STATS 206
Applied Multivariate Analysis
Lecture 5:
Multivariate Linear Regression Models

Prathapasinghe Dharmawansa

Department of Statistics
Stanford University
Autumn 2013

Agenda

- The classical linear regression model
- Least squares estimation
- Inferences about the regression model
- Inference from the estimated regression function
- Model checking and some other aspects of regression
- Multivariate multiple linear regression

The Classical Linear Regression Model (i)

- Let $z_1, \ldots, z_r$ be r predictor variables thought to be related to a response variable Y. The linear regression model with a single response is:

$$\underbrace{Y}_{\text{response}} = \underbrace{\beta_0 + \beta_1 z_1 + \ldots + \beta_r z_r}_{\text{mean (depending on } z_1,\ldots,z_r)} + \underbrace{\varepsilon}_{\text{error}}$$

  – "Linear": the mean part is linear in unknown parameters $\beta_0, \beta_1, \ldots, \beta_r$

The Classical Linear Regression Model (ii)

- With n independent observations on a single response, the complete multiple linear regression model is:

$$Y_i = \beta_0 + \beta_1 z_{i1} + \ldots + \beta_r z_{ir} + \varepsilon_i, \quad i = 1, \ldots, n$$

The error terms $\{\varepsilon_i\}_{i=1}^n$ satisfy:

1. $\mathrm{E}(\varepsilon_i) = 0$; 2. $\mathrm{Var}(\varepsilon_i) = \sigma^2$ (constant); 3. $\mathrm{Cov}(\varepsilon_j, \varepsilon_k) = 0, j \neq k$

- In matrix notation

$$\underbrace{\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}}_{\mathbf{Y}_{n\times 1}} = \underbrace{\begin{bmatrix} 1 & z_{11} & z_{12} & \ldots & z_{1r} \\ 1 & z_{21} & z_{22} & \ldots & z_{2r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & z_{n2} & \ldots & z_{nr} \end{bmatrix}}_{\mathbf{Z}_{n\times(r+1)}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_r \end{bmatrix}}_{\boldsymbol{\beta}_{(r+1)\times 1}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}}_{\boldsymbol{\varepsilon}_{n\times 1}}$$

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\mathbf{Z}: \text{ design matrix})$$

and 1. $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$; 2. $\mathrm{Cov}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}$

(In the above, $\boldsymbol{\beta}, \sigma^2$ are unknown parameters.)

Example 1: Fitting a Straight-line Regression Model

Determine the linear regression model for fitting a straight line: Mean response $= \mathrm{E}(Y) = \beta_0 + \beta_1 z_1$ to the data

z_1	0	1	2	3	4
y	1	4	3	8	9

- Random errors $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_5]' \Longrightarrow \mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

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

$$\Longrightarrow \mathbf{y} = \begin{bmatrix} 1 \\ 4 \\ 3 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

Least Squares Estimation (i)

- Problem: fitting the linear regression model to observed y_i based on known $1, z_{j1}, \ldots, z_{jr}$.
- Method of least squares: to select $\boldsymbol{\beta}$ which minimizes the sum of the squares of differences

$$
\begin{aligned}
S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 z_{i1} - \ldots - \beta_r z_{ir})^2 \\
&= (\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})
\end{aligned}
$$

  - The *minimizing* $\boldsymbol{\beta}$: the least squares estimate of $\boldsymbol{\beta}$, denoted as $\widehat{\boldsymbol{\beta}}$
  - Define the residuals:

$$
\widehat{\varepsilon}_i = y_i - \widehat{\beta}_0 - \widehat{\beta}_1 z_{i1} - \ldots - \widehat{\beta}_r z_{ir}, \quad i = 1, \ldots, n
$$

  The vector of residuals: $\widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}$

Least Squares Estimation (ii)

The Least Squares Estimate (Main Results)

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\mathbf{Z}\text{: full rank } (r+1) \le n)$$

- The least squares estimate of $\boldsymbol{\beta}$ in the above model is given by:
$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$$

- Orthogonality principle: Let $\widehat{\mathbf{y}} \triangleq \mathbf{Z}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$, $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and we have $\widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \widehat{\mathbf{y}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$. Then the following holds:
$$\mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}, \quad \widehat{\mathbf{y}}'\widehat{\boldsymbol{\varepsilon}} = 0$$

- Further, the residual sum of squares are given by the following
$$\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 z_{i1} - \ldots - \widehat{\beta}_r z_{ir})^2$$
$$= \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}\widehat{\boldsymbol{\beta}}$$

Least Squares Estimation (iii)

The Least Squares Estimate: Proof

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}, \quad \mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0},\ \widehat{\mathbf{y}}'\widehat{\boldsymbol{\varepsilon}} = 0$$

- Proof of orthogonality: let $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ as asserted; then

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

$\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$: "hat" matrix

(1) Both $(\mathbf{I} - \mathbf{H})$ and $\mathbf{H}$ are *symmetric*:

$$(\mathbf{I} - \mathbf{H})' = \mathbf{I} - \mathbf{H}; \qquad \mathbf{H}' = \mathbf{H}$$

(2) Both $(\mathbf{I} - \mathbf{H})$ and $\mathbf{H}$ are *idempotent*:

$$(\mathbf{I} - \mathbf{H})^2 = (\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) = \mathbf{I} - \mathbf{H}; \qquad \mathbf{H}^2 = \mathbf{H}$$

(3) $\mathbf{Z}'(\mathbf{I} - \mathbf{H}) = \mathbf{Z}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'] = \mathbf{Z}' - \mathbf{Z}' = \mathbf{0}$

Thus, $\mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{Z}'(\mathbf{y} - \widehat{\mathbf{y}}) = \mathbf{Z}'(\mathbf{I} - \mathbf{H})\mathbf{y} \overset{(3)}{=} \mathbf{0} \Longrightarrow \widehat{\mathbf{y}}'\widehat{\boldsymbol{\varepsilon}} = \widehat{\boldsymbol{\beta}}'\mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = 0$

- In addition: $\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{y}'(\mathbf{I} - \mathbf{H})'(\mathbf{I} - \mathbf{H})\mathbf{y} \overset{(1)(2)}{=} \mathbf{y}'(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}\widehat{\boldsymbol{\beta}}$

Least Squares Estimation (iv)

The Least Squares Estimate: Proof

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}, \quad \mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}, \ \widehat{\mathbf{y}}'\widehat{\boldsymbol{\varepsilon}} = 0$$

- Proof for $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$

  Let $\mathbf{y} - \mathbf{Z}\boldsymbol{\beta} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}} + \mathbf{Z}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\boldsymbol{\beta} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}} + \mathbf{Z}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$

$$\begin{aligned} \Longrightarrow \quad S(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{Z}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{Z}\boldsymbol{\beta}) \\ &= \underbrace{(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})}_{\text{not depending on } \boldsymbol{\beta}} + \underbrace{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{Z}'\mathbf{Z}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}_{\text{squared length of } \mathbf{Z}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})} \\ &\quad + 2\underbrace{(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'\mathbf{Z}}_{\widehat{\boldsymbol{\varepsilon}}'\mathbf{Z} = \mathbf{0}'}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{aligned}$$

  $\mathbf{Z}$ full-rank $\Longrightarrow$ the 2nd term $\mathbf{Z}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \neq \mathbf{0}$ if $\boldsymbol{\beta} \neq \widehat{\boldsymbol{\beta}}$
  (i.e., the squared length term is 0 (minimum) if and only if $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$)

  Summary: unique minimum $S(\boldsymbol{\beta})$ achieved at $\boldsymbol{\beta} = \underbrace{\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}}_{\text{Least Squares Estimate}}$ ■

Example 2: The LS estimates (a)

Calculate the LS estimates $\widehat{\boldsymbol{\beta}}$, the residuals $\widehat{\boldsymbol{\varepsilon}}$, and the residual sum of squares for a straight-line model $Y_j = \beta_0 + \beta_1 z_{j1} + \varepsilon_j$ to the data (Example 1)

z_1	0	1	2	3	4
y	1	4	3	8	9

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

- Here

$$\underbrace{\begin{bmatrix}1\\4\\3\\8\\9\end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix}1 & 0\\1 & 1\\1 & 2\\1 & 3\\1 & 4\end{bmatrix}}_{\mathbf{Z}} \underbrace{\begin{bmatrix}\beta_0\\\beta_1\end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\end{bmatrix}}_{\boldsymbol{\varepsilon}} \Longrightarrow \mathbf{Z}'\mathbf{y} = \begin{bmatrix}25\\70\end{bmatrix}$$

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix}5 & 10\\10 & 30\end{bmatrix}^{-1} = \begin{bmatrix}0.6 & -0.2\\-0.2 & 0.1\end{bmatrix}, \; (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \begin{bmatrix}1\\2\end{bmatrix}$$

Example 2: The LS estimates (b)

- (Cont'd)

$$\textbf{LS estimates}: \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\Longrightarrow \textbf{Fitted equation}: \widehat{y} = 1 + 2z$$

$$\text{Fitted value}: \widehat{\mathbf{y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \end{bmatrix}; \textbf{residuals}: \widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \widehat{\mathbf{y}} = \begin{bmatrix} 0 \\ 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}$$

$$\textbf{Residual sum of squares}: \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} = [0\ \ 1\ \ -2\ \ 1\ \ 0][0\ \ 1\ \ -2\ \ 1\ \ 0]' = 6$$

Least Squares Estimation (v)

Sum-of-squares Decomposition

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \qquad \text{Previously shown: } \mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0},\ \widehat{\mathbf{y}}'\widehat{\boldsymbol{\varepsilon}} = 0$$

- Since $\widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \widehat{\mathbf{y}}$ and $\widehat{\mathbf{y}}'\widehat{\boldsymbol{\varepsilon}} = 0$

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

- Since $\mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}$ and the first column of $\mathbf{Z}$ is $\mathbf{1}$

$$0 = \mathbf{1}'\widehat{\boldsymbol{\varepsilon}} = \sum_{i=1}^{n} \widehat{\varepsilon}_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \widehat{y}_i \Longrightarrow \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n} \widehat{y}_i \quad (\overline{y} = \overline{\widehat{y}})$$

From the above: $\mathbf{y}'\mathbf{y} - n\overline{y}^2 = \widehat{\mathbf{y}}'\widehat{\mathbf{y}} - n\overline{\widehat{y}}^2 + \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$, or

$$\underbrace{\sum_{i=1}^{n}(y_i - \overline{y})^2}_{\text{total sum of squares about mean}} = \underbrace{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}_{\text{regression sum of squares}} + \underbrace{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}_{\text{residual (error) sum of squares}}$$

Least Squares Estimation (vi)
Sum-of-squares Decomposition

(Cont'd)

$$\underbrace{\sum_{i=1}^{n}(y_i - \overline{y})^2}_{\text{total sum of squares about mean}} = \underbrace{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}_{\text{regression sum of squares}} + \underbrace{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}_{\text{residual (error) sum of squares}}$$

- Coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

  – R^2 is the proportion of the total variation in the y_i's "explained" by the predictors $z_1, z_2, \ldots, z_r$;
  – R^2: measuring the quality of fitting

Least Squares Estimation (vii)

Sampling Properties of the Least Squares Estimators

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathsf{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I})$$

(1) The least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ satisfies:

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

$$\text{Proof}: \ \mathsf{E}(\widehat{\boldsymbol{\beta}}) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\underline{\mathsf{E}(\mathbf{Y})} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\underline{\mathbf{Z}\boldsymbol{\beta}} = \boldsymbol{\beta}$$

$$\begin{aligned} \mathsf{Cov}(\widehat{\boldsymbol{\beta}}) &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\underline{\mathsf{Cov}(\mathbf{Y})}((\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')' \\ &= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\underline{\sigma^2\mathbf{I}})\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1} \end{aligned}$$

(2) The residuals $\widehat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}$ satisfies: [Note: $\widehat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$]

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

$$\text{Proof}: \ \mathsf{E}(\widehat{\boldsymbol{\varepsilon}}) = \mathsf{E}(\mathbf{Y}) - \mathbf{Z}\mathsf{E}(\widehat{\boldsymbol{\beta}}) = \mathbf{Z}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\beta} = \mathbf{0}$$

$$\Longrightarrow \mathsf{Cov}(\widehat{\boldsymbol{\varepsilon}}) = (\underline{\mathbf{I} - \mathbf{H}})\underline{\underline{\mathsf{Cov}(\mathbf{Y})}(\mathbf{I} - \mathbf{H})'} = \underline{\underline{\sigma^2}}(\underline{\mathbf{I} - \mathbf{H}})$$

Least Squares Estimation (viii)

Sampling Properties of the Least Squares Estimators

(3) $\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}) = (n-r-1)\sigma^2$; define $s^2 = \frac{\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}}{n-r-1} = \frac{\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}}{n-r-1} \Longrightarrow \mathsf{E}(s^2) = \sigma^2$

Proof: Fact: $\mathrm{trace}(\mathbf{AB}) = \mathrm{trace}(\mathbf{BA})$

$$\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}) = \mathsf{E}\{\mathrm{trace}(\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}})\} = \mathsf{E}\{\mathrm{trace}(\widehat{\boldsymbol{\varepsilon}}\widehat{\boldsymbol{\varepsilon}}')\} = \mathrm{trace}\{\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}})\}$$

$$\overset{\mathsf{E}(\widehat{\boldsymbol{\varepsilon}})=\mathbf{0}}{=} \mathrm{trace}\{\mathsf{Cov}(\widehat{\boldsymbol{\varepsilon}})\} = \mathrm{trace}\{\sigma^2(\mathbf{I}_n - \mathbf{H})\}$$

$$= \sigma^2[\mathrm{trace}(\mathbf{I}_n) - \mathrm{trace}(\mathbf{H})] = \sigma^2(n-(r+1)) \quad \text{where we used}$$

$$\mathrm{trace}(\mathbf{H}) \overset{\mathbf{H}=\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'}{=} \mathrm{trace}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Z}] = \mathrm{trace}(\mathbf{I}_{r+1}) = r+1$$

(4) $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\varepsilon}}$: <u>uncorrelated</u>

Proof: $\mathsf{Cov}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\varepsilon}}) = \mathsf{Cov}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}, (\mathbf{I}-\mathbf{H})\mathbf{Y})]$

$$= (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'[\mathsf{Cov}(\mathbf{Y})](\mathbf{I}-\mathbf{H})' = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I}-\mathbf{H}) \overset{\mathbf{Z}'(\mathbf{I}-\mathbf{H})=\mathbf{0}}{=} \mathbf{0}$$

Gauss' Least Squares Theorem

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad [\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0},\ \ \mathsf{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I};\ \ \mathbf{Z}: \text{full-rank } (r+1)]$$

For any $\mathbf{c}$, the estimator

$$\mathbf{c}'\widehat{\boldsymbol{\beta}} = c_0\widehat{\beta}_0 + c_1\widehat{\beta}_1 + \ldots + c_r\widehat{\beta}_r$$

of $\mathbf{c}'\boldsymbol{\beta}$ has the smallest possible variance among all linear estimators of the form $\mathbf{a}'\mathbf{Y} = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n$ that are unbiased for $\mathbf{c}'\boldsymbol{\beta}$.

Proof:

1) For any fixed $\mathbf{c}$, let $\mathbf{a}'\mathbf{Y}$ be any unbiased estimator of $\mathbf{c}'\boldsymbol{\beta}$. Then

$$\mathsf{E}(\mathbf{a}'\mathbf{Y}) = \mathbf{a}'\mathbf{Z}\boldsymbol{\beta} = \mathbf{c}'\boldsymbol{\beta} \text{ (regardless the value of } \boldsymbol{\beta})$$

$$\Longrightarrow (\mathbf{c} - \mathbf{Z}'\mathbf{a})'\boldsymbol{\beta} = \mathbf{0} \ \text{ for all } \boldsymbol{\beta}, \text{ including } \boldsymbol{\beta} = \mathbf{c} - \mathbf{Z}'\mathbf{a}$$

$$\Longrightarrow \mathbf{c} = \mathbf{Z}'\mathbf{a} \ \text{ for any unbiased estimator}$$

2) Furthermore, $\mathbf{c}'\widehat{\boldsymbol{\beta}} = \mathbf{c}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} = \mathbf{a}^{\star\prime}\mathbf{Y}$ with $\mathbf{a}^{\star} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{c}$.
From previous results, $\mathsf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$\Longrightarrow \mathbf{c}'\widehat{\boldsymbol{\beta}} = \mathbf{a}^{\star\prime}\mathbf{Y} \text{ is an unbiased estimator of } \mathbf{c}'\boldsymbol{\beta}$$

Gauss' Least Squares Theorem

For any $\mathbf{c}$, the estimator

$$\mathbf{c}'\widehat{\boldsymbol{\beta}} = c_0\widehat{\beta}_0 + c_1\widehat{\beta}_1 + \ldots + c_r\widehat{\beta}_r$$

of $\mathbf{c}'\boldsymbol{\beta}$ has the smallest possible variance among all linear estimators of the form $\mathbf{a}'\mathbf{Y} = a_1Y_1 + a_2Y_2 + \ldots + a_nY_n$ that are unbiased for $\mathbf{c}'\boldsymbol{\beta}$.

Proof: (Cont'd)

3) Due to unbiasedness condition [1) $\mathbf{c}' = \mathbf{a}'\mathbf{Z}$], for any $\mathbf{a}$, $(\mathbf{a}-\mathbf{a}^\star)'\mathbf{Z} = \mathbf{a}'\mathbf{Z} - \mathbf{a}^{\star'}\mathbf{Z} = \mathbf{c}' - \mathbf{c}' = \mathbf{0}' \Longrightarrow (\mathbf{a}-\mathbf{a}^\star)'\mathbf{a}^\star = (\mathbf{a}-\mathbf{a}^\star)'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{c} = 0$

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

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

Since $\mathbf{a}^\star$ is fixed and $(\mathbf{a}-\mathbf{a}^\star)'(\mathbf{a}-\mathbf{a}^\star) > 0$ unless $\mathbf{a} = \mathbf{a}^\star$, Var$(\mathbf{a}'\mathbf{Y})$ is minimized by choosing $\mathbf{a} = \mathbf{a}^\star$ and then $\mathbf{a}^{\star'}\mathbf{Y} = \mathbf{c}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\mathbf{Y} = \mathbf{c}'\widehat{\boldsymbol{\beta}}$.■

(The estimator $\mathbf{c}'\widehat{\boldsymbol{\beta}}$: referred to as the best (minimum-variance) linear unbiased estimator (BLUE) of $\mathbf{c}'\boldsymbol{\beta}$)

Inferences about the Regression Model (i)

Inferences Concerning the Regression Parameters

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})] \text{ New!}$$

- Previously assumed: $\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathsf{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}$; now: $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$
- Let $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ $[\mathbf{Z}$: full-rank $(r+1)]$ and $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$. Then

  i) Maximum likelihood estimator of $\boldsymbol{\beta}$ = Least squares estimator $\widehat{\boldsymbol{\beta}}$

  ii) $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \sim N_{r+1}(\boldsymbol{\beta}, \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1})$

  iii) $\widehat{\boldsymbol{\beta}}$ is independent of the residuals $\widehat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}}$

  iv) Let $\widehat{\sigma}^2$ be the maximum likelihood estimate of σ^2. Then

$$n\widehat{\sigma}^2 = \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} \sim \sigma^2\chi^2_{n-r-1}$$

  Proof: Both $\boldsymbol{\beta}$ and σ^2 are unknown parameters whose ML estimators are given by $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$ and $\widehat{\sigma}^2 = (\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})/n = \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}/n$, respectively. The rest is based on the analysis we have used so far.

Inferences about the Regression Model (ii)

Inferences Concerning the Regression Parameters

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [\mathbf{Z} : \text{full-rank } (r+1),\ \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})]$$

For the **Gaussian** linear regression model (above):

- A $100(1-\alpha)$% confidence region for $\boldsymbol{\beta}$ is given by

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

where $s^2 = \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}/(n-r-1)$ and $F_{r+1,n-r-1}(\alpha)$ is the upper (100α)-th percentile of an F dist. with $r+1$ and $n-r-1$ d.f.

- Simultaneous $100(1-\alpha)$% confidence intervals for the β_i are given by

$$\widehat{\beta}_i \pm \sqrt{\widehat{\text{Var}}(\widehat{\beta}_i)}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}, \quad i = 0, 1, \ldots, r$$

where $\widehat{\text{Var}}(\widehat{\beta}_i)$: the diagonal element of $s^2(\mathbf{Z}'\mathbf{Z})^{-1}$ corresponding to $\widehat{\beta}_i$.

(Proof: see the next two slides)

Inferences about the Regression Model (iii)

Inferences Concerning the Regression Parameters

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [\mathbf{Z}: \text{full-rank } (r+1),\ \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})]$$

A $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$:

$$\underbrace{(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})'(\mathbf{Z}'\mathbf{Z})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})}_{\mathbf{V}'\mathbf{V}} \le (r+1)s^2 F_{r+1,n-r-1}(\alpha)$$

- Outline of proof:

1) Let $\mathbf{V} = (\mathbf{Z}'\mathbf{Z})^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow \mathbf{V} \sim N_{r+1}(\mathbf{0}, \sigma^2\mathbf{I}), \quad \mathbf{V}'\mathbf{V} \sim \sigma^2\chi^2_{r+1}$
2) Previously: $(n-r-1)s^2 \triangleq \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} \sim \sigma^2\chi^2_{n-r-1}$
3) Previously: $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\varepsilon}}$ independent $\Longrightarrow s^2, \mathbf{V}'\mathbf{V}$ independent

$$\overset{1)\,2)\,3)}{\Longrightarrow} \frac{\mathbf{V}'\mathbf{V}/(r+1)}{s^2} = \frac{\chi^2_{r+1}/(r+1)}{\chi^2_{n-r-1}/(n-r-1)} \sim F_{r+1,n-r-1}. \quad \blacksquare$$

Inferences about the Regression Model (iv)

Inferences Concerning the Regression Parameters

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad [\mathbf{Z} : \text{full-rank } (r+1),\ \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})]$$

Simultaneous $100(1-\alpha)\%$ confidence intervals for the β_i:

$$\widehat{\beta}_i \pm \sqrt{\widehat{\text{Var}}(\widehat{\beta}_i)}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}, \quad i = 0, 1, \ldots, r$$

- Outline of proof:

  Previous page: A $100(1-\alpha)\%$ confidence region for $\boldsymbol{\beta}$:

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

$$\Longrightarrow (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})' \left[s^2(\mathbf{Z}'\mathbf{Z})^{-1}\right]^{-1} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \le \underbrace{(r+1)F_{r+1,n-r-1}(\alpha)}_{c^2}$$

  Take $\mathbf{a}_i = [0, \ldots, 0, 1, 0, \ldots, 0]'$ (1 at the i-th position). Then the simultaneous $100(1-\alpha)\%$confidence intervals are

$$|\beta_i - \widehat{\beta}_i|^2 \le c^2 \left[s^2(\mathbf{Z}'\mathbf{Z})^{-1}\right]_{ii} \quad \text{or} \quad |\beta_i - \widehat{\beta}_i|^2 \le c^2\, \widehat{\text{Var}}(\widehat{\beta}_i), \quad \forall i. \qquad \blacksquare$$

Inferences about the Regression Model (v)

Likelihood Ratio Tests for the Regression Parameters (1)

- Part of regression analysis: to assess the effects of particular predictor variables on the response variable.
- Here our null hypothesis is that $z_{q+1}, z_{q+2}, \ldots, z_r$ do not influence Y:

  $H_0 : \beta_{q+1} = \beta_{q+2} = \ldots = \beta_r = 0$ or $\boldsymbol{\beta}_{(2)} = [\beta_{q+1}, \beta_{q+2}, \ldots, \beta_r]' = \mathbf{0}$

  versus $H_1 : \beta_i \neq 0$ for some i, $q+1 \leq i \leq r$

  Under $H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0}$, the model is $\mathbf{Y} = \mathbf{Z}_1\boldsymbol{\beta}_{(1)} + \boldsymbol{\varepsilon}$ and under H_1, the model is $\mathbf{Y} = \mathbf{Z}_1\boldsymbol{\beta}_{(1)} + \mathbf{Z}_2\boldsymbol{\beta}_{(2)} + \boldsymbol{\varepsilon}$, where

$$\mathbf{Z} = \left[\begin{array}{c:c} \underbrace{\mathbf{Z}_1}_{n\times(q+1)} & \underbrace{\mathbf{Z}_2}_{n\times(r-q)} \end{array} \right], \quad \boldsymbol{\beta} = \left[\begin{array}{c} \underbrace{\boldsymbol{\beta}_{(1)}}_{(q+1)\times 1} \\ \hdashline \underbrace{\boldsymbol{\beta}_{(2)}}_{(r-q)\times 1} \end{array} \right]$$

Inferences about the Regression Model (vi)

Likelihood Ratio Tests for the Regression Parameters (2)

Define the following:

$$\mathrm{SS}_{\mathrm{res}}(\mathbf{Z}_1) \triangleq (\mathbf{y} - \mathbf{Z}_1\widehat{\boldsymbol{\beta}}_{(1)})'(\mathbf{y} - \mathbf{Z}_1\widehat{\boldsymbol{\beta}}_{(1)}), \;\; \widehat{\boldsymbol{\beta}}_{(1)} = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{y}$$

$$\mathrm{SS}_{\mathrm{res}}(\mathbf{Z}) \triangleq (\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})$$

- Main result here:

Consider: $\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ $[\mathbf{Z}$: full-rank $(r+1)$, $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})]$.

The likelihood ratio test of $H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0}$ rejects H_0 at level α if

$$\frac{[\mathrm{SS}_{\mathrm{res}}(\mathbf{Z}_1) - \mathrm{SS}_{\mathrm{res}}(\mathbf{Z})]/(r-q)}{s^2} > F_{r-q,n-r-1}(\alpha)$$

where as before we have $s^2 = (\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})/(n-r-1)$.

Inferences about the Regression Model (vii)

Likelihood Ratio Tests for the Regression Parameters (3)

Proof:

1) Given the data and the normality assumption, the likelihood function with unknown $\boldsymbol{\beta}, \sigma^2$ is

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

At maximum: $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$, $\widehat{\sigma}^2 = (\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})/n$

2) Under H_0, $\mathbf{Y} = \mathbf{Z}_1\boldsymbol{\beta}_{(1)} + \boldsymbol{\varepsilon}$.

$$L\left(\boldsymbol{\beta}_{(1)}, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{(\mathbf{y}-\mathbf{Z}_1\boldsymbol{\beta}_{(1)})'(\mathbf{y}-\mathbf{Z}_1\boldsymbol{\beta}_{(1)})}{2\sigma^2}}$$

$$\leq \max_{\boldsymbol{\beta}_{(1)},\sigma^2} L\left(\boldsymbol{\beta}_{(1)}, \sigma^2\right) = \frac{e^{-n/2}}{(2\pi\widehat{\sigma}_1^2)^{n/2}}$$

At max.: $\widehat{\boldsymbol{\beta}}_{(1)} = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{y}$, $\widehat{\sigma}_1^2 = (\mathbf{y} - \mathbf{Z}_1\widehat{\boldsymbol{\beta}}_{(1)})'(\mathbf{y} - \mathbf{Z}_1\widehat{\boldsymbol{\beta}}_{(1)})/n$

Inferences about the Regression Model (viii)

Likelihood Ratio Tests for the Regression Parameters (4)

Proof: (Cont'd)

3) The likelihood ratio is given by

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

which leads to the test statistic $(\widehat{\sigma}_1^2 - \widehat{\sigma}^2)/\widehat{\sigma}^2$ or its scaled version

$$\frac{n(\widehat{\sigma}_1^2 - \widehat{\sigma}^2)/(r-q)}{n\widehat{\sigma}^2/(n-r-1)} = \frac{\frac{\mathsf{SS}_{\mathsf{res}}(\mathbf{Z}_1) - \mathsf{SS}_{\mathsf{res}}(\mathbf{Z})}{(r-q)}}{s^2} \sim F_{r-q,n-r-1}. \quad \blacksquare$$

Inferences from the Estimated Regression Function (i)

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathsf{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I})$$

$\mathbf{z}_0 = [1, z_{01}, \ldots, z_{0r}]'$: selected values for predictor variables
(or a specific point in the design matrix space)

- Estimating the regression function at $\mathbf{z}_0$

Y_0: the response at $\mathbf{z}_0 \Longrightarrow \mathsf{E}(Y_0|\mathbf{z}_0) = \mathbf{z}_0'\boldsymbol{\beta}$ (LS estimate: $\mathbf{z}_0'\widehat{\boldsymbol{\beta}}$)

i) $\mathbf{z}_0'\widehat{\boldsymbol{\beta}}$: the unbiased estimator of $\mathbf{z}_0'\boldsymbol{\beta}$ with minimum variance
(due to Gauss' LS theorem)

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

iii) If ε is normal $\Rightarrow$ a $100(1-\alpha)\%$ confidence interval for $\mathsf{E}(Y_0|\mathbf{z}_0) = \mathbf{z}_0'\boldsymbol{\beta}$:

$$\mathbf{z}_0'\widehat{\boldsymbol{\beta}} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{[\mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0]s^2}$$

Inferences from the Estimated Regression Function (ii)

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (\mathsf{E}(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \mathsf{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I})$$

$\mathbf{z}_0 = [1, z_{01}, \ldots, z_{0r}]'$: selected values for predictor variables

(or a specific point in the design matrix space)

- Forecasting a new observation at $\mathbf{z}_0$ (The model: $Y_0 = \mathbf{z}_0'\boldsymbol{\beta} + \varepsilon_0$)

  Y_0, a new observation at $\mathbf{z}_0$, is predicted as

$$\mathbf{z}_0'\widehat{\boldsymbol{\beta}} = \widehat{\beta}_0 + \widehat{\beta}_1 z_{01} + \ldots + \widehat{\beta}_r z_{0r}$$

a) $\mathbf{z}_0'\widehat{\boldsymbol{\beta}}$: unbiased predictor

b) Forecast error: $(Y_0 - \mathbf{z}_0'\widehat{\boldsymbol{\beta}})$, $\quad \mathsf{Var}(Y_0 - \mathbf{z}_0'\widehat{\boldsymbol{\beta}}) = \sigma^2[1 + \mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0]$

c) If $\boldsymbol{\varepsilon}$ is normal $\Rightarrow$ a $100(1-\alpha)\%$ prediction interval for Y_0:

$$\mathbf{z}_0'\widehat{\boldsymbol{\beta}} \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{[1 + \mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0]s^2}$$

Example 3: Computer Data (a)

Interval Estimates for Mean and Future Responses

- Companies considering purchasing computers:
  - Assessing future needs to determine suitable equipment
- Data collected from $n = 7$ companies
  - To develop a forecast equation of hardware requirements for inventory management
- The data (see the table on next page)
  - z_1: customer orders (in thousands)
  - z_2: add-delete item count (in thousands)
  - Y: CPU (central processing unit) time (in hours)

Example 3: Computer Data (b)

Interval Estimates for Mean and Future Responses

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	0.815	146.5
151.5	1.061	172.8
136.2	8.603	160.1
92	1.125	108.5

- Construct a 95% confidence interval for the mean CPU time: $\mathsf{E}(Y_0|\mathbf{z}_0) = \beta_0 + \beta_1 z_{01} + \beta_2 z_{02}$ at $\mathbf{z}_0 = [1,\ 130,\ 7.5]'$
- Find a 95% prediction interval for a new facility's CPU requirement corresponding to the same $\mathbf{z}_0$

Table 7.3 in the textbook

Example 3: Computer Data (c)
Interval Estimates for Mean and Future Responses

Analysis

1. First construct the estimated regression function:

$$\underbrace{\begin{bmatrix}141.5\\168.9\\154.8\\146.5\\172.8\\160.1\\108.5\end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix}1 & 123.5 & 2.108\\1 & 146.1 & 9.213\\1 & 133.9 & 1.905\\1 & 128.5 & 0.815\\1 & 151.5 & 1.061\\1 & 136.2 & 8.603\\1 & 92 & 1.125\end{bmatrix}}_{\mathbf{Z}} \underbrace{\begin{bmatrix}\beta_0\\\beta_1\\\beta_2\end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix}\varepsilon_1\\\varepsilon_2\\\varepsilon_3\\\varepsilon_4\\\varepsilon_5\\\varepsilon_6\\\varepsilon_7\end{bmatrix}}_{\boldsymbol{\varepsilon}}$$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = \begin{bmatrix}8.42\\1.08\\0.42\end{bmatrix} \Longrightarrow \widehat{y} = 8.42 + 1.08z_1 + 0.42z_2$$

Example 3: Computer Data (d)
Interval Estimates for Mean and Future Responses

Analysis (Cont'd)

Estimated regression function: $\widehat{y} = 8.42 + 1.08z_1 + 0.42z_2$ and

$$(\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 8.1797 & -0.0641 & 0.0883 \\ -0.0641 & 0.0005 & -0.0011 \\ 0.0883 & -0.0011 & 0.0144 \end{bmatrix} \qquad (n = 7, r = 2)$$

$$s^2 = \frac{(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'}{7-2-1} \Longrightarrow s = 1.2039$$

2. Thus

$$\mathbf{z}_0'\widehat{\boldsymbol{\beta}} = 8.42 + 1.08(130) + 0.42(7.5) = 151.97, \;\; t_4(0.025) = 2.776$$

$$s\sqrt{\mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0} = 0.71, \;\; s\sqrt{1 + \mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0} = 1.40$$

Example 3: Computer Data (e)

Interval Estimates for Mean and Future Responses

Analysis (Cont'd)

3. The 95% confidence interval for the mean CPU time at $\mathbf{z}_0$ is

$$\mathbf{z}_0'\widehat{\boldsymbol{\beta}} \pm t_4(0.025)s\sqrt{\mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0} = 151.97 \pm 2.776(0.71)$$

or (150.00, 153.94)

4. A 95% prediction interval for the CPU time at a new facility with condition $\mathbf{z}_0$ is

$$\mathbf{z}_0'\widehat{\boldsymbol{\beta}} \pm t_4(0.025)s\sqrt{1 + \mathbf{z}_0'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z}_0} = 151.97 \pm 2.776(1.40)$$

or (148.08, 155.86)

Model Checking: A Residual Analysis

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

- Need to examine the adequacy of the model before using it
- Check the model by checking the residuals:

  Recall

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

  $$\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}) = \mathbf{0}, \ \mathsf{Cov}(\widehat{\boldsymbol{\varepsilon}}) = \sigma^2(\mathbf{I} - \mathbf{H}), \ \mathsf{Var}(\widehat{\varepsilon}_j) = \sigma^2(1 - h_{jj}), j = 1, \ldots, n$$

  where h_{jj}: the j-th diagonal element of $\mathbf{H}$ (known as leverage)

  Use $s^2 = \frac{\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}}{n-r-1}$ as an estimate of σ^2 (recall $\mathsf{E}(s^2) = \sigma^2$):

  $$\widehat{\mathsf{Var}}(\widehat{\varepsilon}_j) = s^2(1 - h_{jj}), \ j = 1, 2, \ldots, n$$

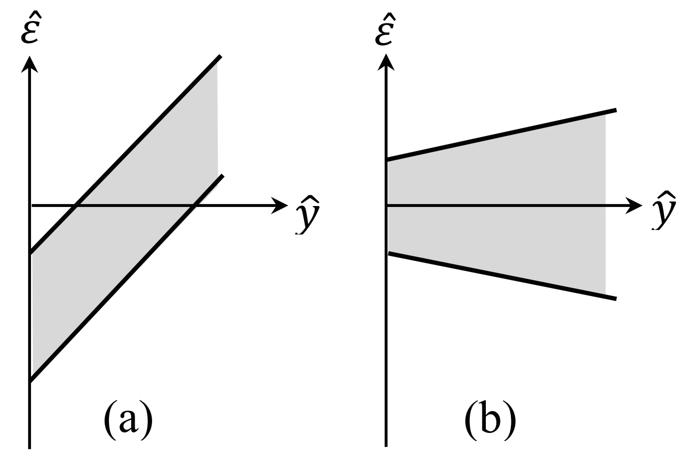
  $$\text{Studentized residuals}: \ \widehat{\varepsilon}_j^* = \frac{\widehat{\varepsilon}_j}{\sqrt{s^2(1 - h_{jj})}}, \ j = 1, 2, \ldots, n$$

- If the model fits, we expect $\widehat{\varepsilon}_j^*$'s to look like i.i.d $N(0, 1)$ random variables

Model Checking: A Residual Analysis

Plotting the Residuals ($\widehat{\varepsilon}_j$ or $\widehat{\varepsilon}_j^*$)

1. Plot residual $\widehat{\varepsilon}_j$ vs. $\widehat{y}_j$ ($\widehat{y}_j = \widehat{\beta}_0 + \widehat{\beta}_1 z_{j1} + \ldots + \widehat{\beta}_r z_{jr}$); be aware of:
   (a) Dependence of $\widehat{\varepsilon}_j$ on $\widehat{y}_j$
   (b) Non-constant variance



2. Plot $\widehat{\varepsilon}_j$ vs. a predictor variable, such as z_1
3. Q-Q plots of $\widehat{\varepsilon}_j^*$ or $\widehat{\varepsilon}_j$; check normality; detect unusual observations

Model Checking: A Residual Analysis

Plotting the Residuals ($\widehat{\varepsilon}_j$ or $\widehat{\varepsilon}_j^*$)

4. Plot residuals vs. time (i.e., check the assumption of independence)

   Assuming chronological data, construct a test of independence from the first-order auto-correlation

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

   Then use the Durbin-Watson test based on the following statistic:

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

   Compare the obtained result with a table of critical values (details omitted here)

Model Checking: Leverage and Influence

- Residual analysis: useful but may not be enough
  $\Longrightarrow$ Further check leverage and influential observations
- Leverage: h_{jj} (the j-th diagonal element of $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$)
  - Measuring the distance of the j-th observ. to the rest $(n-1)$ ones
    Example: consider the simple model with one variable z $(\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon})$

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, h_{jj} = \frac{1}{n} + \frac{(z_j - \overline{z})^2}{\sum_{k=1}^{n}(z_k - \overline{z})^2}$$

  - Measuring the contribution of y_j to $\widehat{y}_j$ (hence the name "leverage")
    Recall: $\widehat{\mathbf{y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y} \Longrightarrow \widehat{y}_j = h_{jj}y_j + \sum_{k \neq j} h_{jk}y_k$; If h_{jj} is large relative to other h_{jk}, then y_j will a major contributor to $\widehat{y}_j$.
- Influential observations: Methods for assessing influence are typically based on the change in the least squares estimate $\widehat{\boldsymbol{\beta}}$ when observations are deleted from the data. (Details omitted here)

Other Aspects of Linear Regression

Predictor Variables Selection

- Methods for selecting predictor variables
  1. Mallow's C_p statistic   (p: the number of variables)

     $$C_p = \left(\frac{\text{residual sum of squares for subset model with } p \text{ parameters}}{\text{residual variance for full model}} \right) - (n - 2p)$$

     Plot (p, C_p) for each subset of predictors; choose the one with (p, C_p) coordinates near the 45° line
  2. If the list of predictors is long, use stepwise regression to select important ones without considering all possibilities.
  3. Information-criterion based approaches, e.g.:
     Akaike's information criterion (AIC) for selecting p

     $$\text{AIC}(p) = n \ln \left(\frac{\text{residual sum of squares for subset model with } p \text{ parameters}}{n} \right) + 2p$$

     Select models with smaller AIC values

Multivariate Multiple Linear Regression (i)

- Modeling m multiple linear regressions using the same design matrix $\mathbf{Z}$

  The multivariate multiple linear regression model:

$$\begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1m} \\ Y_{21} & Y_{22} & \dots & Y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{nm} \end{bmatrix} = \begin{bmatrix} z_{10} & z_{11} & \dots & z_{1r} \\ z_{20} & z_{21} & \dots & z_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n0} & z_{n1} & \dots & z_{nr} \end{bmatrix} \begin{bmatrix} \beta_{01} & \beta_{02} & \dots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{r1} & \beta_{r2} & \dots & \beta_{rm} \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \dots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \dots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \dots & \varepsilon_{nm} \end{bmatrix}$$

$(\text{Note}: \ z_{10} = z_{20} = \ldots = z_{n0} = 1)$

Multivariate Multiple Linear Regression (ii)

(Cont'd) In a more compact form:

$$\mathbf{Y}_{n\times m} = \mathbf{Z}_{n\times(r+1)}\boldsymbol{\beta}_{(r+1)\times m} + \boldsymbol{\varepsilon}_{n\times m}$$

$$\mathbf{Y} = \left[\mathbf{Y}_{(1)}|\mathbf{Y}_{(2)}|\dots|\mathbf{Y}_{(m)}\right]$$

$$\boldsymbol{\beta} = \left[\boldsymbol{\beta}_{(1)}|\boldsymbol{\beta}_{(2)}|\dots|\boldsymbol{\beta}_{(m)}\right], \qquad \boldsymbol{\varepsilon} = \left[\boldsymbol{\varepsilon}_{(1)}|\boldsymbol{\varepsilon}_{(2)}|\dots|\boldsymbol{\varepsilon}_{(m)}\right]$$

$$\left[\mathbf{Y}_{(1)}|\mathbf{Y}_{(2)}|\dots|\mathbf{Y}_{(m)}\right] = \mathbf{Z}\left[\boldsymbol{\beta}_{(1)}|\boldsymbol{\beta}_{(2)}|\dots|\boldsymbol{\beta}_{(m)}\right] + \left[\boldsymbol{\varepsilon}_{(1)}|\boldsymbol{\varepsilon}_{(2)}|\dots|\boldsymbol{\varepsilon}_{(m)}\right]$$

In the above,

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

From the model: $\mathbf{Y}_{(i)} = \mathbf{Z}\boldsymbol{\beta}_{(i)} + \boldsymbol{\varepsilon}_{(i)}, i = 1, \dots, m$, with $\mathsf{Cov}(\boldsymbol{\varepsilon}_{(i)}) = \sigma_{ii}\mathbf{I}$.
But the errors for different responses can be **correlated**.

Multivariate Multiple Linear Regression (iii)

- Thus, assuming $\mathbf{Z}$: full-rank (rank $(r+1) < n$)

  LS estimate of $\boldsymbol{\beta}_{(i)}$: $\widehat{\boldsymbol{\beta}}_{(i)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}_{(i)}, \quad i = 1, 2, \ldots, m$

The least squares estimate of matrix $\boldsymbol{\beta}$

$$\boxed{\widehat{\boldsymbol{\beta}} = \left[\hat{\boldsymbol{\beta}}_{(1)}|\ldots|\hat{\boldsymbol{\beta}}_{(m)}\right] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\left[\mathbf{Y}_{(1)}|\ldots|\mathbf{Y}_{(m)}\right] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}}$$

Remark:

Choose parameters $\mathbf{B} = \left[\mathbf{b}_{(1)} \mid \ldots \mid \mathbf{b}_{(m)}\right] \Longrightarrow$ error matrix: $\mathbf{Y} - \mathbf{Z}\mathbf{B}$

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

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

$\widehat{\boldsymbol{\beta}}$ <u>minimizes</u> trace$[(\mathbf{Y} - \mathbf{Z}\mathbf{B})'(\mathbf{Y} - \mathbf{Z}\mathbf{B})]$ and $|(\mathbf{Y} - \mathbf{Z}\mathbf{B})'(\mathbf{Y} - \mathbf{Z}\mathbf{B})|$.

Multivariate Multiple Linear Regression (iv)

- Using the LS estimate $\widehat{\beta}$, we have
  - Fitted values: $\widehat{\mathbf{Y}} = \mathbf{Z}\widehat{\boldsymbol{\beta}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$
  - Residuals: $\widehat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y}$
- Important results:

$$\boxed{\text{Orthogonality principle:} \quad \mathbf{Z}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}, \quad \widehat{\mathbf{Y}}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{0}}$$

  - Consequently: $\mathbf{Y}'\mathbf{Y} = \widehat{\mathbf{Y}}'\widehat{\mathbf{Y}} + \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$ or $\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{Z}'\mathbf{Z}\widehat{\boldsymbol{\beta}}$

Multivariate Multiple Linear Regression (v)

- Sampling properties of the LS estimate $\widehat{\boldsymbol{\beta}}$:
  Assuming $\mathbf{Z}$: full-rank (rank $(r+1) < n$)

> - $\mathsf{E}(\widehat{\boldsymbol{\beta}}_{(i)}) = \boldsymbol{\beta}_{(i)}$, i.e., $\mathsf{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
> - $\mathsf{Cov}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(\mathbf{Z}'\mathbf{Z})^{-1}, \quad i, k = 1, \ldots, m$
> - $\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}_{(i)}) = \mathbf{0}$ and $\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}) = \mathbf{0}$
> - $\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}'_{(i)}\widehat{\boldsymbol{\varepsilon}}_{(k)}) = (n-r-1)\sigma_{ik}$ and $\mathsf{E}(\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}) = (n-r-1)\boldsymbol{\Sigma}$
> - $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\varepsilon}}$: uncorrelated

Note: In the above: $\boldsymbol{\Sigma}_{m\times m} = \{\sigma_{ik}\}$

Multivariate Multiple Linear Regression (vi)

- **Gaussian/Normal** multivariate multiple linear regression

  Assuming (1) $\mathbf{Z}$: full-rank ($\text{rank}(\mathbf{Z}) = (r+1), n \geq (r+1)+m$), (2) $\boldsymbol{\varepsilon}$: multivariate normal and (3) $\boldsymbol{\Sigma}$: positive definite

- The LS estimate $\widehat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\mathbf{Y}$ is the ML estimator of $\boldsymbol{\beta}$
- $\widehat{\boldsymbol{\beta}}$: normal distribution, $\text{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, $\text{Cov}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\boldsymbol{\beta}}_{(k)}) = \sigma_{ik}(\mathbf{Z}'\mathbf{Z})^{-1}$
- $\widehat{\boldsymbol{\beta}}$: independent of $\widehat{\boldsymbol{\Sigma}}$, where $\widehat{\boldsymbol{\Sigma}}$ is the ML estimator of $\boldsymbol{\Sigma}$:
  $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}} = \frac{1}{n}(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{Z}\widehat{\boldsymbol{\beta}})$ and $n\widehat{\boldsymbol{\Sigma}} \sim \mathbf{W}_{m,n-r-1}(\boldsymbol{\Sigma})$
- The maximum likelihood $L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}) = (2\pi)^{-mn/2} \left|\widehat{\boldsymbol{\Sigma}}\right|^{-n/2} e^{-mn/2}$

Multivariate Multiple Linear Regression (vii)

Likelihood Ratio Test for Regression Parameters (1)

- Testing $H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0}$, where $\boldsymbol{\beta} = \begin{bmatrix} \underbrace{\boldsymbol{\beta}_{(1)}}_{(q+1)\times m} \\ \underbrace{\boldsymbol{\beta}_{(2)}}_{(r-q)\times m} \end{bmatrix}$, vs. $H_1 : \boldsymbol{\beta}_{(2)} \neq \mathbf{0}$

  Partitioning $\mathbf{Z} = [\ \underbrace{\mathbf{Z}_1}_{n\times(q+1)} \ |\ \underbrace{\mathbf{Z}_2}_{n\times(r-q)}\ ] \Longrightarrow \mathbf{Y} = \mathbf{Z}_1\boldsymbol{\beta}_{(1)} + \mathbf{Z}_2\boldsymbol{\beta}_{(2)} + \boldsymbol{\varepsilon}$

  Under H_0: $\mathbf{Y} = \mathbf{Z}_1\boldsymbol{\beta}_{(1)} + \boldsymbol{\varepsilon}$; assuming normality,

$$\widehat{\boldsymbol{\beta}}_{(1)} = (\mathbf{Z}_1'\mathbf{Z}_1)^{-1}\mathbf{Z}_1'\mathbf{Y}, \qquad \widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{n}(\mathbf{Y} - \mathbf{Z}_1\widehat{\boldsymbol{\beta}}_{(1)})'(\mathbf{Y} - \mathbf{Z}_1\widehat{\boldsymbol{\beta}}_{(1)})$$

  The likelihood ratio:

$$\Lambda = \frac{\max_{\boldsymbol{\beta}_{(1)},\boldsymbol{\Sigma}} L\left(\boldsymbol{\beta}_{(1)}, \boldsymbol{\Sigma}\right)}{\max_{\boldsymbol{\beta},\boldsymbol{\Sigma}} L(\boldsymbol{\beta}, \boldsymbol{\Sigma})} = \frac{L(\widehat{\boldsymbol{\beta}}_{(1)}, \widehat{\boldsymbol{\Sigma}}_1)}{L(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}})} = \left(\frac{|\widehat{\boldsymbol{\Sigma}}|}{|\widehat{\boldsymbol{\Sigma}}_1|}\right)^{n/2}$$

Multivariate Multiple Linear Regression (viii)

Likelihood Ratio Test for Regression Parameters (2)

Main result: Assuming $\mathbf{Z}$ having full rank $(r+1), (r+1)+m \leq n$, and $\boldsymbol{\varepsilon}$: normally distributed. Under $H_0 : \boldsymbol{\beta}_{(2)} = \mathbf{0}$, $n\widehat{\mathbf{\Sigma}} \sim \mathbf{W}_{m,n-r-1}(\mathbf{\Sigma})$ and $n\widehat{\mathbf{\Sigma}}$ is independent of $n(\widehat{\mathbf{\Sigma}}_1 - \widehat{\mathbf{\Sigma}})$, where $n(\widehat{\mathbf{\Sigma}}_1 - \widehat{\mathbf{\Sigma}}) \sim \mathbf{W}_{m,r-q}(\mathbf{\Sigma})$. The likelihood ratio test is equivalent to rejecting H_0 for large values of

$$-2\ln\Lambda = -n\ln\frac{|\widehat{\mathbf{\Sigma}}|}{|\widehat{\mathbf{\Sigma}}_1|} = -n\ln\frac{|n\widehat{\mathbf{\Sigma}}|}{|n\widehat{\mathbf{\Sigma}} + n(\widehat{\mathbf{\Sigma}}_1 - \widehat{\mathbf{\Sigma}})|}$$

For large n, the modified statistic

$$-\left[n - r - 1 - \frac{1}{2}(m - r + q + 1)\right]\ln\left(\frac{|\widehat{\mathbf{\Sigma}}|}{|\widehat{\mathbf{\Sigma}}_1|}\right)$$

can be shown to be approximately $\chi^2_{m(r-q)}$ distributed.

(Remark: Other test statistics used for testing H_0: e.g., Wilks' lambda, Pillai's trace, Hotelling-Lawley trace, and Roy's greatest root)