2003/09/30
線性模式

Linear Models

★ The transpose of a matrix A is denoted by A' or $A^{'}$.
★ Assume X_1 and X_2, $(X_1,\ X_2)$ have full rank.
★ *Your answers should be calculated to the simplest form.*

1. Consider the linear model

$$E(Y) = X_1\beta_1 + X_2\beta_2$$

where X_1 is $n \times p$, X_2 is $n \times q$, and β_2 is the parameter of interest. Let $H_1 = X_1(X_1'X_1)^{-1}X_1'$, $D_1 = I - H_1$, $H_2 = D_1X_2(X_2'D_1X_2)^{-1}X_2'D_1$, where I is the $n \times n$ identity matrix and $Var(Y) = \sigma^2 I$.

a. [10%] Calculate the expected value of the LSE, say, $\widehat{\beta}_2$, derived from regressing Y on X_2 when X_1 is ignored, and derive the expected value of the corresponding residual sum of squares.

b. [10%] Calculate the expected value of the LSE, say, $\widehat{\beta}_2^*$, derived from regressing Y on $(I - H_1)X_2$, and derive the expected value of the corresponding residual sum of squares.

c. [10%] Calculate the expected value of the LSE, say, $\widehat{\beta}_2$, derived from regressing $(I - H_1)Y$ on $(I - H_1)X_2$, and derive the expected value of the corresponding residual sum of squares.

d. [10%] Consider the partition

$$Y'Y = Y'H_1Y + Y'H_2Y + Y'(I - H_1 - H_2)Y$$

show that $H_1H_2 = H_1(I - H_1 - H_2) = H_2(I - H_1 - H_2) = 0$. Now assume normality, give a test for $H_0 : \beta_2 = 0$. (You should provide theorem that supports the legitimacy of your test)

2. Consider b_0 and b_1, the least squares estimates of β_0 and β_1 derived under the model $E(\mathbf{Y}) = \beta_0 + \mathbf{X}_1\beta_1$. *Now suppose the true model is* $E(\mathbf{Y}) = \beta_0 + \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2$.

a. [10%] What is the expected value of $(b_0,\ b_1)$?

b. [10%] What is the expected value of $\sum(y_i - \hat{y}_i)^2$, where $\hat{y}_i$ is the predicted value based on b_0 and b_1?

c. [10%] Is $E(\widehat{\mathbf{Y}})$ equal to $E(\mathbf{Y})$?

A

3. (30%) Consider the following two ANOVA tables and fill in all the missing values in □.

Sequential ANOVA

Source	df	Sum of Squares	F
A	2	41.7	*F* statistic= □ df= □
B	3	□	*F* statistic= □ df= □
C	2	23.2	*F* statistic= □ df= □
Error	18	100.5	
Total (corrected)		221.0	

Partial ANOVA

Source	df	Sum of Squares	F
A	2	51.7	*F* statistic= □ df= □
B	□	65.2	*F* statistic= □ df= □
C	2	□	*F* statistic= □ df= □
Error	18	100.5	
Total (corrected)		221.0	

Linear Models (Qualifying Exam)

9/30/03

1. Consider the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$, and $\mathbf{X}$ is $n \times p$ of rank p.
   (a) Partition $\mathbf{X}$ into $[\mathbf{X}_1 \mid \mathbf{X}_2]$ and show that SSE_F, the error sum of squares for the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, can not be larger than SSE_R, the error sum of squares for the model $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$.
   (b) Under the full model (i.e. $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$), find the distribution of SSE_R as defined in part (a).

2. Let $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_3(\mathbf{0}, \sigma^2\mathbf{I}_3)$. Assume that $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$ and

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 4 \end{bmatrix}.$$

   (a) Find the variance of $\hat{\beta}_1$, the OLS estimator of β_1.

   (b) Prove or disprove that $\breve{\beta}_1 = (Y_3 - Y_1)/3$ is unbiased, but has a variance that is larger than the variance of $\hat{\beta}_1$.

B

1. Each of the following problems could occur in fitting a multiple linear regression model. For each of the problems suggest a possible solution. Please justify and explain your answers.

(1) The $X^T X$ matrix contains some very large values.

(2) Multicollinearity exists.

(3) The residuals are uncorrelated but have a non-constant variance.

(4) The overall model is statistically significant, but none of the individual parameter estimates is significant.

2. Consider the two models $Y_1 = X_1\beta_1 + \epsilon_1$ and $Y_2 = X_2\beta_2 + \epsilon_2$, where the X_i's are $n_i \times p$ matrices and $i = 1, 2$. Suppose that $\epsilon_i \sim N(0, \sigma_i^2 I)$ and that ϵ_1 and ϵ_2 are independent.

(1) Assuming that the σ_i's are known, obtain a test for the hypothesis $\beta_1 = \beta_2$.

(2) Assume that $\sigma_1 = \sigma_2$ but they are unknown. Derive a test for the hypothesis $\beta_1 = \beta_2$.

3. Consider a linear mixed-effect model

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i$$

where

$$\begin{aligned} b_i &\sim N(0, D) \\ \epsilon_i &\sim N(0, \Sigma_i), \end{aligned}$$

and $b_1, \cdots, b_N, \epsilon_1, \cdots, \epsilon_N$ are independent. Here Y_i is the n_i-dimensional response vector for subject i, $i = 1, \cdots, N$, N is the number of subjects, X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates, β is a p-dimensional vector containing the fixed effects, b_i is the q-dimensional vector containing the random effects, ϵ_i is an n_i-dimensional vector of residual components, D is a general $(q \times q)$ covariance matrix, and Σ_i is a $(n_i \times n_i)$ covariance matrix.

(1) Please find the marginal distribution of Y_i.

(2) Find the MLE of β, denoted by $\hat{\beta}$, based on part (1).

(3) Please also show $Var(\hat{\beta})$.

(4) Find $\hat{b}_i$ by using the fact that $\hat{b}_i = E(b_i|Y_i = y_i)$.

(5) Please show $Var(\hat{b}_i)$.

C