$ vs. $H_1 : \beta_k \neq 0$ in the model $\boldsymbol{y} = \sum_1^k \beta_j \boldsymbol{x}_j + \boldsymbol{\epsilon}$ is

$$t = \frac{\hat{\beta}_k}{\sqrt{s^2/||\boldsymbol{x}_k^{\perp}||^2}},$$

and

$$t^2 = \frac{(ESS_{k-1} - ESS_k)}{ESS_k}(n-k),$$

where ESS denotes the error sum of squares.

(2) Show the partial correlation coefficient of $\boldsymbol{y}$ with $\boldsymbol{x}_k$ with the effects of $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{k-1}$ removed is

$$r = \frac{t/\sqrt{n-k}}{\sqrt{1 + t^2/(n-k)}}.$$

(3) Please show that $R_k^2 = R_{k-1}^2 + \frac{d}{1+d}(1 - R_{k-1}^2)$, where $d = t^2/(n-k)$.

(4) Show that

$$r^2 = \frac{R_k^2 - R_{k-1}^2}{1 - R_{k-1}^2},$$

and also comment on this result.

(5) Let $\hat{\boldsymbol{y}}_k$ be the predicted value of $\boldsymbol{y}$ based on all k independent variables in the model and let $\hat{\boldsymbol{y}}_{k-1}$ be the predicted value of $\boldsymbol{y}$ based on only $k-1$ independent variables. Show that $Var(\hat{\boldsymbol{y}}_k) - Var(\hat{\boldsymbol{y}}_{k-1}) \geq 0$.

2. **(25%)** Under the model $\boldsymbol{Y} = \boldsymbol{\theta} + \boldsymbol{\epsilon}$, for $\boldsymbol{\theta} \in V$, $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, please show that

(1) $\hat{\boldsymbol{\theta}} \sim N_n(\boldsymbol{\theta}, \boldsymbol{P}_V \sigma^2)$,

(2) $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{Y}} - \hat{\boldsymbol{\theta}}$ are independent random vectors, and

(3) $||\hat{\boldsymbol{Y}} - \hat{\boldsymbol{\theta}}||/\sigma^2 \sim \chi^2_{n-dim(V)}$.

3. **(15%)** Suppose that $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. If the Newton method is used with a starting value $\hat{\boldsymbol{\beta}}^0$, what are $\hat{\boldsymbol{\beta}}^1$, $\hat{\boldsymbol{\beta}}^2, \cdots$?

第三卷

1. Consider the following heteroskedastic model:
$Y_t = a + bx_t + e_t$ where e_t are independent, E $(e_t) = 0$ and $\text{Var}(e_t) = \sigma^2 x_t, t = 1, \ldots N$.

(a). Transform the Y_t's so that the response follows the usual constant-variance model.

(b). Compute the $X'X$ matrix for this transformed model.

(c). Suppose $x_t = t$. What is the behavior of this matrix when N is large?

(d). Is the least squares intercept estimator $\hat{a}$ consistent in this transformed model?

(Hint: |The harmonic series $\sum_{k=1}^{N} k^{-1} \approx \log N$).

(20%)

2. Consider the linear model $\underset{\sim}{y} = X\underset{\sim}{b} + \underset{\sim}{e} = X_0 \underset{\sim 0}{b} + X_1 \underset{\sim 1}{b} + \underset{\sim}{e}$, where X: $N \times p$, $X_0 : N \times p_0$ and $X_1 : N \times (p - p_0)$ are fixed (known) matrices and $\underset{\sim}{e} \sim N(0, \sigma^2 I_N)$. Let $r = \text{rank}(X) > r_0 = \text{rank}(X_0)$. Also, let $P_{X_0} = X_0(X_0'X_0)^- X_0'$ and $P_X = X(X'X)^- X'$. Consider the sums of squares $R(\underset{\sim 0}{b}) = SS_0 = \underset{\sim}{y}' P_{X_0} \underset{\sim}{y}$; $R(\underset{\sim 1}{b} \mid \underset{\sim 0}{b}) = SS_1 = \underset{\sim}{y}' P_X \underset{\sim}{y} - SS_0$ and $SSE = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}' P_X \underset{\sim}{y}$.

(a). Assume that $\underset{\sim 0}{b}$ and $\underset{\sim 1}{b}$ are fixed but unknown parameters.

(1). Give algebraic expressions (involving σ^2, $\underset{\sim}{b}$ and X) for $E(SS_0)$, $E(SS_1)$ and E(SSE).

(2). Specify completely the joint distribution of SS_0, SS_1 and SSE (properly normalized). Carefully, prove your answer. State any relevant results.

(b). Suppose $\underset{\sim 0}{b}$ is fixed but unknown; $\underset{\sim 1}{b} \sim N(0, \sigma_1^2 I_{p-p_0})$ and $\underset{\sim 1}{b}$ and $\underset{\sim}{e}$ are independent. Find an unbiased estimator of σ_1^2 that is a linear function of SS_1 and SSE. (You must show that your estimator is unbiased.)

(25%)

3. For the linear model $y = X\beta + \varepsilon$ with $\varepsilon \sim N(0, \sigma^2 I)$ and X being $n \times p$ of rank $r < p$, we can write $SSR = y'XGX'y$, where G is any generalized inverse of $X'X$.

   (a) Verify that $SSR/\sigma^2 \sim \chi^2[r(X), \beta' X' X \beta / (2\sigma^2)]$, that is, SSR/σ^2 has a noncentral χ^2 distribution with df $= r(X)$ and noncentrality parameter $\beta' X' X \beta / (2\sigma^2)$.

   (b) Show that XGX' in SSR is invariant to G, i.e., does not depend on which generalized inverse one uses.

(20%)

4. Consider the model $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, where X is an $N \times p$ matrix of rank $r < p < N$; $\mathbf{b}$ is a $p \times 1$ vector of unknown parameters; $\mathbf{e} \sim N(0, \mathbf{V})$ and $\mathbf{V}$ is nonsingular. Let $\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y}$ be a particular solution to the normal equations. Define $s^2 = \hat{\mathbf{e}}'\hat{\mathbf{e}}/(N - r)$ where $\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y}$ and $\hat{\mathbf{e}} = (\mathbf{I} - \mathbf{P}_X)\mathbf{y}$. Suppose $\lambda'\mathbf{b}$ is estimable.

(a). Give the distribution of $\lambda'\hat{\mathbf{b}}$.

(b). Show that if $\mathbf{VX} = \mathbf{XQ}$ for some matrix $\mathbf{Q}$, then $\lambda'\hat{\mathbf{b}}$ and $\hat{\mathbf{e}} = (\mathbf{I} - \mathbf{P}_X)\mathbf{y}$ are independent.

(c). Assume that $\mathbf{V} = \sigma^2(\mathbf{I} + \mathbf{P}_X)$ for some matrix $\mathbf{Q}$, Show that for some constant k, kt has a student's t-distribution where:

$$t = \frac{(\lambda'\hat{\mathbf{b}} - \lambda'\mathbf{b})}{[\lambda'(\mathbf{X}'\mathbf{X})^{+}\lambda s^2]^{1/2}}.$$

Find k.

(d). If $\mathbf{V} = \sigma^2\mathbf{I}$, give the BLUE for $\lambda'\mathbf{b}$.

(e). If $\mathbf{V} = \sigma^2(\mathbf{I} + \mathbf{P}_X)$, $\sigma^2 > 0$, is $\lambda'\hat{\mathbf{b}}$ the BLUE for $\lambda'\mathbf{b}$? Why or why not?

(35%)