第一卷

Linear Models 97.9.

1.(8%) Suppose that $\mathbf{X}=\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix}\right)$, and $P(X_1 - 2X_2 + 3X_3 - 5 \le c) = 0.90$. What is c?

2.(10%) The vector space V, defined as $V=L(x_1, x_2, \ldots, x_k)$, has dimension k. Find simple formulae for the coefficients a_j in $p(y|V)=\sum_j a_j x_j^{\perp}$, and show that $L(x_1^{\perp}, x_2^{\perp}, \ldots, x_k^{\perp})=V$.

3.(18%) Suppose the model $\mathbf{y}=\sum_{j=1}^{3}\beta_j x_j + \varepsilon$ is to be fitted to $\mathbf{x}_1=[1, 1, 1, 1, 1]'$, $\mathbf{x}_2=[1, 1, 1, 0, 0]'$, $\mathbf{x}_3=[1, 0, 0, 1, 1]'$, and $\mathbf{y}=[20, -6, -14, 8, 7]'$.

a.(6%) Compute the estimates for β_j, j=1, 2, 3.

b.(6%) Compute $\mathbf{x}_j^{\perp}$, j=1, 2, 3.

c.(6%) Use the alternative model $\mathbf{y}=\sum_{j=1}^{3}\beta_j^* x_j^{\perp} + \varepsilon$ to estimate β_j^*, and thus β_j. Compare the results with what you got in part a.

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

4.(12%) Let $\mathbf{y}=\sum_{j=1}^{k}\beta_j x_j + \varepsilon$, where $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ denote linearly independent vectors, and $E[\varepsilon]=\mathbf{0}$, $D[\varepsilon]=\sigma^2 I$. Let $\eta=\sum_{j=1}^{k} c_j \beta_j$, and η^* be a linear unbiased estimator of η, then $\mathrm{var}(\eta^*) \ge \mathrm{var}(\hat{\eta})$, where $\hat{\eta}=\sum_{j=1}^{k} c_j \hat{\beta}_j$.

5.(18%) Suppose $\Omega=R_4$, $\mathbf{x}_1=[1, 1, 1, 1]'$, $\mathbf{x}_2=[0, 1, 0, 1]'$, $\mathbf{x}_3=[1, 0, -1, 0]'$, $\mathbf{x}_4=[0, 3, 1, 3]'$, $V=L(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ be a supspace of Ω, and $\eta=\beta_1+\beta_2+2\beta_3+2\beta_4$.

a.(6%) Verify that η is estimable. Compute the vector **a** such that $\hat{\eta}=(\mathbf{a}, \mathbf{y})$ is the BLUE of η.

b.(6%) Give a vector **d** different from the vector **a** such that $\eta^*=(\mathbf{d}, \mathbf{y})$ be another linear unbiased estimator of η. Show that $p(\mathbf{d}|V)=\mathbf{a}$, and $\mathrm{var}(\eta^*)-\mathrm{var}(\hat{\eta})>0$.

c.(6%) Now it is given that $\mathbf{y}=[4, 8, 5, 6]'$, compute $\hat{\eta}$ and η^*.

6.(16%)The joint probability mass function of the pair of random variables (X, Y) is tabulated as follows.

		X		
		0	1	2
Y	0	0	0.35	0.2
	1	0.15	0.15	0.15

Find the least squares predictor $g(\mathbf{X})=E[\mathbf{Y}|\mathbf{X}]$, and the linear least squares predictor $h(\mathbf{X})=\hat{Y}=\mu_Y+\rho\sigma_Y(X-\mu_X)/\sigma_X$. Show that $g(\mathbf{X})$ and $h(\mathbf{X})$ are unbiased estimators of $E[Y]$. Find $var(g(\mathbf{X}))$, $E[(\mathbf{Y}-g(\mathbf{X}))^2]$, $var(\hat{Y})$, and $E[(\mathbf{Y}-\hat{Y})^2]$.

7.(18%)The function g(T) is used to denote the expected resulted amount of substance A produced from a standard procedure under temperature T with 500 grams of input material. It is known that nothing will be produced when the temperature is below 30 degrees. The amount is a linear function of T when T comes between 30 and 120 degrees. The amount will still be another linear function of T when T goes above 120 degrees. The experiment points are T= 40, 70, 100, 130, 190, and 250.

a.(6%) Work out the appropriate linear model.

b.(12%) Suppose that the amounts corresponding to the set temperatures are 38, 136, 251, 302, 166, and 63, respectively. Please compute the estimates for the parameters in your model especially the variance σ^2.

第二卷

1. (20%) Consider the model $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, \cdots, n$. Let x_0 be the value at which the function $E(y) = \beta_0 + \beta_1 x + \beta_2 x^2$ is maximized (or minimized).

   (1) Find the maximum likelihood estimator of x_0.

   (2) Find a $(1-\alpha)100\%$ confidence interval for x_0.

2. (30%) Suppose that we collected data (y_i, x_i) $i = 1, \cdots, n+m$ and that we know that the line changes phase between x_n and x_{n+1}. The model $y_i = \beta_{10} + \beta_{11} x_i + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, \cdots, n$ applies to the first phase and the model $y_i = \beta_{20} + \beta_{21} x_i + \epsilon_i$, $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, $i = n+1, \cdots, n+m$ applies to the second phase. Let γ be the value of x at which the lines intersect.

   (1) Find the maximum likelihood estimators for β_{10}, β_{11}, β_{20}, β_{21}, σ^2, and γ. (Hint: γ is a function of other parameters.)

   (2) Find a $(1-\alpha)100\%$ confidence interval for γ.

3. (30%) Consider a set of seemingly unrelated regression equations

$$Y_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(0, \sigma^2 I), \quad i = 1, 2, \cdots, r$$

   where $\boldsymbol{X}_i$ is an $n_i \times p$ matrix and the $\boldsymbol{\epsilon}_i$'s are independent. Find the test for $H_0 : \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_r$.

4. (20%) Let $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$, where

$$\begin{aligned} \mu_{ij} &= \bar{\mu}_{..} + (\bar{\mu}_{i.} - \bar{\mu}_{..}) + (\bar{\mu}_{ij} - \bar{\mu}_{i.}) \\ &= \mu + \alpha_i + \beta_{ij}, \end{aligned}$$

   where $\epsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2)$, $i = 1, \cdots, I$, $j = 1, \cdots, J$, and $k = 1, \cdots, K$. Please obtain the test statistics for testing hypotheses $H_0 : \beta_{ij} = 0$ for all i, j.

第三卷

Linear Models (Qualifying Exam)

9/9/08

1. (50%) Let $\mathbf{y} \sim N_2(\mathbf{0}, \mathbf{I}_2 + \mathbf{J}_2)$, where $\mathbf{J}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Let $U_1 = \mathbf{y}'\mathbf{A}_1\mathbf{y}$ and $U_2 = \mathbf{y}'\mathbf{A}_2\mathbf{y}$ with

$$\mathbf{A}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } \mathbf{A}_2 = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$$

(a) Are U_1 and U_2 independent?

(b) Find $P(U_1/U_2 > 1)$. What is U_1/U_2 distributed?

2. (50%) Let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where **X** is $n \times p$ of rank *p*, and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Partition **X** into $\mathbf{X} = [\mathbf{X}_1 \,|\, \mathbf{X}_2 \,|\, \mathbf{X}_3]$, where $\mathbf{X}_i$ is $n \times p_i$ of rank p_i, and let $\boldsymbol{\beta}' = [\boldsymbol{\beta}_1' \,|\, \boldsymbol{\beta}_2' \,|\, \boldsymbol{\beta}_3']$. Hence, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_3\boldsymbol{\beta}_3 + \boldsymbol{\varepsilon}$.

(a) Show that successively fitting the models $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$, $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, and $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ yields SS's for $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and $\boldsymbol{\beta}_3$ which are orthogonal.

(b) Prove that the SSE for the model containing $\boldsymbol{\beta}$ must be at least as small as the SSE for the model with only $\boldsymbol{\beta}_1$.

(c) Now, let $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{V})$, $\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, and $\hat{\boldsymbol{\beta}}_{wls} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Show that $\text{cov}(\hat{\boldsymbol{\beta}}_{ols}) - \text{cov}(\hat{\boldsymbol{\beta}}_{wls}) \geq 0$ (i.e. non-negative definite).

(Hints: (1) $\mathbf{A} \geq \mathbf{B}$ iff $\mathbf{B}^{-1} \geq \mathbf{A}^{-1}$. (2) If $\mathbf{C} > 0$, then $\mathbf{CAC} \geq \mathbf{B}$ iff $\mathbf{A} \geq \mathbf{C}^{-1}\mathbf{B}\mathbf{C}^{-1}$.)