卷一

Linear Models(Qualifying Exam)

9/8/2006

1. Consider the model $y_{ijk}=\mu+\tau_i+\varepsilon_{ij}$, $i=1,2,3,\ j=1,2,3.$
   (a) Write $\mathbf{X}$, $\mathbf{X}'\mathbf{X}$, $\mathbf{X}'\mathbf{y}$, and the normal equations.
   (b) What is the rank of $\mathbf{X}$ or $\mathbf{X}'\mathbf{X}$? Find a set of linearly independent estimable functions.
   (c) Define an appropriate side condition and find the resulting solution to the normal equations.
   (d) Show that $H_0:\tau_1=\tau_2=\tau_3$ is testable. Find $\hat{\boldsymbol\beta}'\mathbf{X}'\mathbf{y}=\mathrm{SS}(\mu,\tau)$ and $\hat{\boldsymbol\beta}_2'\mathbf{X}_2'\mathbf{y}=\mathrm{SS}(\mu)$.
   (e) Construct an analysis of variance table for the test of $H_0:\tau_1=\tau_2=\tau_3$.

2. Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from the bivariate normal distribution with parameters $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2$, and ρ. Find a constant K so that

$$T=K\frac{(\bar{X}-\bar{Y})-\delta}{\{\sum_{i=1}^{n}[(X_i-Y_i)-(\bar{X}-\bar{Y})]^2\}^{1/2}}\sim t_m(\theta).$$

Express m and θ as a function of the parameters and constant δ.
(Hint: Let $D_i=X_i-Y_i$. Express T as a function of the D_i.)

統模資格考

卷(三)

2006/9/8

1. Suppose the true model is $Y = X_p\beta_p + X_q\beta_q + \varepsilon$, and $\varepsilon \sim N_n(0, \sigma^2 I_n)$, where he dimension of X_p is $n \times p$ and X_q is $n \times q$. Moreover, the columns of X_p and X_q are linearly independent..

Let the subset model be $Y = X_p\beta_p + \varepsilon$.

(a) Let $\hat{\sigma}_S{}^2$ be the usual estimator for σ^2 under the subset model.

Is $\hat{\sigma}_S{}^2$ unbiased? If not, find its bias. (10%)

(b) Let $\hat{\beta}_{p,T}$ and $\hat{\beta}_{q,T}$ be the LSE of the true model, while $\hat{\beta}_{p,S}$ be the LSE of the subset model.

Show that $\hat{\beta}_{p,T} = \hat{\beta}_{p,S} - (X_p{}'X_p)^{-1}X_p{}'X_q\hat{\beta}_{q,T}$ (10%)

(Hint: If $M = \begin{bmatrix} A & B \\ B' & D \end{bmatrix}$, then $M^{-1} = \begin{bmatrix} A^{-1} + A^{-1}BWB'A^{-1} & -A^{-1}BW \\ -WB'A^{-1} & W \end{bmatrix}$,

where $W = (D - B'A^{-1}B)^{-1}$.)

(c) Let e_S be the residuals vector under the subset model. Regress e_S on X_q, and denote the LSE of slope as $\tilde{\beta}_q$. Is $\tilde{\beta}_q$ unbiased? What is the relationship between $\hat{\beta}_{q,T}$ and $\tilde{\beta}_q$? (15%)

(d) Derive the test statistic for testing $\beta_q = 0$ vs. $\beta_q \neq 0$, then write down its distribution, and specify its parameters. (15%)

(e) For an individual with $(\tilde{x}_p, \tilde{x}_q)$, we want to estimate $\eta = \tilde{x}_p\beta_p + \tilde{x}_q\beta_q$. Using the subset model, then $\hat{\eta}_S = \tilde{x}_p\hat{\beta}_{p,S}$, while under the true model $\hat{\eta}_T = \tilde{x}_p\hat{\beta}_{p,T} + \tilde{x}_q\hat{\beta}_{q,T}$. Find the condition that $\hat{\eta}_S$ has smaller mean square errors than $\hat{\eta}_T$. (10%)

2. Suppose the true model is $Y = X_p\beta_p + \varepsilon$, $\varepsilon \sim N_n(0, \sigma^2 I_n)$. The postulated model takes $E(Y) = X_p\beta_p + X_q\beta_q$. The dimension of X_p is $n \times p$ and X_q is $n \times q$. The columns of X_p and X_q are linearly independent.

(a) Show that LSE of the postulated model(i.e. $\hat{\beta}_{p,P}$ and $\hat{\beta}_{q,P}$) are unbiased for true parameters. (10%)

(b) Let $\eta = c'\beta_p$. Under the postulated model $\hat{\eta}_P = c'\hat{\beta}_{p,P}$, while under the true model $\hat{\eta}_T = c'\hat{\beta}_{p,T}$. What is $V(\hat{\eta}_P) - V(\hat{\eta}_T)$? (10%)

3. Let $Y_i = X_i\beta_i + \varepsilon_i$, where X_i's are $n_i \times p$ matrices, $\varepsilon_i \sim N_n(0, \sigma^2 I_n)$, i=1,2; and $\mathrm{Cov}(\varepsilon_1, \varepsilon_2) = 0$. Obtain a test statistic for testing $\beta_1 = \beta_2$ vs. $\beta_1 \neq \beta_2$, and specify its distribution. (20%)

卷三 2006.9.8

Linear Models (95.9)

1.(14%) Let Ω be the vector space of all 2x3 matrices, and $\mathbf{y}=\begin{bmatrix}13 & 9 & 8\\ 11 & 3 & 10\end{bmatrix}$. Find a set of orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_6\}$ such that $V_0=L(\mathbf{v}_1)$, $V_R=L(\mathbf{v}_2)$, $V_C=L(\mathbf{v}_3, \mathbf{v}_4)$, and $V^{\perp}=L(\mathbf{v}_5, \mathbf{v}_6)$.

a.(8%) Assume that $\hat{y}_i=p(y|V_i)$, i=0, R, C. Compute $\hat{y}_0$, $\hat{y}_R$, $\hat{y}_C$, $\hat{y}=\hat{y}_0+\hat{y}_R+\hat{y}_C$, and the residual $y-\hat{y}$.

b.(6%) Compute the squared lengths of the vectors in a. Verify that the Pythagorean equation applies.

2.(10%) Let $\mathbf{x}_1=\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_2=\begin{bmatrix}1\\1\\1\\0\end{bmatrix}$, $\mathbf{x}_3=\begin{bmatrix}0\\1\\1\\-1\end{bmatrix}$, $V_1=L(\mathbf{x}_1)$, $V_2=L(\mathbf{x}_2, \mathbf{x}_3)$, and $V=V_1\oplus V_2$.

Obtain the vectors V_1, V_2, V, $V_1^{\perp}$, $V_2^{\perp}$, and $V^{\perp}$.

3.(8%) Let V_1 and V_2 be subspaces of Ω. Show that $(V_1+V_2)^{\perp}=V_1^{\perp}\cap V_2^{\perp}$.

4.(12%) Let $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be random samples taken from normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively. Specify the sampling distributions of the following statistics.

a.(3%) $[(\bar{X}-\bar{Y})-\delta]/[S_2\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}]$. Assume that $\sigma_1^2=\sigma_2^2=\sigma^2$, and δ denotes a constant.

b.(3%) $\frac{1}{\sigma^2}[\Sigma(X_i-\mu_0)^2+\Sigma(Y_i-\mu_0)^2]$. Assume that $\sigma_1^2=\sigma_2^2=\sigma^2$.

c.(3%) $K[\Sigma(X_i-\mu_0)^2/\Sigma(Y_i-\bar{Y})^2]$. Please also specify constant K.

d.(3%) $K\frac{X_1+X_2}{|Y_1-Y_2|}$. Please also specify constant K.

5.(10%) Let $x_1=[1, 1, 1, 1]'$, $x_2=[1, 0, 1, 0]'$, $x_3=3x_1-2x_2$. Is $2\beta_1+\beta_2+4\beta_3$ estimable for the model $Y=\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\varepsilon$? Find conditions for $c=[c_1, c_2, c_3]'$ such that $\eta=c_1\beta_1+c_2\beta_2+c_3\beta_3$ be estimable for the same model.

$2,\ldots, n_i$, i=1, 2, 3, n_1=4, n_2=5, and n_3=4. Furthermore, let $\mathbf{J}_1=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ & 0 & \end{bmatrix}$,

$\mathbf{J}_2=\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ & 1 & \end{bmatrix}$, and $\mathbf{J}_3=\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ & 0 & \end{bmatrix}$ denote the condition effect indicator

matrices, and the entry x_{ij} in the matrix $\mathbf{x}=\begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \\ x_{14} & x_{24} & x_{34} \\ & x_{25} & \end{bmatrix}$ denotes the

amount of fertilization for the blueberry of ith condition on the jth plot. We are now fitting the model to the yield of blueberry.

$$\mathbf{y}=\sum_{j=1}^{3}\beta_j \mathbf{J}_j+\beta\mathbf{x}+\boldsymbol{\varepsilon}.$$

And now we have $\mathbf{y}=\begin{bmatrix} 6 & 10 & 13 \\ 9 & 7 & 21 \\ 11 & 15 & 18 \\ 10 & 10 & 20 \\ & 13 & \end{bmatrix}$, and $\mathbf{x}=\begin{bmatrix} 2 & 2 & 2 \\ 2 & 1 & 4 \\ 5 & 5 & 3 \\ 3 & 3 & 3 \\ & 4 & \end{bmatrix}$.

a.(6%) Estimate β_1, β_2, β_3, and β.

b.(6%) Compute $\hat{y}$ and S^2.

c.(6%) Use α=0.05 to test H_0: $\beta_1=\beta_2=\beta_3$ against the appropriate alternative.

6.(10%) Assume that $\mathbf{x}_1=[1, 1, 1, 1, 1]'$, $\mathbf{x}_2=[1, 1, 1, 0, 0]'$, $\mathbf{x}_3=[1, 0, 0, 1, 0]'$, and y $=[2, -2, -1, 7, 4]'$, and we are fitting the model $y=\sum_{j=1}^{3}\beta_j x_j+\varepsilon$ to the data.

a.(5%) Use $\alpha=0.05$ to test $H_0: \beta_2+2\beta_3=0$ against the appropriate alternative.

b.(5%) Use $\alpha=0.05$ to test $H_0: \beta_1=0$ and $\beta_2+2\beta_3=0$ against the appropriate alternative.

(Matrices for you to use: $\begin{bmatrix} 5 & 3 & 2 \\ 3 & 3 & 1 \\ 2 & 1 & 2 \end{bmatrix}^{-1}=\frac{1}{7}\begin{bmatrix} 5 & -4 & -3 \\ -4 & 6 & 1 \\ -3 & 1 & 6 \end{bmatrix}$.)

7.(18%) A chemist wishes to determine the percentages of impurities β_1 and β_2 in two 100g containers (1 and 2) of potassium chloride. The process she uses is able to measure the weight in grams of impurities in any 2g sample of the solution with mean equal to the true weight of the impurities and standard deviation 0.006g. She makes three measurements. Measurement #1 is on a 2g sample from container 1. Measurement #2 is on a 2g sample from container 2. Measurement #3 is on a mixture of a 1g sample from container 1 and a 1g sample from container 2.

a.(3%) Obtain the linear model.

b.(4%) Give unbiased estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ for β_1 and β_2, respectively.

c.(5%) Compute the covariance matrix of $\hat{\beta}_1$ and $\hat{\beta}_2$.

d.(6%) Assume that $Y=[0.056, 0.036, 0.058]'$, compute $\hat{\beta}_1$, $\hat{\beta}_2$, and S^2.

8.(18%) We are modeling yields of blueberry of three different conditions with several plots within each condition. Let $\mathbf{y}=\begin{bmatrix} y_{11} & y_{21} & y_{31} \\ y_{12} & y_{22} & y_{32} \\ y_{13} & y_{23} & y_{33} \\ y_{14} & y_{24} & y_{34} \\ & y_{25} & \end{bmatrix}$, in which the entry y_{ij} denotes the yield of blueberry of the ith condition on the jth plot, where j=1,