

Multivariate Linear Regression Models

Shyh-Kang Jeng

Department of Electrical Engineering/
Graduate Institute of Communication/
Graduate Institute of Networking and
Multimedia

1

Outline

- Introduction
- The Classical Linear Regression Model
- Least Square Estimation
- Inference about the Regression Model
- Inference from the Estimated Regression Function

2

Outline

- Model Checking and Other Aspects of Regression
- Multivariate Multiple Regression
- The Concept of Linear Regression
- Comparing the Two Formulations of the Regression Model
- Multiple Regression Models with Time Dependent Errors

3

Outline

- Introduction
- The Classical Linear Regression Model
- Least Square Estimation
- Inference about the Regression Model
- Inference from the Estimated Regression Function

4

Questions

- What is regression analysis?

5

Regression Analysis

- A statistical methodology
 - For predicting value of one or more response (dependent) variables
 - Predict from a collection of predictor (independent) variable values

6

Example 7.1 Fitting a Straight Line

Observed data

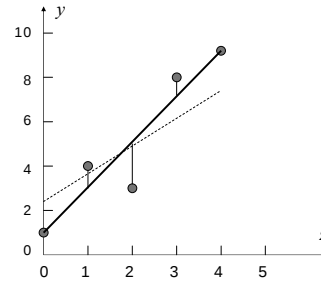
z_i	0	1	2	3	4
y	1	4	3	8	9

Linear regression model

$$\text{Mean response} = E(Y) = \beta_0 + \beta_1 z_1$$

7

Example 7.1 Fitting a Straight Line



8

Outline

- Introduction
- The Classical Linear Regression Model
- Least Square Estimation
- Inference about the Regression Model
- Inference from the Estimated Regression Function

9

Questions

- What is the classical regression model?
- How to treat a one-way ANOVA problem as the classical regression model?

10

Classical Linear Regression Model

$$Y = \beta_0 + \beta_1 z_1 + \cdots + \beta_r z_r + \varepsilon$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_{11} & z_{12} & \cdots & z_{1r} \\ 1 & z_{21} & z_{22} & \cdots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \cdots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad z_{j0} = 1$$

11

Classical Linear Regression Model

$$E(\varepsilon_j) = 0$$

$$\text{Var}(\varepsilon_j) = \sigma^2$$

$$\text{Cov}(\varepsilon_j, \varepsilon_k) = 0, \quad j \neq k$$

$$\Rightarrow$$

$$E(\boldsymbol{\varepsilon}) = \mathbf{0}$$

$$\text{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}$$

12

Example 7.1

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_5 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} 1 & z_{11} \\ 1 & z_{21} \\ \vdots & \vdots \\ 1 & z_{51} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_5 \end{bmatrix}$$

$$\mathbf{y}' = [1 \quad 4 \quad 3 \quad 8 \quad 9]$$

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}$$

13

Examples 6.7 & 6.8

$$\begin{pmatrix} 9 & 6 & 9 \\ 0 & 2 \\ 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 4 \\ 4 & 4 \\ 4 & 4 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -3 & -3 \\ -2 & -2 & -2 \end{pmatrix} + \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 \\ 1 & -1 & 0 \end{pmatrix}$$

$$SS_{obs} = 216, \quad SS_{mean} = 128$$

$$SS_{tr} = 78, \quad \text{d.f.} = 3 - 1 = 2$$

$$SS_{res} = 10, \quad \text{d.f.} = (3 + 2 + 3) - 3 = 5$$

$$F = \frac{SS_{tr} / (g - 1)}{SS_{res} / (\sum n_t - g)} = \frac{78 / 2}{10 / 5} = 19.5 > F_{2,5}(0.01) = 13.27$$

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0 \text{ is rejected at the 1\% level}$$

14

Example 7.2 One-Way ANOVA

$$X_{1j} = \mu + \tau_1 + e_{1j}, \quad X_{2j} = \mu + \tau_2 + e_{2j}, \quad X_{3j} = \mu + \tau_3 + e_{3j}$$

$$Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \beta_3 z_{j3} + \varepsilon_j$$

$$\beta_0 = \mu, \quad \beta_1 = \tau_1, \quad \beta_2 = \tau_2, \quad \beta_3 = \tau_3$$

$$z_j = \begin{cases} 1 & \text{if the observation is from population } j \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbf{Y}' = [9 \quad 6 \quad 9 \quad 0 \quad 2 \quad 3 \quad 1 \quad 2]$$

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

15

Outline

- Introduction
- The Classical Linear Regression Model
- Least Square Estimation
- Inference about the Regression Model
- Inference from the Estimated Regression Function

16

Questions

- What is the method of least squares?
- What is the least square estimation about the assumed coefficients in the classical regression model? (Result 7.1)
- What is the coefficient of determination?
- How to explain the results of the least square estimation through geometry?

17

Questions

- What is the projection matrix?
- What are the expectation of the estimated coefficients and the residual? What are the covariance matrix and the variance of the residual? (Result 7.2)
- What is the Gauss least square theorem? (Result 7.3)

18

Method of Least Squares

Selects $\mathbf{b}$ so as to minimize

$$S(\mathbf{b}) = \sum_{j=1}^n (y_j - b_0 - b_1 z_{j1} - \cdots - b_r z_{jr})^2$$

$$= (\mathbf{y} - \mathbf{Zb})(\mathbf{y} - \mathbf{Zb})'$$

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} S(\mathbf{b})$$

$$\text{residuals} = \hat{\varepsilon}_j = y_j - \hat{\beta}_0 - \hat{\beta}_1 z_{j1} - \cdots - \hat{\beta}_r z_{jr}$$

$$\hat{\varepsilon} = \mathbf{y} - \mathbf{Z}\hat{\mathbf{b}} = \mathbf{y} - \hat{\mathbf{y}}, \quad \text{fitted } \mathbf{y} = \hat{\mathbf{y}} = \mathbf{Z}\hat{\mathbf{b}}$$

19

Result 7.1

$$\mathbf{Z} \text{ has full rank } r+1 \leq n, \quad \hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

$$\hat{\mathbf{y}} = \mathbf{Z}\hat{\mathbf{b}} = \mathbf{H}\mathbf{y}, \quad \mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$$

$$\hat{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$$

$$\mathbf{Z}'\hat{\varepsilon} = 0, \quad \hat{\mathbf{y}}'\hat{\varepsilon} = 0, \quad \mathbf{Z}'\mathbf{y} = \mathbf{Z}'\hat{\mathbf{y}}$$

$$\hat{\varepsilon}'\hat{\varepsilon} = \mathbf{y}'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}\hat{\mathbf{b}}$$

20

Proof of Result 7.1

$$\hat{\mathbf{b}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

$$\mathbf{y} - \mathbf{Zb} = \mathbf{y} - \mathbf{Z}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{b}} - \mathbf{Zb} = \mathbf{y} - \mathbf{Z}\hat{\mathbf{b}} + \mathbf{Z}(\hat{\mathbf{b}} - \mathbf{b})$$

$$S(\mathbf{b}) = (\mathbf{y} - \mathbf{Zb})(\mathbf{y} - \mathbf{Zb})'$$

$$= (\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}})(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}})' + (\hat{\mathbf{b}} - \mathbf{b})'\mathbf{Z}'\mathbf{Z}(\hat{\mathbf{b}} - \mathbf{b})$$

$$+ 2(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}})\mathbf{Z}(\hat{\mathbf{b}} - \mathbf{b})'$$

$$(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}})\mathbf{Z} = \mathbf{y}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Z}$$

$$= \mathbf{y}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{Z} = \mathbf{y}'(\mathbf{Z} - \mathbf{Z}) = 0$$

$$\hat{\mathbf{b}} = \arg \min_{\mathbf{b}} S(\mathbf{b})$$

21

Proof of Result 7.1

$$\mathbf{Z}'\hat{\varepsilon} = \mathbf{Z}'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{Z}'(\mathbf{y} - \mathbf{Z}\hat{\mathbf{b}})$$

$$= \mathbf{Z}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{y} = 0$$

$$\hat{\mathbf{y}}'\hat{\varepsilon} = \hat{\mathbf{b}}'\mathbf{Z}'\hat{\varepsilon} = 0$$

$$\hat{\varepsilon}'\hat{\varepsilon} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$$

$$= \mathbf{y}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{y}$$

$$= \mathbf{y}'(\mathbf{I} - 2\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' + \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{y}$$

$$= \mathbf{y}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\hat{\mathbf{y}}$$

22

Example 7.1 Fitting a Straight Line

✦ Observed data

z_i	0	1	2	3	4
y	1	4	3	8	9

✦ Linear regression model

$$\text{Mean response} = E(Y) = \beta_0 + \beta_1 z_1$$

23

Example 7.3

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} 5 & 10 \\ 10 & 30 \end{bmatrix}, \quad (\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$$

$$\mathbf{Z}'\mathbf{y} = \begin{bmatrix} 25 \\ 70 \end{bmatrix}, \quad \hat{\mathbf{b}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\hat{\mathbf{y}} = 1 + 2z, \quad \hat{\mathbf{y}} = \mathbf{Z}\hat{\mathbf{b}} = \begin{bmatrix} 1 & 3 & 5 & 7 & 9 \end{bmatrix}'$$

$$\hat{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 0 & 1 & -2 & 1 & 0 \end{bmatrix}'$$

$$\text{residual sum of equations } \hat{\varepsilon}'\hat{\varepsilon} = 6$$

24

Coefficient of Determination

$$\hat{\mathbf{y}}' \hat{\boldsymbol{\varepsilon}} = 0$$

$$\mathbf{y}' \mathbf{y} = (\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}})'(\hat{\mathbf{y}} + \mathbf{y} - \hat{\mathbf{y}}) = (\hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}})'(\hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}}) = \hat{\mathbf{y}}' \hat{\mathbf{y}} + \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}$$

$$\mathbf{Z}' \hat{\boldsymbol{\varepsilon}} = 0 \Rightarrow 0 = \mathbf{1}' \hat{\boldsymbol{\varepsilon}} = \sum_{j=1}^n \hat{\varepsilon}_j = \sum_{j=1}^n y_j - \sum_{j=1}^n \hat{y}_j \Rightarrow \bar{y} = \bar{\hat{y}}$$

$$\mathbf{y}' \mathbf{y} - n \bar{y}^2 = \hat{\mathbf{y}}' \hat{\mathbf{y}} - n(\bar{\hat{y}})^2 + \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}$$

$$\sum_{j=1}^n (y_j - \bar{y})^2 = \sum_{j=1}^n (\hat{y}_j - \bar{\hat{y}})^2 + \sum_{j=1}^n \hat{\varepsilon}_j^2$$

$$R^2 = 1 - \frac{\sum_{j=1}^n \hat{\varepsilon}_j^2}{\sum_{j=1}^n (y_j - \bar{y})^2} = \frac{\sum_{j=1}^n (\hat{y}_j - \bar{\hat{y}})^2}{\sum_{j=1}^n (y_j - \bar{y})^2}$$

25

Geometry of Least Squares

$$E(\mathbf{Y}) = \mathbf{Z}\boldsymbol{\beta} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} z_{11} \\ z_{21} \\ \vdots \\ z_{n1} \end{bmatrix} + \cdots + \beta_r \begin{bmatrix} z_{1r} \\ z_{2r} \\ \vdots \\ z_{nr} \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

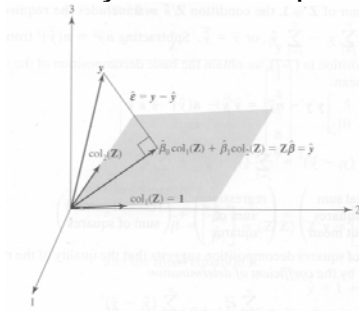
$\mathbf{y} - \mathbf{Z}\mathbf{b}$ = (observed vector) - (vector in model plane)

$$S(\mathbf{b}) = (\mathbf{y} - \mathbf{Z}\mathbf{b})'(\mathbf{y} - \mathbf{Z}\mathbf{b})$$

$$\hat{\boldsymbol{\beta}} = \arg \min_{\mathbf{b}} S(\mathbf{b}), \quad \hat{\mathbf{y}} = \mathbf{Z}\hat{\boldsymbol{\beta}} \text{ on model plane, } \hat{\boldsymbol{\varepsilon}} \perp \text{ model plane}$$

26

Geometry of Least Squares



27

Projection Matrix

$$\mathbf{Z}'\mathbf{Z} = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \cdots + \lambda_{r+1} \mathbf{e}_r \mathbf{e}_r'$$

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \frac{1}{\lambda_1} \mathbf{e}_1 \mathbf{e}_1' + \frac{1}{\lambda_2} \mathbf{e}_2 \mathbf{e}_2' + \cdots + \frac{1}{\lambda_{r+1}} \mathbf{e}_r \mathbf{e}_r'$$

$$\mathbf{q}_i = \lambda_i^{-1/2} \mathbf{Z} \mathbf{e}_i \Rightarrow \mathbf{q}_i' \mathbf{q}_k = \lambda_i^{-1/2} \lambda_k^{-1/2} \mathbf{e}_i' \mathbf{Z}' \mathbf{Z} \mathbf{e}_k = \lambda_i^{-1/2} \lambda_k^{-1/2} \mathbf{e}_i' \lambda_k \mathbf{e}_k = \delta_{ik}$$

$$\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' = \sum_{i=1}^{r+1} \lambda_i^{-1} \mathbf{Z} \mathbf{e}_i \mathbf{e}_i' \mathbf{Z}' = \sum_{i=1}^{r+1} \mathbf{q}_i \mathbf{q}_i'$$

projection of $\mathbf{y}$ on the model plane constructed by

$\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{r+1}\}$ is

$$\sum_{i=1}^{r+1} \mathbf{q}_i (\mathbf{q}_i' \mathbf{y}) = \left(\sum_{i=1}^{r+1} \mathbf{q}_i \mathbf{q}_i' \right) \mathbf{y} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \mathbf{y} = \mathbf{Z}\hat{\boldsymbol{\beta}}$$

28

Result 7.2

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}$$

$$\Rightarrow E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1}$$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\Rightarrow E(\hat{\boldsymbol{\varepsilon}}) = 0, \quad \text{Cov}(\hat{\boldsymbol{\varepsilon}}) = \sigma^2 [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'] = \sigma^2 [\mathbf{I} - \mathbf{H}]$$

$$E(\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}) = (n - r - 1) \sigma^2$$

$$s^2 = \frac{\hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}}}{n - (r + 1)} = \frac{\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}}{n - r - 1}, \quad E(s^2) = \sigma^2$$

$\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\varepsilon}}$ are uncorrelated

29

Proof of Result 7.2*

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'(\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}')\mathbf{Y} = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}')(\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

$$= (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}')\boldsymbol{\varepsilon}$$

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}$$

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \text{Cov}(\boldsymbol{\varepsilon}) \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1}$$

$$E(\hat{\boldsymbol{\varepsilon}}) = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}') E(\boldsymbol{\varepsilon}) = 0$$

$$\text{Cov}(\hat{\boldsymbol{\varepsilon}}) = (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}') \text{Cov}(\boldsymbol{\varepsilon}) (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}')$$

$$= \sigma^2 (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}')$$

30

Proof of Result 7.2*

$$\begin{aligned}
 \text{Cov}(\hat{\beta}, \hat{\varepsilon}) &= E[(\hat{\beta} - \beta)(\hat{\varepsilon})'] = (Z'Z)^{-1}Z'E(\varepsilon\varepsilon')[\mathbf{I} - Z(Z'Z)^{-1}Z'] \\
 &= \sigma^2(Z'Z)^{-1}Z'[\mathbf{I} - Z(Z'Z)^{-1}Z'] = 0 \\
 \hat{\varepsilon}\hat{\varepsilon}' &= \varepsilon'[\mathbf{I} - Z(Z'Z)^{-1}Z'][\mathbf{I} - Z(Z'Z)^{-1}Z']\varepsilon \\
 &= \varepsilon'[\mathbf{I} - Z(Z'Z)^{-1}Z']\varepsilon = \text{tr}[\varepsilon'(\mathbf{I} - Z(Z'Z)^{-1}Z')\varepsilon] \\
 &= \text{tr}[(\mathbf{I} - Z(Z'Z)^{-1}Z')\varepsilon\varepsilon'] \\
 E(\hat{\varepsilon}\hat{\varepsilon}') &= \text{tr}[(\mathbf{I} - Z(Z'Z)^{-1}Z')E(\varepsilon\varepsilon')] \\
 &= \sigma^2 \text{tr}[(\mathbf{I} - Z(Z'Z)^{-1}Z')] = \sigma^2 \text{tr}(\mathbf{I}) - \sigma^2 \text{tr}[(Z'Z)^{-1}Z'Z] \\
 &= \sigma^2(n - r - 1)
 \end{aligned}$$

31

Result 7.3

Gauss Least Square Theorem

$\mathbf{Y} = \mathbf{Z}\beta + \varepsilon$, $E(\varepsilon) = 0$, $\text{Cov}(\varepsilon) = \sigma^2\mathbf{I}$
 $\mathbf{c}'\hat{\beta} = c_0\hat{\beta}_0 + c_1\hat{\beta}_1 + \dots + c_r\hat{\beta}_r$ as an estimator of $\mathbf{c}'\beta$ has the smallest possible variance among all estimator of the form
 $\mathbf{a}'\mathbf{Y} = a_1Y_1 + a_2Y_2 + \dots + a_nY_n$ that are unbiased for $\mathbf{c}'\beta$

32

Proof of Result 7.3

For $\mathbf{a}'\mathbf{Y}$ as an unbiased estimator of $\mathbf{c}'\beta$,

$$E(\mathbf{a}'\mathbf{Y}) = E(\mathbf{a}'\mathbf{Z}\beta + \mathbf{a}'\varepsilon) = \mathbf{a}'\mathbf{Z}\beta = \mathbf{c}'\beta \Rightarrow \mathbf{a}'\mathbf{Z} = \mathbf{c}'$$

$$E(\mathbf{c}'\hat{\beta}) = \mathbf{c}'\beta$$

$$\mathbf{c}'\hat{\beta} = \mathbf{c}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} = \mathbf{a}^{*'}\mathbf{Y}, \quad \mathbf{a}^{*'}\mathbf{Z} = \mathbf{c}'$$

$$\text{Var}(\mathbf{a}'\mathbf{Y}) = \text{Var}(\mathbf{a}'\mathbf{Z}\beta + \mathbf{a}'\varepsilon) = \text{Var}(\mathbf{a}'\varepsilon) = \mathbf{a}'\mathbf{I}\sigma^2\mathbf{a}$$

$$= \sigma^2(\mathbf{a} - \mathbf{a}^* + \mathbf{a}^*)(\mathbf{a} - \mathbf{a}^* + \mathbf{a}^*)$$

$$= \sigma^2[(\mathbf{a} - \mathbf{a}^*)(\mathbf{a} - \mathbf{a}^*) + \mathbf{a}^{*'}\mathbf{a}^*]$$

is minimum when $\mathbf{a}'\mathbf{Y} = \mathbf{a}^{*'}\mathbf{Y} = \mathbf{c}'\hat{\beta}$ (BLUE)

33

Outline

- Introduction
- The Classical Linear Regression Model
- Least Square Estimation
- Inference about the Regression Model
- Inference from the Estimated Regression Function

34

Questions

- What are the maximum likelihood estimator to the coefficients and the assumed variance? (Result 7.4)
- What are the confidence region and the simultaneous confidence intervals for the assumed coefficients? (Result 7.5)
- How to know that the number of the coefficients has been enough? (Result 7.6)

35

Questions

- How to modify Result 7.6 when the rank of the $\mathbf{Z}$ matrix is not full?
- How to generalize Result 7.6 to the case that the coefficient vector is multiplied by a matrix?

36

Result 7.4

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} : N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

maximul likelihood estimator of $\boldsymbol{\beta}$ is the same as

the least squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}$

$$\hat{\boldsymbol{\beta}} : N_{r+1}(\boldsymbol{\beta}, \sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1})$$

independent of $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$

$\hat{\sigma}^2$: maximum likelihood estimator of σ^2

$$n\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} : \sigma^2 \chi_{n-r-1}^2$$

37

Proof of Result 7.4*

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= \prod_{j=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\varepsilon_j^2/2\sigma^2} = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/2\sigma^2} \\ &= \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-(\mathbf{y}-\mathbf{Z}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{Z}\boldsymbol{\beta})/2\sigma^2} \end{aligned}$$

For fixed σ^2 ,

$$\arg \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \sigma^2) = \arg \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})$$

$$= \hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}, \text{ independent of } \sigma$$

38

Proof of Result 7.4*

$$\begin{aligned} \hat{\sigma}^2 &= \arg \max_{\sigma^2} L(\hat{\boldsymbol{\beta}}, \sigma^2) \\ &= \frac{(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}})}{n} \\ &= \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n} \end{aligned}$$

39

Proof of Result 4.11*

Exponent of $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\begin{aligned} &-\frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right) \right] - \frac{1}{2} n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \\ \Rightarrow \hat{\boldsymbol{\mu}} &= \bar{\mathbf{x}} \end{aligned}$$

$$L(\hat{\boldsymbol{\mu}}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2} |\boldsymbol{\Sigma}|^{n/2}} e^{-\frac{1}{2} \text{tr} \left[\boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right) \right]}$$

$$\Rightarrow \hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' = \frac{n-1}{n} \mathbf{S}$$

40

Proof of Result 7.4*

$$\begin{aligned} \begin{bmatrix} \boldsymbol{\beta} \\ - \\ \hat{\boldsymbol{\varepsilon}} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\beta} \\ - \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \\ - \\ \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \end{bmatrix} \boldsymbol{\varepsilon} = \mathbf{A} + \mathbf{A}\boldsymbol{\varepsilon} \\ \text{Cov} \begin{bmatrix} \boldsymbol{\beta} \\ - \\ \hat{\boldsymbol{\varepsilon}} \end{bmatrix} &= \mathbf{A} \text{Cov}(\boldsymbol{\varepsilon}) \mathbf{A}' \\ &= \sigma^2 \begin{bmatrix} (\mathbf{Z}'\mathbf{Z})^{-1} & | & \mathbf{0} \\ - & & \\ \mathbf{0}' & | & \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \end{bmatrix} \end{aligned}$$

41

Proof of Result 7.4*

$$(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}') \mathbf{e} = \lambda \mathbf{e}$$

$$(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}')^2 = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'$$

$$\lambda^2 \mathbf{e} = \lambda \mathbf{e} \Rightarrow \lambda = 0, 1$$

$$\text{tr}(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}') = n - r - 1$$

$$\text{tr}(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}') = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

$$\lambda_1 = \lambda_2 = \dots = \lambda_{n-r-1} = 1, \quad \lambda_{n-r} = \lambda_{n-r+1} = \dots = \lambda_n = 0$$

$$\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' = \mathbf{e}_1 \mathbf{e}_1' + \mathbf{e}_2 \mathbf{e}_2' + \dots + \mathbf{e}_{n-r-1} \mathbf{e}_{n-r-1}'$$

42

Proof of Result 7.4*

$$\mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_{n-r-1} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1' \\ \mathbf{e}_2' \\ \vdots \\ \mathbf{e}_{n-r-1}' \end{bmatrix} \boldsymbol{\varepsilon} : N_{n-r-1}(\mathbf{0}, \text{Cov}(\mathbf{V}))$$

$$\text{Cov}(V_i, V_k) = \mathbf{e}_i' \text{Cov}(\boldsymbol{\varepsilon}) \mathbf{e}_k = \sigma^2 \delta_{ik}$$

$$V_i : \text{independent } N(0, \sigma^2)$$

$$n\hat{\sigma}^2 = \hat{\boldsymbol{\varepsilon}}' \hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}' (\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}') \boldsymbol{\varepsilon} = \sum_{i=1}^{n-r-1} \boldsymbol{\varepsilon}' \mathbf{e}_i \mathbf{e}_i' \boldsymbol{\varepsilon}$$

$$= V_1^2 + V_2^2 + \dots + V_{n-r-1}^2 : \sigma^2 \chi_{n-r-1}^2$$

43

χ^2 Distribution*

$$X_1 : N(\mu_1, \sigma_1^2), \quad X_2 : N(\mu_2, \sigma_2^2), \quad \dots,$$

$$X_\nu : N(\mu_\nu, \sigma_\nu^2); \quad Z_i = \frac{X_i - \mu_i}{\sigma_i} : N(0, 1)$$

$$\chi^2 = \sum_{i=1}^{\nu} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2, \quad \nu : \text{degrees of freedom (d.f.)}$$

$$f_n(\chi^2) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} (\chi^2)^{n/2-1} e^{-\chi^2/2}, & \chi^2 > 0 \\ 0, & \chi^2 \leq 0 \end{cases}$$

(Gamma distribution with $\alpha = n/2 - 1, \beta = 2$)

44

Result 7.5

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} : N_n(0, \sigma^2 \mathbf{I})$$

100(1- α)% confidence region for $\boldsymbol{\beta}$

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{Z}' \mathbf{Z} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq (r+1)s^2 F_{r+1, n-r-1}(\alpha)$$

Simultaneous 100(1- α)% confidence intervals for β_i

$$\hat{\beta}_i \pm \sqrt{\hat{\text{Var}}(\hat{\beta}_i)} \sqrt{(r+1)F_{r+1, n-r-1}(\alpha)}$$

$\hat{\text{Var}}(\hat{\beta}_i)$: diagonal element of $s^2(\mathbf{Z}'\mathbf{Z})^{-1}$

corresponding to $\hat{\beta}_i$

45

Proof of Result 7.5

$$\mathbf{V} = (\mathbf{Z}'\mathbf{Z})^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \quad E(\mathbf{V}) = \mathbf{0}$$

$$\text{Cov}(\mathbf{V}) = (\mathbf{Z}'\mathbf{Z})^{1/2} \text{Cov}(\hat{\boldsymbol{\beta}}) (\mathbf{Z}'\mathbf{Z})^{1/2} = \sigma^2 \mathbf{I}$$

$$\mathbf{V} : N_{r+1}(0, \sigma^2 \mathbf{I}), \quad \mathbf{V}'\mathbf{V} : \sigma^2 \chi_{r+1}^2$$

$$(n-r-1)s^2 : \sigma^2 \chi_{n-r-1}^2$$

$$\frac{\mathbf{V}'\mathbf{V} / (r+1)}{(n-r-1)s^2 / (n-r-1)} : F_{r+1, n-r-1}$$

Confidence region

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' \mathbf{Z}' \mathbf{Z} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = \mathbf{V}'\mathbf{V} \leq (r+1)s^2 F_{r+1, n-r-1}(\alpha)$$

46

Example 7.4 (Real Estate Data)

- 20 homes in a Milwaukee, Wisconsin, neighborhood
 - Regression model
- $$Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \varepsilon$$
- Y : selling price (thousands of dollars)
- z_1 : total dwelling size (hundreds of squared feet)
- z_2 : assessed value (thousands of dollars)

47

Example 7.4

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 5.1523 & & \\ 0.2544 & 0.0512 & \\ -0.1463 & -0.0172 & 0.0067 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y} = \begin{bmatrix} 30.967 \\ 2.634 \\ 0.045 \end{bmatrix}, \quad s = 3.473$$

$$s\sqrt{5.1523} = 7.88, \quad s\sqrt{0.0512} = 0.785, \quad s\sqrt{0.0067} = 0.285$$

$$\hat{y} = 30.967 + 2.634 z_1 + 0.045 z_2, \quad R^2 = 0.834$$

(7.88) (0.785) (0.285)

$$\hat{\beta}_2 \pm t_{17}(0.025) \sqrt{\hat{\text{Var}}(\hat{\beta}_2)} = 0.045 \pm 2.110 \times 0.285$$

95% confidence interval for β_2 : (-0.556, 0.647)

48

Result 7.6

Likelihood ratio test rejects $H_0: \beta_{(2)} = 0$ if

$$\frac{(SS_{res}(\mathbf{Z}_1) - SS_{res}(\mathbf{Z})) / (r - q)}{S^2} > F_{r-q, n-r-1}(\alpha)$$

$$\beta'_{(2)} = [\beta_{q+1} \quad \beta_{q+2} \quad \cdots \quad \beta_r] \quad \varepsilon: N_n(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$\mathbf{Y} = \mathbf{Z}\beta + \varepsilon = [\mathbf{Z}_1 \quad \mathbf{Z}_2] \begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \end{bmatrix} + \varepsilon = \mathbf{Z}_1 \beta_{(1)} + \mathbf{Z}_2 \beta_{(2)} + \varepsilon$$

$$SS_{res}(\mathbf{Z}_1) - SS_{res}(\mathbf{Z})$$

$$= (\mathbf{y} - \mathbf{Z}_1 \hat{\beta}_{(1)})' (\mathbf{y} - \mathbf{Z}_1 \hat{\beta}_{(1)}) - (\mathbf{y} - \mathbf{Z} \hat{\beta})' (\mathbf{y} - \mathbf{Z} \hat{\beta})$$

$$\hat{\beta}_{(1)} = (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' \mathbf{y}$$

49

Effect of Rank

✦ In situations where $\mathbf{Z}$ is not of full rank, $\text{rank}(\mathbf{Z})$ replaces $r+1$ and $\text{rank}(\mathbf{Z}_1)$ replaces $q+1$ in Result 7.6

50

Proof of Result 7.6

$$\max_{\beta, \sigma^2} L(\beta, \sigma^2) = \max_{\beta, \sigma^2} \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{(\mathbf{y} - \mathbf{Z}\beta)'(\mathbf{y} - \mathbf{Z}\beta)}{2\sigma^2}}$$

$$= \frac{1}{(2\pi)^{n/2} \hat{\sigma}^n} e^{-n/2}$$

which occurs at $\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}$ and $\hat{\sigma}^2 = (\mathbf{y} - \mathbf{Z}\hat{\beta})'(\mathbf{y} - \mathbf{Z}\hat{\beta})$

Under $H_0: \beta_{(2)} = 0$, $\mathbf{Y} = \mathbf{Z}_1 \beta_{(1)} + \varepsilon$ and

$$\max_{\beta_{(1)}, \sigma^2} L(\beta_{(1)}, \sigma^2) = \frac{1}{(2\pi)^{n/2} \hat{\sigma}_1^n} e^{-n/2}, \text{ where the maximum}$$

occurs at $\hat{\beta}_{(1)} = (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' \mathbf{y}$ and $\hat{\sigma}_1^2 = (\mathbf{y} - \mathbf{Z}_1 \hat{\beta}_{(1)})'(\mathbf{y} - \mathbf{Z}_1 \hat{\beta}_{(1)})$

51

Proof of Result 7.6

Reject $H_0: \beta_{(2)} = 0$ for small values of

$$\frac{\max_{\beta_{(1)}, \sigma^2} L(\beta_{(1)}, \sigma^2)}{\max_{\beta, \sigma^2} L(\beta, \sigma^2)} = \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}^2} \right)^{-n/2} = \left(1 + \frac{\hat{\sigma}_1^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \right)^{-n/2}$$

is equivalent to reject H_0 for large $(\hat{\sigma}_1^2 - \hat{\sigma}^2) / \hat{\sigma}^2$ or

$$\frac{n(\hat{\sigma}_1^2 - \hat{\sigma}^2) / (r - q)}{n\hat{\sigma}^2 / (n - r - 1)} = \frac{(SS_{res}(\mathbf{Z}_1) - SS_{res}(\mathbf{Z})) / (r - q)}{S^2} = F$$

$$n\hat{\sigma}^2: \sigma^2 \chi_{n-r-1}^2, \quad n\hat{\sigma}_1^2: \sigma^2 \chi_{n-q-1}^2, \quad F: F_{r-q, n-r-1}$$

52

Wishart Distribution

$$w_{n-1}(\mathbf{A} | \Sigma) = \frac{|\mathbf{A}|^{(n-p-2)/2} e^{-\text{tr}[\mathbf{A}\Sigma^{-1}]/2}}{2^{p(n-1)/2} \pi^{p(p-1)/4} |\Sigma|^{(n-1)/2} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n-i)\right)}$$

$\mathbf{A}$: positive definite

Properties:

$$\mathbf{A}_1: W_{m_1}(\mathbf{A}_1 | \Sigma), \quad \mathbf{A}_2: W_{m_2}(\mathbf{A}_2 | \Sigma) \Rightarrow$$

$$\mathbf{A}_1 + \mathbf{A}_2: W_{m_1+m_2}(\mathbf{A}_1 + \mathbf{A}_2 | \Sigma)$$

$$\mathbf{A}: W_m(\mathbf{A} | \Sigma) \Rightarrow \mathbf{C}\mathbf{A}\mathbf{C}': W_m(\mathbf{C}\mathbf{A}\mathbf{C}' | \mathbf{C}\Sigma\Sigma')$$

53

Generalization of Result 7.6

$\mathbf{C}: (r - q) \times (r + 1)$ matrix

Reject $H_0: \mathbf{C}\beta = \mathbf{0}$ at level α if

$$\frac{(\mathbf{C}\hat{\beta})(\mathbf{C}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta})}{S^2} > (r - q)F_{r-q, n-r-1}$$

since $\mathbf{C}\hat{\beta}: N_{r-q}(\mathbf{C}\beta, \sigma^2 \mathbf{C}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}')$ and

$$(\mathbf{C}\hat{\beta})(\mathbf{C}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta}): \sigma^2 \chi_{n-r}^2$$

54

Example 7.5 (Service Ratings Data)

Location	Gender	Service (Y)
1	0	15.2
1	0	21.2
1	0	27.3
1	0	21.2
1	0	21.2
1	1	36.4
1	1	92.4
2	0	27.3
2	0	15.2
2	0	9.1
2	0	18.2
2	0	50.0
2	1	44.0
2	1	63.6
3	0	15.2
3	0	30.3
3	1	36.4
3	1	40.9

55

Example 7.5: Design Matrix

	constant	location	gender	interaction						
$\mathbf{Z}^{(0)}$	1	1	0	0	1	0	0	0	0	5 responses
	1	1	0	0	1	0	0	0	0	
	1	1	0	0	1	0	0	0	0	
	1	1	0	0	1	0	0	0	0	
	1	1	0	0	1	0	0	0	0	
	1	1	0	0	0	1	0	0	0	2 responses
	1	1	0	0	0	1	0	0	0	
	1	0	1	0	1	0	0	1	0	5 responses
	1	0	1	0	1	0	0	1	0	
	1	0	1	0	1	0	0	1	0	
	1	0	1	0	1	0	0	1	0	
	1	0	1	0	1	0	0	1	0	
	1	0	1	0	0	1	0	0	1	2 responses
	1	0	1	0	0	1	0	0	1	
	1	0	0	1	1	0	0	0	0	2 responses
	1	0	0	1	1	0	0	0	0	
1	0	0	1	0	0	0	0	1	2 responses	
1	0	0	1	0	0	0	0	1		

56

Example 7.5

$$\beta' = [\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \tau_1 \quad \tau_2 \quad \gamma_{11} \quad \gamma_{12} \quad \gamma_{21} \quad \gamma_{22} \quad \gamma_{31} \quad \gamma_{32}]$$

$$\text{rank}(\mathbf{Z}) = 6, \quad \text{SS}_{\text{res}}(\mathbf{Z}) = 2977.4, \quad n - \text{rank}(\mathbf{Z}) = 12$$

$\mathbf{Z}_1$: first six columns of $\mathbf{Z}$

$$\text{SS}_{\text{res}}(\mathbf{Z}_1) = 3419.1, \quad n - \text{rank}(\mathbf{Z}_1) = 18 - 4 = 14$$

$$H_0: \gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = \gamma_{31} = \gamma_{32} = 0$$

$$F = \frac{(\text{SS}_{\text{res}}(\mathbf{Z}_1) - \text{SS}_{\text{res}}(\mathbf{Z})) / (6 - 4)}{s^2} = \frac{(\text{SS}_{\text{res}}(\mathbf{Z}_1) - \text{SS}_{\text{res}}(\mathbf{Z})) / 2}{\text{SS}_{\text{res}}(\mathbf{Z}) / 12}$$

= 0.89: insignificant for any appropriate level α

We can further verify that there is no location effect, but that the gender is significant

57

Outline

- Introduction
- The Classical Linear Regression Model
- Least Square Estimation
- Inference about the Regression Model
- Inference from the Estimated Regression Function

58

Questions

- What is the unbiased estimator of $E(Y_0 | \mathbf{z}_0)$ with minimum variance and its corresponding confidence intervals? (Result 7.7)
- What are the unbiased predictor and its prediction intervals of Y_0 ? (Result 7.8)

59

Result 7.7

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(0, \sigma^2 \mathbf{I})$$

$$\mathbf{z}_0' = [1 \quad z_{01} \quad \cdots \quad z_{0r}] \quad Y_0: \text{response at } \mathbf{z}_0$$

$$E(Y_0 | \mathbf{z}_0) = \beta_0 + \beta_1 z_{01} + \cdots + \beta_r z_{0r} = \mathbf{z}_0' \boldsymbol{\beta}$$

$\mathbf{z}_0' \hat{\boldsymbol{\beta}}$ is the unbiased estimator of $E(Y_0 | \mathbf{z}_0)$

with minimum variance.

$$\text{Var}(\mathbf{z}_0' \hat{\boldsymbol{\beta}}) = \mathbf{z}_0' (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{z}_0 \sigma^2$$

100(1 - α)% confidence interval of $E(Y_0 | \mathbf{z}_0)$ is

$$\mathbf{z}_0' \hat{\boldsymbol{\beta}} \pm t_{n-r-1}(\alpha/2) \sqrt{(\mathbf{z}_0' (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{z}_0) s^2}$$

60

Proof of Result 7.7

$\mathbf{z}_0'\hat{\boldsymbol{\beta}}$ is a linear combination of β_i 's $\Rightarrow$

$\mathbf{z}_0'\hat{\boldsymbol{\beta}}$ is the unbiased estimator of $\mathbf{z}_0'\boldsymbol{\beta}$ with the minimum variance by Result 7.3

$$\text{Var}(\mathbf{z}_0'\hat{\boldsymbol{\beta}}) = \mathbf{z}_0' \text{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{z}_0 = \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0 \sigma^2$$

$$\hat{\boldsymbol{\beta}} : N_{r+1}(\boldsymbol{\beta}, \sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1})$$

which is independent of $s^2 : \chi_{n-r-1}^2 / (n-r-1)$

$$\mathbf{z}_0'\hat{\boldsymbol{\beta}} : N(\mathbf{z}_0'\boldsymbol{\beta}, \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0 \sigma^2)$$

$$\frac{(\mathbf{z}_0'\hat{\boldsymbol{\beta}} - \mathbf{z}_0'\boldsymbol{\beta}) / \sqrt{\sigma^2 \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0}}{\sqrt{s^2 / \sigma^2}} = \frac{(\mathbf{z}_0'\hat{\boldsymbol{\beta}} - \mathbf{z}_0'\boldsymbol{\beta})}{\sqrt{s^2 \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0}} : t_{n-r-1}$$

Result 7.8

$$Y_0 = \mathbf{z}_0'\boldsymbol{\beta} + \varepsilon_0$$

$$\varepsilon_0 : N(0, \sigma^2) \text{ independent of } \boldsymbol{\varepsilon}, \hat{\boldsymbol{\beta}}, s^2$$

unbiased predictor of $Y_0 : \mathbf{z}_0'\hat{\boldsymbol{\beta}}$

$$\text{Var}(Y_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}}) = \sigma^2 (1 + \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0)$$

100(1- α)% prediction interval for Y_0 :

$$\mathbf{z}_0'\hat{\boldsymbol{\beta}} \pm t_{n-r-1}(\alpha/2) \sqrt{s^2 (1 + \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0)}$$

Proof of Result 7.8

$$\text{Forecast error } Y_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}} = \mathbf{z}_0'\boldsymbol{\beta} + \varepsilon_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}} = \varepsilon_0 + \mathbf{z}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$$

$$E(Y_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}}) = E(\varepsilon_0) + E(\mathbf{z}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) = 0$$

$$\begin{aligned} \text{Var}(Y_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}}) &= \text{Var}(\varepsilon_0) + \text{Var}(\mathbf{z}_0'(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) \\ &= \sigma^2 (1 + \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0) \end{aligned}$$

$$(Y_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}}) : N(0, \sigma^2 (1 + \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0))$$

$$(n-r-1)s^2 : \sigma^2 \chi_{n-r-1}^2$$

$$\frac{(Y_0 - \mathbf{z}_0'\hat{\boldsymbol{\beta}})}{\sqrt{s^2 (1 + \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0)}} : t_{n-r-1}$$

Example 7.6 (Computer Data)

z_1 (Orders)	z_2 (Add-delete items)	Y (CPU time)
123.5	2.108	141.5
146.1	9.213	168.9
133.9	1.905	154.8
128.5	.815	146.5
151.5	1.061	172.8
136.2	8.603	160.1
92.0	1.125	108.5

Source: Data taken from H. P. Artis, *Forecasting Computer Requirements: A Forecaster's Dilemma* (Piscataway, NJ: Bell Laboratories, 1979).

Example 7.6

$$\mathbf{z}_0' = [1 \quad 130 \quad 7.5] \quad \hat{y} = 8.42 + 1.08z_1 + 0.42z_2$$

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 8.17969 & & \\ -0.06411 & 0.00052 & \\ 0.08831 & -0.00107 & 0.01440 \end{bmatrix}$$

$$s = 1.204, \quad \mathbf{z}_0'\hat{\boldsymbol{\beta}} = 8.42 + 1.08*130 + 0.42*7.5 = 151.97$$

$$s\sqrt{\mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0} = 0.71, \quad t_4(0.025) = 2.776$$

95% confidence interval for the mean CPU time

$$\mathbf{z}_0'\hat{\boldsymbol{\beta}} \pm t_4(0.025)s\sqrt{\mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0} \text{ or } (150.00, 153.94)$$

95% prediction interval at $\mathbf{z}_0$

$$\mathbf{z}_0'\hat{\boldsymbol{\beta}} \pm t_4(0.025)s\sqrt{1 + \mathbf{z}_0' (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{z}_0} \text{ or } (148.08, 155.86)$$

Outline

- Model Checking and Other Aspects of Regression
- Multivariate Multiple Regression
- The Concept of Linear Regression
- Comparing the Two Formulations of the Regression Model
- Multiple Regression Models with Time Dependent Errors

Questions

- How to know the adequacy of the linear regression model?
- How to test independence of time?
- What is the leverage?
- What is the Mallows's C_p Statistic? How to use it?
- What is the stepwise regression?
- How to treat colinearity?

67

Questions

- What is the bias caused by a mis-specified model?

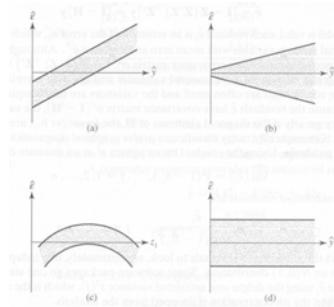
68

Adequacy of the Model

$$\begin{aligned}\hat{\varepsilon}_1 &= y_1 - \hat{\beta}_0 - \hat{\beta}_1 z_{11} - \cdots - \hat{\beta}_r z_{1r} \\ \hat{\varepsilon}_2 &= y_2 - \hat{\beta}_0 - \hat{\beta}_1 z_{21} - \cdots - \hat{\beta}_r z_{2r} \\ &\vdots \\ \hat{\varepsilon}_n &= y_n - \hat{\beta}_0 - \hat{\beta}_1 z_{n1} - \cdots - \hat{\beta}_r z_{nr} \\ \hat{\boldsymbol{\varepsilon}} &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{y} = [\mathbf{I} - \mathbf{H}]\mathbf{y} \\ \hat{\varepsilon}_j &\text{ is an estimate of } \varepsilon_j : N(0, \sigma^2)\end{aligned}$$

69

Residual Plots



70

Q-Q Plots and Histograms

- Used to detect the presence of unusual observations or severe departures from normality that may require special attention in the analysis
- If n is large, minor departures from normality will not greatly affect inferences about $\boldsymbol{\beta}$

71

Test of Independence of Time

Test constructed from the first autocorrelation

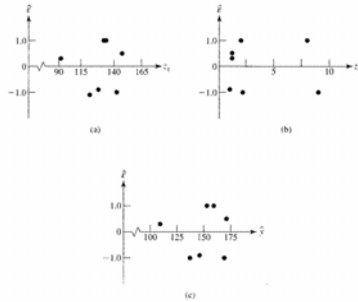
$$r_1 = \frac{\sum_{j=2}^n \hat{\varepsilon}_j \hat{\varepsilon}_{j-1}}{\sum_{j=1}^n \hat{\varepsilon}_j^2}$$

Durbin - Watson Test

$$\frac{\sum_{j=2}^n (\hat{\varepsilon}_j - \hat{\varepsilon}_{j-1})^2}{\sum_{j=1}^n \hat{\varepsilon}_j^2} \approx 2(1 - r_1)$$

72

Example 7.7: Residual Plot



73

Leverage

✦ “Outliers” in either the response or explanatory variables may have a considerable effect on the analysis and determine the fit

✦ Leverage for simple linear regression with one explanatory variable z

$$h_{jj} = \frac{1}{n} + \frac{(z_j - \bar{z})^2}{\sum_{j=1}^n (z_j - \bar{z})^2}, \quad \text{average} = \frac{r+1}{n}$$

74

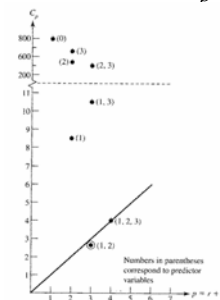
Mallow's C_p Statistic

✦ Select variables from all possible combinations

$$C_p = \left(\frac{\left(\begin{array}{l} \text{residual sum of squares for subset models} \\ \text{with } p \text{ parameters, including an intercept} \end{array} \right)}{\left(\begin{array}{l} \text{residual variance for full model} \end{array} \right)} \right) - (n - 2p)$$

75

Usage of Mallow's C_p Statistic



76

Stepwise Regression

- ✦ 1. The predictor variable that explains the largest significant proportion of the variation in Y is the first variable to enter
- ✦ 2. The next to enter is the one that makes the highest contribution to the regression sum of squares. Use Result 7.6 to determine the significance (F -test)

77

Stepwise Regression

- ✦ 3. Once a new variable is included, the individual contributions to the regression sum of squares of the other variables already in the equation are checked using F -tests. If the F -statistic is small, the variable is deleted
- ✦ 4. Steps 2 and 3 are repeated until all possible additions are non-significant and all possible deletions are significant

78

Treatment of Colinearity

- ✦ If $\mathbf{Z}$ is not of full rank, $\mathbf{Z}'\mathbf{Z}$ does not have an inverse $\rightarrow$ Colinear
- ✦ Not likely to have exact colinearity
- ✦ Possible to have a linear combination of columns of $\mathbf{Z}$ that are nearly 0
- ✦ Can be overcome somewhat by
 - Delete one of a pair of predictor variables that are strongly correlated
 - Relate the response Y to the principal components of the predictor variables

79

Bias Caused by a Misspecified Model

$$\mathbf{Z} = [\mathbf{Z}_1 \quad \mathbf{Z}_2]$$

$$\mathbf{Y} = [\mathbf{Z}_1 \quad \mathbf{Z}_2] \begin{bmatrix} \beta_{(1)} \\ \beta_{(2)} \end{bmatrix} + \boldsymbol{\varepsilon} = \mathbf{Z}_1 \beta_{(1)} + \mathbf{Z}_2 \beta_{(2)} + \boldsymbol{\varepsilon}$$

$$\hat{\beta}_{(1)} = (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' \mathbf{Y}$$

$$E(\hat{\beta}_{(1)}) = (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' E(\mathbf{Y}) = (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' (\mathbf{Z}_1 \beta_{(1)} + \mathbf{Z}_2 \beta_{(2)})$$

$$= \beta_{(1)} + (\mathbf{Z}_1' \mathbf{Z}_1)^{-1} \mathbf{Z}_1' \mathbf{Z}_2 \beta_{(2)}$$

biased estimator of $\beta_{(1)}$

80

Outline

- ✦ Model Checking and Other Aspects of Regression
- ✦ Multivariate Multiple Regression
- ✦ The Concept of Linear Regression
- ✦ Comparing the Two Formulations of the Regression Model
- ✦ Multiple Regression Models with Time Dependent Errors

81

Questions

- ✦ How to do multivariate multiple regression?
- ✦ What are the expectation of the estimated matrix of coefficients and the covariance matrix of the residuals? (Result 7.9)
- ✦ What is the forecast error?

82

Questions

- ✦ What is the maximum likelihood estimator of the matrix of coefficients? (Result 7.10)
- ✦ How to know that number of variables is enough in the multivariate multiple regression? (Result 7.11)
- ✦ How to do Predictions from Regressions?

83

Example 7.8

- ✦ Observed data

z_1	0	1	2	3	4
y_1	1	4	3	8	9
y_2	-1	-1	2	3	2

- ✦ Regression model

$$Y_1 = \beta_{01} + \beta_{11}z_1 + \varepsilon_1$$

$$Y_2 = \beta_{02} + \beta_{12}z_1 + \varepsilon_2$$

84

Multivariate Multiple Regression

$$Y_1 = \beta_{01} + \beta_{11}z_1 + \dots + \beta_{r1}z_r + \varepsilon_1$$

$$Y_2 = \beta_{02} + \beta_{12}z_1 + \dots + \beta_{r2}z_r + \varepsilon_2$$

$$\vdots$$

$$Y_m = \beta_{0m} + \beta_{1m}z_1 + \dots + \beta_{rm}z_r + \varepsilon_m$$

$$\boldsymbol{\varepsilon}' = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_m], \quad E(\boldsymbol{\varepsilon}) = 0, \quad \text{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$$

$$\mathbf{Y}'_j = [Y_{j1} \quad Y_{j2} \quad \dots \quad Y_{jm}], \quad \boldsymbol{\varepsilon}'_j = [\varepsilon_{j1} \quad \varepsilon_{j2} \quad \dots \quad \varepsilon_{jm}]$$

$$\mathbf{Z} = \begin{bmatrix} z_{10} & z_{11} & \dots & z_{1r} \\ z_{20} & z_{21} & \dots & z_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n0} & z_{n1} & \dots & z_{nr} \end{bmatrix}$$

85

Multivariate Multiple Regression

$$\mathbf{Y} = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1m} \\ Y_{21} & Y_{22} & \dots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nm} \end{bmatrix} = [\mathbf{Y}_{(1)} \quad \mathbf{Y}_{(2)} \quad \dots \quad \mathbf{Y}_{(m)}]$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{01} & \beta_{02} & \dots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \dots & \beta_{rm} \end{bmatrix} = [\boldsymbol{\beta}_{(1)} \quad \boldsymbol{\beta}_{(2)} \quad \dots \quad \boldsymbol{\beta}_{(m)}]$$

86

Multivariate Multiple Regression

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \dots & \varepsilon_{nm} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varepsilon}_{(1)} & \boldsymbol{\varepsilon}_{(2)} & \dots & \boldsymbol{\varepsilon}_{(m)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\varepsilon}'_1 \\ \boldsymbol{\varepsilon}'_2 \\ \vdots \\ \boldsymbol{\varepsilon}'_n \end{bmatrix}$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$E(\boldsymbol{\varepsilon}_{(i)}) = 0, \quad \text{Cov}(\boldsymbol{\varepsilon}_{(i)}, \boldsymbol{\varepsilon}_{(k)}) = \sigma_{ik}\mathbf{I}, \quad i, k = 1, 2, \dots, m$$

$$\boldsymbol{\Sigma} = \{\sigma_{ik}\}$$

87

Multivariate Multiple Regression

$$\mathbf{Y}_{(i)} = \mathbf{Z}\boldsymbol{\beta}_{(i)} + \boldsymbol{\varepsilon}_{(i)}, \quad \text{Cov}(\boldsymbol{\varepsilon}_{(i)}) = \sigma_{ii}\mathbf{I}, \quad i = 1, 2, \dots, m$$

$$\hat{\boldsymbol{\beta}}_{(i)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(i)}$$

$$\hat{\boldsymbol{\beta}} = [\hat{\boldsymbol{\beta}}_{(1)} \quad \hat{\boldsymbol{\beta}}_{(2)} \quad \dots \quad \hat{\boldsymbol{\beta}}_{(m)}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathbf{Y}_{(1)} \quad \mathbf{Y}_{(2)} \quad \dots \quad \mathbf{Y}_{(m)}]$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}, \quad \mathbf{B} = [\mathbf{b}_{(1)} \quad \mathbf{b}_{(2)} \quad \dots \quad \mathbf{b}_{(m)}]$$

$$(\mathbf{Y} - \mathbf{Z}\mathbf{B})(\mathbf{Y} - \mathbf{Z}\mathbf{B})'$$

$$\begin{bmatrix} (\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)})(\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)})' & \dots & (\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)})(\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)})' \\ \vdots & & \vdots \\ (\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)})(\mathbf{Y}_{(1)} - \mathbf{Z}\mathbf{b}_{(1)})' & \dots & (\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)})(\mathbf{Y}_{(m)} - \mathbf{Z}\mathbf{b}_{(m)})' \end{bmatrix}$$

88

Multivariate Multiple Regression

$$\mathbf{b}_{(i)} = \hat{\boldsymbol{\beta}}_{(i)} \text{ minimizes } (\mathbf{Y}_{(i)} - \mathbf{Z}\mathbf{b}_{(i)})(\mathbf{Y}_{(i)} - \mathbf{Z}\mathbf{b}_{(i)})'$$

$$\therefore \text{tr}[(\mathbf{Y} - \mathbf{Z}\mathbf{B})(\mathbf{Y} - \mathbf{Z}\mathbf{B})'] \text{ is minimized by } \mathbf{B} = \hat{\boldsymbol{\beta}}$$

$$\text{Generalized variance } |(\mathbf{Y} - \mathbf{Z}\mathbf{B})(\mathbf{Y} - \mathbf{Z}\mathbf{B})'| \text{ is also minimized by } \mathbf{B} = \hat{\boldsymbol{\beta}}$$

89

Multivariate Multiple Regression

$$\text{Predicted values: } \hat{\mathbf{Y}} = \mathbf{Z}\hat{\boldsymbol{\beta}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$$

$$\text{Residuals: } \hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y}$$

$$\mathbf{Z}'\hat{\boldsymbol{\varepsilon}} = \mathbf{Z}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y} = \mathbf{0}$$

$$\hat{\mathbf{Y}}'\hat{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\beta}}'\mathbf{Z}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y} = \mathbf{0}$$

$$\mathbf{Y}'\mathbf{Y} = (\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}})(\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}})' = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}$$

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\beta}}$$

90

Example 7.8

$$\mathbf{Z}' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 4 \end{bmatrix}, \quad (\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 0.6 & -0.2 \\ -0.2 & 0.1 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}}_{(1)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_{(1)} = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}}_{(2)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_{(2)} = \begin{bmatrix} -1 & 1 \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{(1)} & \hat{\boldsymbol{\beta}}_{(2)} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \begin{bmatrix} \mathbf{y}_{(1)} & \mathbf{y}_{(2)} \end{bmatrix}$$

$$\hat{y}_1 = 1 + 2z_1, \quad \hat{y}_2 = -1 + z_2$$

91

Example 7.8

$$\hat{\mathbf{Y}} = \mathbf{Z}\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 3 & 0 \\ 5 & 1 \\ 7 & 2 \\ 9 & 3 \end{bmatrix}$$

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = \begin{bmatrix} 0 & 1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 1 & -1 \end{bmatrix}, \quad \hat{\boldsymbol{\varepsilon}}'\hat{\mathbf{Y}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{Y}'\mathbf{Y} = \begin{bmatrix} 171 & 43 \\ 43 & 19 \end{bmatrix}, \quad \hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \begin{bmatrix} 165 & 45 \\ 45 & 15 \end{bmatrix}, \quad \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \begin{bmatrix} 6 & -2 \\ -2 & 4 \end{bmatrix}$$

$$\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \text{ is verified}$$

92

Result 7.9

$$E(\hat{\boldsymbol{\beta}}_{(i)}) = \boldsymbol{\beta}_{(i)} \text{ or } E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

$$\text{Cov}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(\mathbf{Z}'\mathbf{Z})^{-1}$$

$$\hat{\boldsymbol{\varepsilon}} = \begin{bmatrix} \hat{\boldsymbol{\varepsilon}}_{(1)} & \hat{\boldsymbol{\varepsilon}}_{(2)} & \cdots & \hat{\boldsymbol{\varepsilon}}_{(m)} \end{bmatrix} = \mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}$$

$$E(\hat{\boldsymbol{\varepsilon}}_{(i)}) = \mathbf{0}, \quad E(\hat{\boldsymbol{\varepsilon}}_{(i)}\hat{\boldsymbol{\varepsilon}}_{(k)}) = (n-r-1)\sigma_{ik}$$

$$E(\hat{\boldsymbol{\varepsilon}}) = \mathbf{0}, \quad E\left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n-r-1}\right) = \boldsymbol{\Sigma}$$

$$\hat{\boldsymbol{\varepsilon}} \text{ and } \hat{\boldsymbol{\beta}} \text{ are uncorrelated}$$

93

Proof of Result 7.9

$$\mathbf{Y}_{(i)} = \mathbf{Z}\boldsymbol{\beta}_{(i)} + \boldsymbol{\varepsilon}_{(i)}, \quad E(\boldsymbol{\varepsilon}_{(i)}) = \mathbf{0}, \quad E(\boldsymbol{\varepsilon}_{(i)}\boldsymbol{\varepsilon}_{(i)}') = \sigma_{ii}\mathbf{I}$$

$$\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(i)} - \boldsymbol{\beta}_{(i)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}_{(i)}$$

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}_{(i)} &= \mathbf{Y}_{(i)} - \hat{\mathbf{Y}}_{(i)} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y}_{(i)} \\ &= [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon}_{(i)} \end{aligned}$$

$$E(\hat{\boldsymbol{\beta}}_{(i)}) = \boldsymbol{\beta}_{(i)}, \quad E(\hat{\boldsymbol{\varepsilon}}_{(i)}) = \mathbf{0}$$

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(k)}) &= E(\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)})(\hat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})' \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E(\boldsymbol{\varepsilon}_{(i)}\boldsymbol{\varepsilon}_{(k)}')\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma_{ik}(\mathbf{Z}'\mathbf{Z})^{-1} \end{aligned}$$

94

Proof of Result 7.9

$\mathbf{U}$: random vector, $\mathbf{A}$: fixed matrix

$$E(\mathbf{U}'\mathbf{A}\mathbf{U}) = E[\text{tr}(\mathbf{U}'\mathbf{A}\mathbf{U})] = \text{tr}[\mathbf{A}E(\mathbf{U}\mathbf{U}')] = \text{tr}[\mathbf{A}\boldsymbol{\Sigma}]$$

$$E(\hat{\boldsymbol{\varepsilon}}_{(i)}\hat{\boldsymbol{\varepsilon}}_{(k)}) = E(\boldsymbol{\varepsilon}_{(i)}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\boldsymbol{\varepsilon}_{(k)})$$

$$= \text{tr}[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')\sigma_{ik}\mathbf{I}]$$

$$= \sigma_{ik} \text{tr}[(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')] = \sigma_{ik}(n-r-1)$$

$$E\left(\frac{\hat{\boldsymbol{\varepsilon}}_{(i)}'\hat{\boldsymbol{\varepsilon}}_{(k)}}{n-r-1}\right) = \boldsymbol{\Sigma}$$

95

Proof of Result 7.9

$$\begin{aligned} \text{Cov}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\varepsilon}}_{(k)}) &= E((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}_{(i)}\boldsymbol{\varepsilon}_{(k)}'(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')) \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'E(\boldsymbol{\varepsilon}_{(i)}\boldsymbol{\varepsilon}_{(k)}')(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\sigma_{ik}\mathbf{I}(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') \\ &= \sigma_{ik}((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' - (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}') = \mathbf{0} \end{aligned}$$

96

Forecast Error

$$\begin{aligned} \mathbf{z}_0' &= [1 \quad z_{01} \quad \cdots \quad z_{0r}] \quad Y_{0i} = \mathbf{z}_0' \boldsymbol{\beta}_{(i)} + \varepsilon_{0i} \\ E(\mathbf{z}_0' \hat{\boldsymbol{\beta}}_{(i)}) &= \mathbf{z}_0' E(\hat{\boldsymbol{\beta}}_{(i)}) = \mathbf{z}_0' \boldsymbol{\beta}_{(i)}, \quad E(\mathbf{z}_0' \hat{\boldsymbol{\beta}}) = \mathbf{z}_0' \boldsymbol{\beta} \\ E[\mathbf{z}_0' (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}_{(i)}) (\hat{\boldsymbol{\beta}}_{(k)} - \hat{\boldsymbol{\beta}}_{(k)})' \mathbf{z}_0] &= \mathbf{z}_0' E[(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}}_{(i)}) (\hat{\boldsymbol{\beta}}_{(k)} - \hat{\boldsymbol{\beta}}_{(k)})'] \mathbf{z}_0 \\ &= \sigma_{ik} \mathbf{z}_0' (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{z}_0 \\ \boldsymbol{\varepsilon}_0' &= [\varepsilon_{01} \quad \varepsilon_{02} \quad \cdots \quad \varepsilon_{0m}] \text{ independent of } \boldsymbol{\varepsilon} \\ E(\varepsilon_{0i}) &= 0, \quad E(\varepsilon_{0i} \varepsilon_{0k}) = \sigma_{ik} \\ Y_{0i} - \mathbf{z}_0' \hat{\boldsymbol{\beta}}_{(i)} &= Y_{0i} - \mathbf{z}_0' \boldsymbol{\beta}_{(i)} + \mathbf{z}_0' \hat{\boldsymbol{\beta}}_{(i)} - \mathbf{z}_0' \boldsymbol{\beta}_{(i)} = \varepsilon_{0i} - \mathbf{z}_0' (\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)}) \\ E(Y_{0i} - \mathbf{z}_0' \hat{\boldsymbol{\beta}}_{(i)}) &= 0 \end{aligned}$$

97

Forecast Error

$$\begin{aligned} E(Y_{0i} - \mathbf{z}_0' \hat{\boldsymbol{\beta}}_{(i)}) (Y_{0k} - \mathbf{z}_0' \hat{\boldsymbol{\beta}}_{(k)}) &= E(\varepsilon_{0i} - \mathbf{z}_0' (\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)})) (\varepsilon_{0k} - \mathbf{z}_0' (\hat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})) \\ &= E(\varepsilon_{0i} \varepsilon_{0k}) + \mathbf{z}_0' E(\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)}) (\hat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)})' \mathbf{z}_0 \\ &\quad - \mathbf{z}_0' E(\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)}) \varepsilon_{0k} - E(\varepsilon_{0i} (\hat{\boldsymbol{\beta}}_{(k)} - \boldsymbol{\beta}_{(k)}))' \mathbf{z}_0 \\ &= \sigma_{ik} (1 + \mathbf{z}_0' (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{z}_0) \\ \hat{\boldsymbol{\beta}}_{(i)} &= \boldsymbol{\beta}_{(i)} + (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \boldsymbol{\varepsilon}_{(i)}, \quad E((\hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}_{(i)}) \varepsilon_{0k}) = 0 \end{aligned}$$

98

Result 7.10

$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}$: normal distribution
 $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$
 $\hat{\boldsymbol{\beta}}$: normal distribution with $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
 $\text{Cov}(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik} (\mathbf{Z}' \mathbf{Z})^{-1}$
 $\hat{\boldsymbol{\beta}}$ is independent of the maximum likelihood estimator of $\boldsymbol{\Sigma}$
 given by $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} = \frac{1}{n} (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}})$
 and $n\hat{\boldsymbol{\Sigma}}$ is distributed as $W_{m, n-r-1}(\boldsymbol{\Sigma})$

99

Result 7.11

$$\begin{aligned} \boldsymbol{\beta} &= \begin{bmatrix} \boldsymbol{\beta}_{(1)} \\ \vdots \\ \boldsymbol{\beta}_{(q+1)} \\ \vdots \\ \boldsymbol{\beta}_{(r)} \end{bmatrix}, \quad H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0} \\ n\hat{\boldsymbol{\Sigma}} &= (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}})' (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\beta}}) : W_{m, n-r-1}(\boldsymbol{\Sigma}) \text{ independent of } \\ n(\hat{\boldsymbol{\Sigma}}_1 - \hat{\boldsymbol{\Sigma}}) &: W_{m, r-q}(\boldsymbol{\Sigma}), \quad n\hat{\boldsymbol{\Sigma}}_1 = (\mathbf{Y} - \mathbf{Z}_1 \hat{\boldsymbol{\beta}}_{(1)})' (\mathbf{Y} - \mathbf{Z}_1 \hat{\boldsymbol{\beta}}_{(1)}) \\ \text{Reject } H_0 &\text{ for large values of} \\ -n \ln \Lambda &= -n \ln \left(\frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}|} \right) = -n \ln \frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}_1 + n(\hat{\boldsymbol{\Sigma}}_1 - \hat{\boldsymbol{\Sigma}})|} \\ \text{For } n \text{ large, } &-[n - r - 1 - (m - r + q + 1)/2] \ln \left(\frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}_1|} \right) \sim \chi_{m(r-q)}^2 \end{aligned}$$

100

Example 7.9

Example 7.5 data plus one more service - quality index.

$$\begin{aligned} n\hat{\boldsymbol{\Sigma}} &= \begin{bmatrix} 2977.39 & 1021.72 \\ 1021.72 & 2050.95 \end{bmatrix} \\ n(\hat{\boldsymbol{\Sigma}}_1 - \hat{\boldsymbol{\Sigma}}) &= \begin{bmatrix} 441.76 & 246.16 \\ 246.16 & 366.12 \end{bmatrix} \\ \boldsymbol{\beta}_{(2)} &: \text{locatin - gender interaction parameters} \\ H_0 : \boldsymbol{\beta}_{(2)} &= \mathbf{0}, \quad \alpha = 0.05, \\ r_1 &= \text{rank}(\mathbf{Z}) - 1 = 5, \quad q_1 = \text{rank}(\mathbf{Z}_1) - 1 = 3 \\ -[n - r_1 - 1 - (m - r_1 + q_1 + 1)/2] \ln \left(\frac{|\hat{\boldsymbol{\Sigma}}_1|}{|\hat{\boldsymbol{\Sigma}}_1 + n(\hat{\boldsymbol{\Sigma}}_1 - \hat{\boldsymbol{\Sigma}})|} \right) &= 3.28 \\ < \chi_{m(r_1 - q_1)}^2(0.05) &= 9.49, \text{ do not reject } H_0 \end{aligned}$$

101

Other Multivariate Test Statistics

$$\begin{aligned} \mathbf{E} &= n\hat{\boldsymbol{\Sigma}}, \quad \mathbf{H} = n(\hat{\boldsymbol{\Sigma}}_1 - \hat{\boldsymbol{\Sigma}}) \\ \eta_1 \geq \eta_2 \geq \cdots \geq \eta_s &: \text{eigenvalues of } \mathbf{H}\mathbf{E}^{-1}, \quad s = \min(p, r - q) \\ \text{Wilk's lambda} &= \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} = \prod_{i=1}^s \frac{1}{1 + \eta_i} \\ \text{Pillai's trace} &= \text{tr}(\mathbf{H}(\mathbf{H} + \mathbf{E})^{-1}) = \sum_{i=1}^s \frac{1}{1 + \eta_i} \\ \text{Hotelling - Lawlay trace} &= \text{tr}(\mathbf{H}\mathbf{E}^{-1}) = \sum_{i=1}^s \eta_i \\ \text{Roy's greatest root} &= \frac{\eta_1}{1 + \eta_1} \end{aligned}$$

102

Predictions from Regressions

Result 7.10 $\Rightarrow \hat{\beta}'z_0 : N_m(\beta'z_0, z_0'(Z'Z)^{-1}z_0\Sigma)$

$n\hat{\Sigma}$: independently distributed as $W_{n-r-1}(\Sigma)$

$$T^2 = \left(\frac{\hat{\beta}'z_0 - \beta'z_0}{\sqrt{z_0'(Z'Z)^{-1}z_0}} \right) \left(\frac{n\hat{\Sigma}}{n-r-1} \right)^{-1} \left(\frac{\hat{\beta}'z_0 - \beta'z_0}{\sqrt{z_0'(Z'Z)^{-1}z_0}} \right)$$

100(1- α)% confidence ellipsoid for $\beta'z_0$:

$$\left(\hat{\beta}'z_0 - \beta'z_0 \right) \left(\frac{n\hat{\Sigma}}{n-r-1} \right)^{-1} \left(\hat{\beta}'z_0 - \beta'z_0 \right) \leq z_0'(Z'Z)^{-1}z_0 \left[\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha) \right]$$

103

Predictions from Regressions

100(1- α)% simultaneous confidence intervals

for $E(Y_i) = z_0'\beta_{(i)}$:

$$z_0'\hat{\beta}_{(i)} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha) \right) z_0'(Z'Z)^{-1}z_0 \left(\frac{n}{n-r-1} \right) \hat{\sigma}_{ii}}$$

$\hat{\sigma}_{ii}$: the i th diagonal element of $\hat{\Sigma}$

104

Predictions from Regressions

$Y_0 = \beta'z_0 + \epsilon_0$,

$Y_0 - \hat{\beta}'z_0 = (\beta - \hat{\beta})'z_0 + \epsilon_0 : N_m(0, (1 + z_0'(Z'Z)^{-1}z_0)\Sigma)$

independently of $n\hat{\Sigma}$

100(1- α)% prediction ellipsoid for Y_0 :

$$\left(Y_0 - \hat{\beta}'z_0 \right) \left(\frac{n}{n-r-1} \hat{\Sigma} \right)^{-1} \left(Y_0 - \hat{\beta}'z_0 \right) \leq (1 + z_0'(Z'Z)^{-1}z_0) \left[\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha) \right]$$

100(1- α)% simultaneous prediction intervals for Y_{0i} :

$$z_0'\hat{\beta}_{(i)} \pm \sqrt{\left[\left(\frac{m(n-r-1)}{n-r-m} F_{m,n-r-m}(\alpha) \right) \left(1 + z_0'(Z'Z)^{-1}z_0 \right) \left(\frac{n}{n-r-1} \hat{\sigma}_{ii} \right) \right]}$$

106

Example 7.10

Example 7.6 data + Y_2 : disk I/O capacity

Fitted equation : $\hat{y}_2 = 14.14 + 2.25z_1 + 5.67z_2$, $s = 1.812$

$\hat{\beta}_{(2)} = [14.14 \quad 2.25 \quad 5.67]'$, $z_0'(Z'Z)^{-1}z_0 = 0.34725$

From Example 7.6, $\hat{\beta}_{(1)} = [8.42 \quad 1.08 \quad 42]'$

$z_0'\hat{\beta}_{(1)} = 151.97$, $z_0'\hat{\beta}_{(2)} = 349.17$

$$n\hat{\Sigma} = \begin{bmatrix} (y_{(1)} - Z\hat{\beta}_{(1)})(y_{(1)} - Z\hat{\beta}_{(1)})' & (y_{(1)} - Z\hat{\beta}_{(1)})(y_{(2)} - Z\hat{\beta}_{(2)})' \\ (y_{(2)} - Z\hat{\beta}_{(2)})(y_{(1)} - Z\hat{\beta}_{(1)})' & (y_{(2)} - Z\hat{\beta}_{(2)})(y_{(2)} - Z\hat{\beta}_{(2)})' \end{bmatrix} = \begin{bmatrix} 5.80 & 5.30 \\ 5.30 & 13.13 \end{bmatrix}$$

Example 7.10

$$\hat{\beta}'z_0 = \begin{bmatrix} \hat{\beta}_{(1)} \\ \hat{\beta}_{(2)} \end{bmatrix} z_0 = \begin{bmatrix} z_0'\hat{\beta}_{(1)} \\ z_0'\hat{\beta}_{(2)} \end{bmatrix} = \begin{bmatrix} 151.97 \\ 349.17 \end{bmatrix}$$

$n = 7$, $r = 2$, $m = 2$

95% confidence ellipse for $\beta'z_0$:

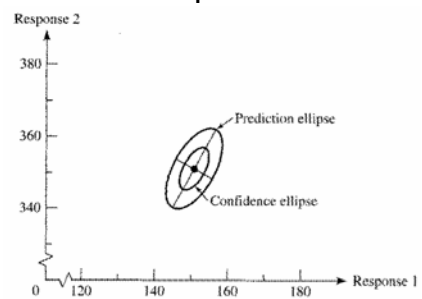
$$4[z_0'\hat{\beta}_{(1)} - 151.97, z_0'\hat{\beta}_{(2)} - 349.17] \begin{bmatrix} 5.80 & 5.30 \\ 5.30 & 13.13 \end{bmatrix}^{-1} \begin{bmatrix} z_0'\hat{\beta}_{(1)} - 151.97 \\ z_0'\hat{\beta}_{(2)} - 349.17 \end{bmatrix} \leq 0.34725 \left[\left(\frac{2 \times 4}{3} \right) F_{2,3}(0.05) \right]$$

95% prediction ellipse : replace $z_0'(Z'Z)^{-1}z_0 = 0.34725$ with

$$1 + z_0'(Z'Z)^{-1}z_0 = 1.34725$$

107

Example 7.10



108

Outline

- Model Checking and Other Aspects of Regression
- Multivariate Multiple Regression
- The Concept of Linear Regression
- Comparing the Two Formulations of the Regression Model
- Multiple Regression Models with Time Dependent Errors

109

Questions

- What are the results if the response Y is also treated as random in regression? (Result 7.12)
- What is the Population Multiple Correlation Coefficient?
- What is the maximum likelihood estimator if the response Y is also treated as random in regression? (Result 7.13, 7.14)

110

Questions

- What is the partial correlation coefficient?

111

Linear Regression

$Y, Z_1, Z_2, \dots, Z_r$: random variables

$f(y, z_1, z_2, \dots, z_r)$: not necessarily normal

mean μ and covariance matrix Σ :

$$\mu = \begin{bmatrix} \mu_Y \\ \mu_Z \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_{YY} & \sigma_{ZY}' \\ \sigma_{ZY} & \Sigma_{ZZ} \end{bmatrix}$$

linear predictor $= b_0 + b_1 Z_1 + b_2 Z_2 + \dots + b_r Z_r = b_0 + \mathbf{b}'\mathbf{Z}$

prediction error $= Y - b_0 - \mathbf{b}'\mathbf{Z}$

mean square error $= E(Y - b_0 - \mathbf{b}'\mathbf{Z})^2$

112

Result 7.12

Linear predictor $\beta_0 + \beta'\mathbf{Z}$ has minimum mean square among all linear predictors of the response Y

$$\beta_0 = \Sigma_{ZZ}^{-1} \sigma_{ZY}, \quad \beta_0 = \mu_Y - \beta'\mu_Z$$

$$E(Y - \beta_0 - \beta'\mathbf{Z})^2 = \sigma_{YY} - \sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}$$

Also, $\beta_0 + \beta'\mathbf{Z}$ is the linear predictor having maximum correlation with Y

$$\text{Corr}(Y, \beta_0 + \beta'\mathbf{Z}) = \max_{b_0, \mathbf{b}} \text{Corr}(Y, b_0 + \mathbf{b}'\mathbf{Z})$$

$$= \sqrt{\frac{\beta' \Sigma_{ZZ} \beta}{\sigma_{YY}}} = \sqrt{\frac{\sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}}{\sigma_{YY}}}$$

113

Proof of Result 7.12

$$b_0 + \mathbf{b}'\mathbf{Z} = b_0 + \mathbf{b}'\mathbf{Z} - (\mu_Y + \mathbf{b}'\mu_Z) + (\mu_Y + \mathbf{b}'\mu_Z)$$

$$E(Y - b_0 - \mathbf{b}'\mathbf{Z})^2$$

$$= E[Y - \mu_Y - \mathbf{b}'(\mathbf{Z} - \mu_Z) + (\mu_Y - b_0 - \mathbf{b}'\mu_Z)]^2$$

$$= E(Y - \mu_Y)^2 + E(\mathbf{b}'(\mathbf{Z} - \mu_Z))^2 + (\mu_Y - b_0 - \mathbf{b}'\mu_Z)^2 - 2E[\mathbf{b}'(\mathbf{Z} - \mu_Z)(Y - \mu_Y)]$$

$$= \sigma_{YY} + \mathbf{b}' \Sigma_{ZZ} \mathbf{b} + (\mu_Y - b_0 - \mathbf{b}'\mu_Z)^2 - 2\mathbf{b}' \sigma_{ZY}$$

$$= \sigma_{YY} + (\mathbf{b} - \Sigma_{ZZ}^{-1} \sigma_{ZY})' \Sigma_{ZZ} (\mathbf{b} - \Sigma_{ZZ}^{-1} \sigma_{ZY}) + (\mu_Y - b_0 - \mathbf{b}'\mu_Z)^2 - \sigma_{ZY}' \Sigma_{ZZ}^{-1} \sigma_{ZY}$$

$$\text{minimized at } \mathbf{b} = \Sigma_{ZZ}^{-1} \sigma_{ZY} = \beta, \quad b_0 = \mu_Y - \beta'\mu_Z = \mu_Y - (\Sigma_{ZZ}^{-1} \sigma_{ZY})'\mu_Z$$

114

Proof of Result 7.12

$$\text{Cov}(b_0 + \mathbf{b}'\mathbf{Z}, Y) = \mathbf{b}'\boldsymbol{\sigma}_{ZY}$$

$$[\text{Corr}(b_0 + \mathbf{b}'\mathbf{Z}, Y)]^2 = \frac{|\mathbf{b}'\boldsymbol{\sigma}_{ZY}|^2}{\sigma_{YY}(\mathbf{b}'\boldsymbol{\Sigma}_{ZZ}\mathbf{b})}$$

Extended Cauchy - Schwartz inequality

$$(\mathbf{b}'\boldsymbol{\sigma}_{ZY})^2 \leq \mathbf{b}'\boldsymbol{\Sigma}_{ZZ}\mathbf{b}\boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY}$$

$$[\text{Corr}(b_0 + \mathbf{b}'\mathbf{Z}, Y)]^2 \leq \frac{\boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY}}{\sigma_{YY}}$$

with equality for $\mathbf{b} = \boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY} = \boldsymbol{\beta}$

$$\boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY} = \boldsymbol{\sigma}_{ZY}'\boldsymbol{\beta} = \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta} = \boldsymbol{\beta}'\boldsymbol{\Sigma}_{ZZ}\boldsymbol{\beta}$$

115

Population Multiple Correlation Coefficient

Population multiple correlation coefficient :

$$\rho_{Y(Z)} = + \sqrt{\frac{\boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY}}{\sigma_{YY}}}$$

Population coefficient of determination : $\rho_{Y(Z)}^2$

Mean square error in using $\beta_0 + \boldsymbol{\beta}'\mathbf{Z}$ to forecast Y :

$$\sigma_{YY} - \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY} = \sigma_{YY}(1 - \rho_{Y(Z)}^2)$$

$$\rho_{Y(Z)}^2 = 0 : \text{no predictive power in } \mathbf{Z}$$

$$\rho_{Y(Z)}^2 = 1 : Y \text{ can be predicted with no error}$$

116

Example 7.11

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \boldsymbol{\mu}_Z \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{YY} & \boldsymbol{\sigma}_{ZY}' \\ \boldsymbol{\sigma}_{ZY} & \boldsymbol{\Sigma}_{ZZ} \end{bmatrix} = \begin{bmatrix} 10 & 1 & -1 \\ - & + & - \\ 1 & 7 & 3 \\ -1 & 3 & 2 \end{bmatrix}$$

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY} = \begin{bmatrix} 1 & -2 \end{bmatrix}, \quad \beta_0 = \mu_Y - \boldsymbol{\beta}'\boldsymbol{\mu}_Z = 3$$

best linear predictor : $\beta_0 + \boldsymbol{\beta}'\mathbf{Z} = 3 + Z_1 - 2Z_2$

$$\text{mean square error} = \sigma_{YY} - \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY} = 7$$

$$\rho_{Y(Z)} = \sqrt{\frac{\boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY}}{\sigma_{YY}}} = \sqrt{\frac{3}{10}} = 0.548$$

117

Linear Predictors and Normality

$$[Y \ Z_1 \ Z_2 \ \cdots \ Z_r] : N_{r+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Conditional distribution of Y with $\mathbf{Z} = \mathbf{z}$:

$$N(\mu_Y + \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}(\mathbf{z} - \boldsymbol{\mu}_Z), \sigma_{YY} - \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY})$$

$$E(Y | \mathbf{z}) = \mu_Y + \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}(\mathbf{z} - \boldsymbol{\mu}_Z) = \beta_0 + \boldsymbol{\beta}'\mathbf{z}$$

(linear regression function)

When the population is not normal, $E(Y | \mathbf{z})$

need not be linear. Nevertheless, it still predicts

Y with the smallest mean square error.

118

Result 7.13

Joint distribution of Y and $\mathbf{Z}$: $N_{r+1}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$\hat{\boldsymbol{\mu}} = \begin{bmatrix} \bar{Y} \\ \bar{\mathbf{Z}} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} s_{YY} & \mathbf{s}_{ZY}' \\ \mathbf{s}_{ZY} & \mathbf{S}_{ZZ} \end{bmatrix}$$

maximum likelihood estimator of the coefficients

$$\hat{\boldsymbol{\beta}} = \mathbf{S}_{ZZ}^{-1}\mathbf{s}_{ZY}, \quad \hat{\beta}_0 = \bar{Y} - \mathbf{s}_{ZY}'\mathbf{S}_{ZZ}^{-1}\bar{\mathbf{Z}} = \bar{Y} - \hat{\boldsymbol{\beta}}'\bar{\mathbf{Z}}$$

maximum likelihood estimator

$$\hat{\beta}_0 + \hat{\boldsymbol{\beta}}'\mathbf{z} = \bar{Y} + \mathbf{s}_{ZY}'\mathbf{S}_{ZZ}^{-1}(\mathbf{z} - \bar{\mathbf{Z}})$$

maximum likelihood estimator of $E[Y - \beta_0 - \boldsymbol{\beta}'\mathbf{Z}]^2$

$$\hat{\sigma}_{YY \cdot \mathbf{Z}} = \frac{n-1}{n}(s_{YY} - \mathbf{s}_{ZY}'\mathbf{S}_{ZZ}^{-1}\mathbf{s}_{ZY})$$

119

Proof of Result 7.13

Apply the invariance property of maximum likelihood, and

$$\beta_0 = \mu_Y - (\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY})'\boldsymbol{\mu}_Z, \quad \boldsymbol{\beta} = \boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY},$$

$$\beta_0 + \boldsymbol{\beta}'\mathbf{z} = \mu_Y + \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}(\mathbf{z} - \boldsymbol{\mu}_Z)$$

$$\text{mean square error} = \sigma_{YY \cdot \mathbf{Z}} = \sigma_{YY} - \boldsymbol{\sigma}_{ZY}'\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\sigma}_{ZY}$$

to get the conclusion by substitution of the maximum likelihood

estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}} = \left(\frac{n-1}{n}\right)\mathbf{S}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

Unbiased estimator for $\sigma_{YY \cdot \mathbf{Z}}$:

$$\left(\frac{n-1}{n-r-1}\right)(s_{YY} - \mathbf{s}_{ZY}'\mathbf{S}_{ZZ}^{-1}\mathbf{s}_{ZY}) = \frac{\sum_{j=1}^n (Y_j - \hat{\beta}_0 - \hat{\boldsymbol{\beta}}'\mathbf{Z}_j)^2}{n-r-1}$$

120

Invariance Property

$\hat{\theta}$: maximum likelihood estimator of θ

$h(\hat{\theta})$: maximum likelihood estimator of $h(\theta)$

Examples:

MLE of $\mu' \Sigma^{-1} \mu = \hat{\mu}' \hat{\Sigma}^{-1} \hat{\mu}$

MLE of $\sqrt{\sigma_{ii}} = \sqrt{\hat{\sigma}_{ii}}$

$\hat{\sigma}_{ii} = \frac{1}{n} \sum_{j=1}^n (X_{ji} - \bar{X}_i)^2 = \text{MLE of } \text{Var}(X_i)$

121

Example 7.12

Example 7.6 data, $n = 7$ observations on Y, Z_1, Z_2

$$\hat{\mu} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} 150.44 \\ 130.24 \\ 3.547 \end{bmatrix},$$

$$\mathbf{S} = \begin{bmatrix} s_{YY} & | & s_{ZY} \\ s_{ZY} & | & s_{ZZ} \end{bmatrix} = \begin{bmatrix} 467.913 & | & 418.763 & 35.983 \\ \text{---} & | & \text{---} & \text{---} \\ 418.763 & | & 377.200 & 28.034 \\ 35.983 & | & 28.034 & 13.657 \end{bmatrix}$$

122

Example 7.12

$$\hat{\beta} = \mathbf{S}_{ZZ}^{-1} \mathbf{s}_{ZY} = \begin{bmatrix} 1.079 \\ 0.420 \end{bmatrix}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}' \bar{z} = 8.421$$

estimated regression function

$$\hat{\beta}_0 + \hat{\beta}' z = 8.42 - 1.08 z_1 + 0.42 z_2$$

maximum likelihood estimate of the mean square error

$$\left(\frac{n-1}{n} \right) (s_{YY} - \mathbf{s}_{ZY}' \mathbf{S}_{ZZ}^{-1} \mathbf{s}_{ZY}) = 0.894$$

123

Prediction of Several Variables

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{Z} \end{bmatrix} : N_{m+r}(\mu, \Sigma), \quad \mu = \begin{bmatrix} \mu_Y \\ \mu_Z \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{YY} & | & \Sigma_{YZ} \\ \Sigma_{ZY} & | & \Sigma_{ZZ} \end{bmatrix}$$

multivariate regression of $\mathbf{Y}$ and $\mathbf{Z}$:

$$E[\mathbf{Y} | \mathbf{z}] = \mu_Y + \Sigma_{YZ} \Sigma_{ZZ}^{-1} (\mathbf{z} - \mu_Z)$$

composed of m univariate regressions. For example,

$$E[Y_i | \mathbf{z}] = \mu_{Y_i} + \Sigma_{Y_i Z} \Sigma_{ZZ}^{-1} (\mathbf{z} - \mu_Z)$$

minimizes the mean square error for the prediction of Y_i

$\beta = \Sigma_{YZ} \Sigma_{ZZ}^{-1}$: matrix of regression coefficients

124

Result 7.14

$\mathbf{Y}$ and $\mathbf{Z}$: $N_{m+r}(\mu, \Sigma)$

regression of $\mathbf{Y}$ and $\mathbf{Z}$: $\beta_0 + \beta z = \mu_Y + \Sigma_{YZ} \Sigma_{ZZ}^{-1} (\mathbf{z} - \mu_Z)$

$$E(\mathbf{Y} - \beta_0 - \beta \mathbf{Z})(\mathbf{Y} - \beta_0 - \beta \mathbf{Z})' = \Sigma_{YY \cdot Z} = \Sigma_{YY} - \Sigma_{YZ} \Sigma_{ZZ}^{-1} \Sigma_{ZY}$$

maximum likelihood estimators

$$\hat{\beta}_0 + \hat{\beta} z = \bar{Y} + \mathbf{S}_{YZ} \mathbf{S}_{ZZ}^{-1} (\mathbf{z} - \bar{Z})$$

$$\hat{\Sigma}_{YY \cdot Z} = \left(\frac{n-1}{n} \right) (\mathbf{S}_{YY} - \mathbf{S}_{YZ} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{ZY})$$

125

Example 7.13

Data of Example 7.6 and 7.10.

$$\hat{\mu} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} 150.44 \\ 327.79 \\ 130.24 \\ 3.547 \end{bmatrix}$$

$$\mathbf{S} = \begin{bmatrix} 467.913 & 1148.556 & | & 418.763 & 35.983 \\ 1148.556 & 3072.491 & | & 1008.976 & 140.558 \\ \text{---} & \text{---} & | & \text{---} & \text{---} \\ 418.763 & 1008.976 & | & 377.200 & 28.034 \\ 35.983 & 140.558 & | & 28.034 & 13.657 \end{bmatrix}$$

126

Example 7.13

$$\begin{aligned}\hat{\beta}_0 + \hat{\beta}z &= \bar{y} + S_{YZ}S_{ZZ}^{-1}(z - \bar{z}) \\ &= \begin{bmatrix} 150.44 \\ 327.79 \end{bmatrix} + \begin{bmatrix} 1.079(z_1 - 130.24) + 0.420(z_2 - 3.547) \\ 2.254(z_1 - 130.24) + 5.665(z_2 - 3.547) \end{bmatrix} \\ &\text{maximum likelihood estimate of } \Sigma_{YY \cdot Z} \\ \left(\frac{n-1}{n} \right) (S_{YY} - S_{YZ}S_{ZZ}^{-1}S_{ZY}) &= \begin{bmatrix} 0.894 & 0.893 \\ 0.893 & 2.205 \end{bmatrix}\end{aligned}$$

127

Partial Correlation Coefficient

Pair of errors :

$$Y_1 - \mu_{Y_1} - \Sigma_{Y_1Z}\Sigma_{ZZ}^{-1}(Z - \mu_Z), Y_2 - \mu_{Y_2} - \Sigma_{Y_2Z}\Sigma_{ZZ}^{-1}(Z - \mu_Z)$$

Error covariance matrix

$$\Sigma_{YY \cdot Z} = \Sigma_{YY} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}$$

Partial correlation coefficient between Y_1 and Y_2 , eliminating the effects of Z :

$$\rho_{Y_1Y_2 \cdot Z} = \frac{\sigma_{Y_1Y_2 \cdot Z}}{\sqrt{\sigma_{Y_1Y_1 \cdot Z}} \sqrt{\sigma_{Y_2Y_2 \cdot Z}}}$$

Sample partial correlation coefficient

$$r_{Y_1Y_2 \cdot Z} = \frac{s_{Y_1Y_2 \cdot Z}}{\sqrt{s_{Y_1Y_1 \cdot Z}} \sqrt{s_{Y_2Y_2 \cdot Z}}}$$

128

Example 7.14

Example 7.13 data

$$S_{YY} - S_{YZ}S_{ZZ}^{-1}S_{ZY} = \begin{bmatrix} 1.043 & 1.042 \\ 1.042 & 2.572 \end{bmatrix}$$

$$r_{Y_1Y_2 \cdot Z} = \frac{s_{Y_1Y_2 \cdot Z}}{\sqrt{s_{Y_1Y_1 \cdot Z}} \sqrt{s_{Y_2Y_2 \cdot Z}}} = \frac{1.042}{\sqrt{1.043} \sqrt{2.572}} = 0.64$$

$$r_{Y_1Y_2} = 0.96$$

Correlation between Y_1 and Y_2 has been sharply reduced after eliminating the effects of Z on both responses

129

Outline

- Model Checking and Other Aspects of Regression
- Multivariate Multiple Regression
- The Concept of Linear Regression
- Comparing the Two Formulations of the Regression Model
- Multiple Regression Models with Time Dependent Errors

130

Questions

- What is the mean corrected form for multivariate multiple regressions?
- Compare the classical regression model and the approach that treats the result as a conditional expectation?

131

Mean Corrected Form of the Regression Model

$$\begin{aligned}Y_j &= \beta_0 + \beta_1 z_{j1} + \cdots + \beta_r z_{jr} + \varepsilon_j \\ &= \beta_0 + \beta_1(z_{j1} - \bar{z}_1) + \cdots + \beta_r(z_{jr} - \bar{z}_r) + \varepsilon_j\end{aligned}$$

mean corrected design matrix

$$Z_c = \begin{bmatrix} 1 & z_{11} - \bar{z}_1 & \cdots & z_{1r} - \bar{z}_r \\ 1 & z_{21} - \bar{z}_1 & \cdots & z_{2r} - \bar{z}_r \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} - \bar{z}_1 & \cdots & z_{nr} - \bar{z}_r \end{bmatrix} = [1 | Z_{c2}], \quad Z_{c2}'1 = 0$$

$$Z_c'Z_c = \begin{bmatrix} 1'1 & 1'Z_{c2} \\ Z_{c2}'1 & Z_{c2}'Z_{c2} \end{bmatrix} = \begin{bmatrix} n & 0' \\ 0 & Z_{c2}'Z_{c2} \end{bmatrix}$$

132

Mean Corrected Form of the Regression Model

$$\begin{aligned} \begin{bmatrix} \hat{\beta}_* \\ \hat{\beta}_c \end{bmatrix} &= (\mathbf{Z}'_c \mathbf{Z}_c)^{-1} \mathbf{Z}'_c \mathbf{y} \\ &= \begin{bmatrix} 1/n & \mathbf{0}' \\ \mathbf{0} & (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}' \mathbf{y} \\ \mathbf{Z}'_{c2} \mathbf{y} \end{bmatrix} = \begin{bmatrix} \bar{y} \\ (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} \mathbf{Z}'_{c2} \mathbf{y} \end{bmatrix} \\ \hat{y} &= \hat{\beta}_* + \hat{\beta}'_c (\mathbf{z} - \bar{\mathbf{z}}) = \bar{y} + \mathbf{y}' \mathbf{Z}_{c2} (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \\ \begin{bmatrix} \text{Var}(\hat{\beta}_*) & \text{Cov}(\hat{\beta}_*, \hat{\beta}_c) \\ \text{Cov}(\hat{\beta}_c, \hat{\beta}_*) & \text{Cov}(\hat{\beta}_c) \end{bmatrix} &= (\mathbf{Z}'_c \mathbf{Z}_c)^{-1} \sigma^2 \\ &= \begin{bmatrix} \sigma^2/n & \mathbf{0}' \\ \mathbf{0} & (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} \sigma^2 \end{bmatrix} \end{aligned}$$

133

Mean Corrected Form for Multivariate Multiple Regressions

least square estimates of the coefficient vectors for the i th response :

$$\begin{aligned} \hat{\beta}_{(i)} &= \begin{bmatrix} \bar{y}_{(i)} \\ (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} \mathbf{Z}'_{c2} \mathbf{y}_{(i)} \end{bmatrix} \\ \text{standardized input variables} \\ (z_{ji} - \bar{z}_i) / \sqrt{(n-1)s_{z_i z_i}} \\ \tilde{\beta}_i &= \beta_i \sqrt{(n-1)s_{z_i z_i}} \\ \hat{\tilde{\beta}}_i &= \hat{\beta}_i \sqrt{(n-1)s_{z_i z_i}} \end{aligned}$$

134

Relating the Formulations

Result 7.13: $\hat{\beta}_0 + \hat{\beta}'_c \mathbf{z} = \bar{y} + \mathbf{s}'_{ZY} \mathbf{S}_{ZZ}^{-1} (\mathbf{z} - \bar{\mathbf{z}})$

mean corrected form :

$$\begin{aligned} \hat{y} &= \hat{\beta}_* + \hat{\beta}'_c (\mathbf{z} - \bar{\mathbf{z}}) = \bar{y} + \mathbf{y}' \mathbf{Z}_{c2} (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} (\mathbf{z} - \bar{\mathbf{z}}) \\ \hat{\beta}_* &= \bar{y} = \hat{\beta}_0, \quad \hat{\beta}'_c = \mathbf{y}' \mathbf{Z}_{c2} (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} = \mathbf{s}'_{ZY} \mathbf{S}_{ZZ}^{-1} = \hat{\beta}'_c \\ \because \mathbf{y}' \mathbf{Z}_{c2} &= (\mathbf{y} - \bar{y} \mathbf{1})' \mathbf{Z}_{c2} + \bar{y} \mathbf{1}' \mathbf{Z}_{c2} = (\mathbf{y} - \bar{y} \mathbf{1})' \mathbf{Z}_{c2} \\ \mathbf{y}' \mathbf{Z}_{c2} (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} &= (\mathbf{y} - \bar{y} \mathbf{1})' \mathbf{Z}_{c2} (\mathbf{Z}'_{c2} \mathbf{Z}_{c2})^{-1} \\ &= (n-1) \mathbf{s}'_{ZY} [(n-1) \mathbf{S}_{ZZ}]^{-1} = \mathbf{s}'_{ZY} \mathbf{S}_{ZZ}^{-1} \end{aligned}$$

135

Example 7.15

- ✦ Example 7.6, classical linear regression model
- ✦ Example 7.12, joint normal distribution, best predictor as the conditional mean
- ✦ Both approaches yielded the same predictor of Y_1

$$\hat{y} = 8.42 + 1.08z_1 + 0.42z_2$$

136

Remarks on Both Formulation

- ✦ Conceptually different
- ✦ Classical model
 - Input variables are set by experimenter
 - Optimal among linear predictors
- ✦ Conditional mean model
 - Predictor values are random variables observed with the response values
 - Optimal among all choices of predictors

137

Outline

- ✦ Model Checking and Other Aspects of Regression
- ✦ Multivariate Multiple Regression
- ✦ The Concept of Linear Regression
- ✦ Comparing the Two Formulations of the Regression Model
- ✦ Multiple Regression Models with Time Dependent Errors

138

Example 7.16 Natural Gas Data

Y Sendout	Z_1 DHD	Z_2 DHDLag	Z_3 Windspeed	Z_4 Weekend
227	32	30	12	1
236	31	32	8	1
228	30	31	8	0
252	34	30	8	0
238	28	34	12	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
333	46	41	8	0
266	33	46	8	0
280	38	33	18	0
386	52	38	22	0
415	57	52	18	0

139

Example 7.16 : First Model

$$\text{Sendout} = 1.858 + 5.874 \text{ DHD} + 1.405 \text{ DHDLag} \\ + 1.315 \text{ Windspeed} - 15.857 \text{ Weekend}$$

$$R^2 = 0.952$$

All coefficients are significant, except the intercept

But,

$$\text{lag1 autocorrelation} = r_1(\hat{\varepsilon}) = \frac{\sum_{j=2}^n \hat{\varepsilon}_j \hat{\varepsilon}_{j-1}}{\sum_{j=1}^n \hat{\varepsilon}_j^2} = 0.52$$

140

Example 7.16 : Second Model

Replace the independent errors with an autoregressive noise

$$N_j = \phi_j N_{j-1} + \phi_7 N_{j-7} + \varepsilon_j$$

Apply SAS to get a fitted model as

$$\text{Sendout} = 2.130 + 5.810 \text{ DHD} + 1.426 \text{ DHDLag} \\ + 1.207 \text{ Windspeed} - 10.109 \text{ Weekend}$$

$$N_j = 0.470 N_{j-1} + 0.240 N_{j-7} + \varepsilon_j$$

$$\hat{\sigma}^2 = 228.89$$

auto correlations of the residuals are all negligible

141