8.1

$$Y = XB + U, \quad X \; n\times p, \quad \text{rank } X = p, \quad B \; p\times d$$

$$EU = 0$$

$\hat{B} = (X^TX)^{-1}X^TY$ cf. page 400 line 1

$(Y - X\hat{B})^T(Y - X\hat{B}) \leq (Y - XB)^T(Y - XB)$ cf. formula (8.10)

$\Rightarrow \operatorname{tr}((Y - X\hat{B})^T(Y - X\hat{B})) \leq \operatorname{tr}((Y - XB)^T(Y - XB))$ cf. A 7.9 (6)

$\operatorname{tr}((Y - XB)^T(Y - XB)) = \operatorname{tr}((Y - X\hat{B})^T(Y - X\hat{B}))$

$\Leftrightarrow \operatorname{tr}((Y - XB)^T(Y - XB) - (Y - X\hat{B})^T(Y - X\hat{B})) = 0$

$\Leftrightarrow \operatorname{tr}((XB - X\hat{B})^T(XB - X\hat{B})) = 0$ cf. formula (8.10) rearranged*

$\Leftrightarrow \operatorname{tr}((B - \hat{B})^TX^TX(B - \hat{B})) = 0, \quad X^TX > 0$ cf. A 5.7

$\Leftrightarrow B - \hat{B} = 0$

$\Leftrightarrow B = \hat{B}$

i.e. equality only for $B = \hat{B}$

* $(Y - \Theta)^T(Y - \Theta) = (Y - \hat{\Theta})^T(Y - \hat{\Theta}) + (\hat{\Theta} - \Theta)^T(\hat{\Theta} - \Theta)$

$\Leftrightarrow (Y - \Theta)^T(Y - \Theta) - (Y - \hat{\Theta})^T(Y - \hat{\Theta}) = (\Theta - \hat{\Theta})^T(\Theta - \hat{\Theta})$

8.2

$$\underset{\sim}{y}_i \sim N_d(\underset{\sim}{\theta}_i, \Sigma) \quad , \quad i = 1,2,3 \quad , \quad \text{independent}$$

$$\underset{\sim}{\theta}_1 = \underset{\sim}{\beta}_1 \qquad \underset{\sim}{\theta}_2 = \underset{\sim}{\beta}_1 - \underset{\sim}{\beta}_2 \qquad \underset{\sim}{\theta}_3 = \underset{\sim}{\beta}_1 + \underset{\sim}{\beta}_2$$

$$Y = XB + U = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta}_1^T \\ \underset{\sim}{\beta}_2^T \end{bmatrix} + U$$

$$\hat{B} = (X^T X)^{-1} X^T Y$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \quad , \quad (X^T X)^{-1} = \frac{1}{6} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \frac{1}{6} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} Y \quad , \qquad \hat{B}^T = Y^T \begin{bmatrix} \frac{1}{3} & 0 \\ \frac{1}{3} & -\frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} \end{bmatrix}$$

$$\hat{\underset{\sim}{\beta}}_1 = \frac{1}{3}(\underset{\sim}{y}_1 + \underset{\sim}{y}_2 + \underset{\sim}{y}_3)$$

$$\hat{\underset{\sim}{\beta}}_2 = \frac{1}{2}(\underset{\sim}{y}_3 - \underset{\sim}{y}_2)$$

$$\hat{\Theta} = X\hat{B} = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} Y = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{5}{6} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{1}{6} & \frac{5}{6} \end{bmatrix} Y$$

$$\hat{\Theta}^T = Y^T \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{5}{6} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{1}{6} & \frac{5}{6} \end{bmatrix}$$

$$\hat{\underset{\sim}{\theta}}_1 = \frac{1}{3}(\underset{\sim}{y}_1 + \underset{\sim}{y}_2 + \underset{\sim}{y}_3) \qquad (= \hat{\underset{\sim}{\beta}}_1)$$

$$\hat{\underset{\sim}{\theta}}_2 = \frac{1}{3}\underset{\sim}{y}_1 + \frac{5}{6}\underset{\sim}{y}_2 - \frac{1}{6}\underset{\sim}{y}_3 \qquad (= \hat{\underset{\sim}{\beta}}_1 - \hat{\underset{\sim}{\beta}}_2)$$

$$\hat{\underset{\sim}{\theta}}_3 = \frac{1}{3}\underset{\sim}{y}_1 - \frac{1}{6}\underset{\sim}{y}_2 + \frac{5}{6}\underset{\sim}{y}_3 \qquad (= \hat{\underset{\sim}{\beta}}_1 + \hat{\underset{\sim}{\beta}}_2)$$

8.3

$\underset{\sim}{y}_i \sim N_d(\underset{\sim}{\mu}_i, \Sigma)$, $i = 1, \dots, n$, independent

$\underset{\sim}{\mu}_i = B^T \underset{\sim}{x}_i$, B $p \times d$

$X = [\underset{\sim}{x}_1 \ \underset{\sim}{x}_2 \ \dots \ \underset{\sim}{x}_n]^T$ $n \times p$, rank $X = p$

$\underset{\sim}{y}_i = B^T \underset{\sim}{x}_i + \underset{\sim}{u}_i$, $\underset{\sim}{u}_i \sim N_d(\underset{\sim}{0}, \Sigma)$, $i = 1, \dots, n$, independent

$Y = XB + U$, i.e. the usual model

Consider $\hat{\underset{\sim}{y}}_0 = \hat{B}^T \underset{\sim}{x}_0$, $\hat{B} = (X^TX)^{-1}X^TY$

$\hat{y}_{0j} = \hat{\underset{\sim}{\beta}}^{(j)T} \underset{\sim}{x}_0 = \underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}}^{(j)}$

$$\mathrm{Cov}(\hat{y}_{0j}, \hat{y}_{0k}) = \mathrm{Cov}(\underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}}^{(j)}, \underset{\sim}{x}_0^T \hat{\underset{\sim}{\beta}}^{(k)}) = \underset{\sim}{x}_0^T \mathrm{Cov}(\hat{\underset{\sim}{\beta}}^{(j)}, \hat{\underset{\sim}{\beta}}^{(k)}) \underset{\sim}{x}_0$$

$$= \underset{\sim}{x}_0^T \sigma_{jk} (X^TX)^{-1} \underset{\sim}{x}_0 = \underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \, \sigma_{jk}$$

cf. formula (8.22)

Hence $\mathrm{Var}\, \hat{\underset{\sim}{y}}_0 = \underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \Sigma$

As Σ is estimated by S $\left(= \frac{E}{n-p}\right)$ we see that $\mathrm{Var}\, \hat{\underset{\sim}{y}}_0$ is estimated by $\underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 S$

Consider $\underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 = \underset{\sim}{a}^T \hat{B}^T \underset{\sim}{x}_0$, lin. comb. of normal dist. var.

$E[\underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0] = \underset{\sim}{a}^T E \hat{B}^T \underset{\sim}{x}_0 = \underset{\sim}{a}^T B^T \underset{\sim}{x}_0 = \underset{\sim}{a}^T \underset{\sim}{\mu}_0$

$\mathrm{Var}(\underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0) = \underset{\sim}{a}^T \underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \Sigma \underset{\sim}{a} = \underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \underset{\sim}{a}^T \Sigma \underset{\sim}{a}$

$\underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 \sim N(\underset{\sim}{a}^T \underset{\sim}{\mu}_0, \underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \underset{\sim}{a}^T \Sigma \underset{\sim}{a})$

$$\frac{\underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 - \underset{\sim}{a}^T \underset{\sim}{\mu}_0}{\sqrt{\underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \underset{\sim}{a}^T S \underset{\sim}{a}}} \sim t(n-p)$$

Hence the confidence interval for $\underset{\sim}{a}^T \underset{\sim}{\mu}_0$ with confidence level $1-\alpha$

$$\underset{\sim}{a}^T \underset{\sim}{\mu}_0 = \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 \pm t_{1-\frac{\alpha}{2}}(n-p) \sqrt{\underset{\sim}{x}_0^T (X^TX)^{-1} \underset{\sim}{x}_0 \underset{\sim}{a}^T S \underset{\sim}{a}}$$

cont.

8.3 continued

Let $\underset{\sim}{y}_0$ now indicate a new observation at $\underset{\sim}{x}_0$ independent of the ealier observations $\underset{\sim}{y}_1, \dots, \underset{\sim}{y}_n$

$\underset{\sim}{y}_0 = B^T \underset{\sim}{x}_0 + \underset{\sim}{u}_0$, $\underset{\sim}{u}_0 \sim N_d(\underset{\sim}{0}, \Sigma)$

Consider $\underset{\sim}{a}^T \underset{\sim}{y}_0 - \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 = \underset{\sim}{a}^T \underset{\sim}{y}_0 - \underset{\sim}{a}^T \hat{B}^T \underset{\sim}{x}_0$, lin. comb. of normal dist. var.

$$E[\underset{\sim}{a}^T \underset{\sim}{y}_0 - \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0] = \underset{\sim}{a}^T B^T \underset{\sim}{x}_0 - \underset{\sim}{a}^T B^T \underset{\sim}{x}_0 = 0$$

$$\begin{aligned} Var(\underset{\sim}{a}^T \underset{\sim}{y}_0 - \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0) &= \underset{\sim}{a}^T \Sigma \underset{\sim}{a} + \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0 \underset{\sim}{a}^T \Sigma \underset{\sim}{a} \\ &= (1 + \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0) \underset{\sim}{a}^T \Sigma \underset{\sim}{a} \end{aligned}$$

$$\underset{\sim}{a}^T \underset{\sim}{y}_0 - \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 \sim N(0, (1 + \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0) \underset{\sim}{a}^T \Sigma \underset{\sim}{a})$$

$$\frac{\underset{\sim}{a}^T \underset{\sim}{y}_0 - \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 - 0}{\sqrt{(1 + \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0) \underset{\sim}{a}^T S \underset{\sim}{a}}} \sim t(n-p)$$

Hence the prediction interval for $\underset{\sim}{a}^T \underset{\sim}{y}_0$ with confidence level $1-\alpha$

$$\underset{\sim}{a}^T \underset{\sim}{y}_0 = \underset{\sim}{a}^T \hat{\underset{\sim}{y}}_0 \pm t_{1-\frac{\alpha}{2}}(n-p) \sqrt{(1 + \underset{\sim}{x}_0^T (X^T X)^{-1} \underset{\sim}{x}_0) \underset{\sim}{a}^T S \underset{\sim}{a}}$$

8.4

$$\underset{\sim}{y}_i \sim N_d(\underset{\sim}{\mu}, \Sigma) \ , \quad i = 1, \dots, n \ , \quad \text{independent}$$

$Y = \underset{\sim}{1}_d \underset{\sim}{\mu}^T + U$, $\underset{\sim}{u}_i \sim N_d(\underset{\sim}{0}, \Sigma)$, $i = 1, \dots, n$, indep.

$$\hat{\underset{\sim}{\mu}}^T = (\underset{\sim}{1}_d^T \underset{\sim}{1}_d)^{-1} \underset{\sim}{1}_d^T Y = \frac{1}{d} \sum_{i=1}^{d} \underset{\sim}{y}_i^T = \bar{\underset{\sim}{y}}^T$$

$$\hat{\underset{\sim}{\mu}} = \bar{\underset{\sim}{y}}$$