STAT 8260
Spring 2008

STAT 8260 Homework #4 – Solutions

6.4 Since $var(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

choosing the x_i values to minimize $var(\hat{\beta}_1)$ is equivalent to maximizing $\sum_{i=1}^{n}(x_i-\bar{x})^2$.

This can be done by placing all of the x_i's as far away from their mean as possible. So, all of the x_i's must be equal to a or b, the only question remaining being what proportion of the x_i's should be placed at a & what proportion of the x_i's should be placed at b?

Suppose k of the x_i's are equal to a and the remaining $n-k$ x_i's are equal to b. Then

$\bar{x} = \dfrac{ka+(n-k)b}{n}$ and

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 = k\left[a - \frac{ka+(n-k)b}{n}\right]^2 + (n-k)\left[b - \frac{ka+(n-k)b}{n}\right]^2$$

$$= k\left[\frac{n(a-b)-k(a-b)}{n}\right]^2 + (n-k)\left[\frac{k(b-a)}{n}\right]^2$$

$$= \frac{k}{n^2}\left[(n-k)(a-b)\right]^2 + \frac{n-k}{n^2}\left[-k(a-b)\right]^2 = \frac{(a-b)^2}{n^2}\left[k(n-k)^2+k^2(n-k)\right]$$

$$= \frac{(a-b)^2}{n^2}k(n-k)(n-k+k) = \frac{(a-b)^2 k(n-k)}{n}$$

Differentiating w/ respect to k and setting $=0$ yields

$$\frac{(a-b)^2}{n}\frac{\partial}{\partial k}k(n-k) = \frac{(a-b)^2}{n}\left[k(-1)+(n-k)\right] = 0$$

$$\Rightarrow \frac{(a-b)^2}{n}(n-2k) = 0 \Rightarrow (a-b)^2 = \frac{2(a-b)^2}{n}k \Rightarrow k = \frac{(a-b)^2 n}{2(a-b)^2} = \frac{n}{2}$$

So $n/2$ of the x_i's $= a$, $n/2$ of them at b minimizes the variance of $\hat{\beta}_1$.

#2 (6.14, worked using birthweight data) Model: $y_i = \beta_0 + \beta_1 x_i + e_i$
$i = 1, \dots, 12$, $e_1, \dots, e_n \overset{iid}{\sim} N(0, \sigma^2)$

a) $$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$= \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} = 130.4$$

(see the Matlab script hwk4-2.m or the SAS program hwk4-2.sas for details of the calculations)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = -2141.7$$

b.) $$t = \frac{\hat{\beta}_1}{s \sqrt{\sum_i (x_i - \bar{x})^2}}$$

$$SSE = \underbrace{\sum (y_i - \bar{y})^2}_{SST} - \underbrace{\frac{\left[\sum (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\sum (x_i - \bar{x})^2}}_{SSR}$$

$$SST = 865127$$
$$SSR = S_{xy} \hat{\beta}_1 = 616401$$
$$SSE = SST - SSR = 248726$$
$$\Rightarrow s^2 = MSE = SSE/(n-2) = \frac{248726}{10} = 24872.6$$

$$t = \frac{130.4}{\sqrt{24872.6 / S_{xx}}} = 4.98$$

$$\hat{\imath} = 3.2955$$

Using a 2-sided alternative $H_1: \beta_1 \neq 0$, we compare this test statistic vs $t_{1-\alpha/2}(n-2) = t_{.975}(10) = 2.228$

Since $t = 4.89 > t_{.975}(10) = 2.228$, we reject $H_0: \beta_1 = 0$ at level $\alpha = .05$

Equivalently, we can compute the p-value,

$Pr\left(|t(10)| > 4.98 \mid H_0 \text{ is true}\right) = .00055 < .05$

c.) The standard error of $\hat{\beta}_1$ was the denominator of our t-test in part (b.):

$$s.e.(\hat{\beta}_1) = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} = 26.19$$

A $100(1-\alpha)\%$ CI for β_1 is given by $\hat{\beta}_1 \pm t_{1-\alpha/2}(n-2)\, s.e.(\hat{\beta}_1)$

For $\alpha = .05$, this yields

$$130.4 \pm 2.228(26.19) = (72.04, 188.8)$$

d.) $R^2 = \frac{SSR}{SST} = .7125$

#3 (7.5 in our text)

Show that $\operatorname{var}(\hat{\beta}_0) = \dfrac{\sigma^2 \left(\sum x_i^2/n\right)}{\sum (x_i - \bar{x})^2} = \sigma^2\left[\dfrac{1}{n} + \dfrac{\bar{x}^2}{\sum_i (x_i - \bar{x})^2}\right]$

$$\sigma^2\left[\frac{1}{n} + \frac{\bar{x}}{\sum (x_i - \bar{x})^2}\right] = \sigma^2\left[\frac{\sum (x_i - \bar{x})^2 + n\bar{x}^2}{n \sum (x_i - \bar{x})^2}\right]$$

$$= \sigma^2\left[\frac{\sum x_i^2 - n\bar{x}^2 + n\bar{x}^2}{n \sum (x_i - \bar{x})^2}\right]$$

because $\sum (x_i - \bar{x})^2 = \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$

$= \sum x_i^2 - 2\bar{x}\sum x_i + \bar{x}^2 n$

$= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum x_i^2 - n\bar{x}^2$

$$= \sigma^2 \frac{\sum x_i^2/n}{\sum (x_i - \bar{x})^2}$$

#4. (7.14 from our text)

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + e_i$$

$$= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + e_i - \beta_1 \bar{X}_1 + \beta_1 \bar{X} - \beta_2 \bar{X}_2 + \beta_2 \bar{X}_2$$

$$\cdots - \beta_k \bar{X}_k + \beta_k \bar{X}_k$$

$$= \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 + \cdots + \beta_k \bar{X}_k + \beta_1 (X_{i1} - \bar{X}_1) + \beta_2 (X_{i2} - \bar{X}_2)$$

$$+ \cdots + \beta_k (X_{ik} - \bar{X}_k) + e_i$$

7.17 a) In the uncentered model, $y = X\beta + e$, where $\underset{\sim}{X} = [j_n, X_1]$. The normal equations for this model are ← and $\underset{\sim}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$

$$X^T X \beta = X^T y$$

$$\Rightarrow [j_n, X_1]^T [j_n, X_1] \beta = [j_n, X_1]^T y$$

$$\Rightarrow \begin{pmatrix} j_n^T j_n & j_n^T X_1 \\ X_1^T j_n & X_1^T X_1 \end{pmatrix} \beta = \begin{pmatrix} j_n^T y \\ X_1^T y \end{pmatrix}$$

$$\Rightarrow \begin{cases} n\beta_0 + j_n^T X_1 \beta_1 = n\bar{y} & (1) \\ X_1^T j_n \beta_0 + X_1^T X_1 \beta_1 = X_1^T y & (2) \end{cases}$$

Recall that in the centered model, ~~$y = [j_n, \beta$~~

$$y = [j_n, X_c] \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} + e$$

where $\alpha = \beta_0 + \beta_1 \bar{x}_1 + \dots + \beta_k \bar{x}_k = \beta_0 + \underbrace{\frac{1}{n} j_n^T X_1}_{= \underset{\sim}{\bar{x}}^T} \beta_1$ where $\underset{\sim}{\bar{x}} = \begin{pmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{pmatrix}$

so, dividing through by n in (1) yields

$$\beta_0 + \frac{1}{n} j_n^T X_1 \beta_1 = \bar{y} \quad \text{or} \quad \alpha = \bar{y} \quad (1')$$

which is satisfied by $\hat{\alpha} = \bar{y}$

and (2) $\Leftrightarrow \quad n\underline{\bar{X}}\beta_0 + X_1^T X_1 \beta_1 = X_1^T \underline{y}$ (2')

The normal equations for the centered model take the form

$$[j_n, X_c]^T [j_n, X_c] \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} = [j_n, X_c]^T y$$

$$\Rightarrow \begin{pmatrix} j_n^T j_n & j_n^T X_c \\ X_c^T j_n & X_c^T X_c \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} = \begin{matrix} j_n^T y \\ X_c^T y \end{matrix}$$

$$\Rightarrow \begin{pmatrix} n & 0 \\ 0 & X_c^T X_c \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \end{pmatrix} = \begin{pmatrix} n\bar{y} \\ X_c^T y \end{pmatrix}$$

$$\Rightarrow \begin{cases} n\alpha = n\bar{y} & (3) \text{ equivalent to } (1') \\ X_c^T X_c \beta_1 = X_c^T y & (4) \end{cases}$$

Recall $X_c = (I - \frac{1}{n} J_{n,n}) X_1 = P_{V_0^\perp} X_1$ where $V_0 = \mathcal{L}(j_n)$

$\Rightarrow$

$$X_c^T X_c = X_1^T P_{V_0^\perp} X_1 = X_1^T X_1 - X_1^T P_{V_0} X_1 = X_1^T X_1 - \frac{1}{n} X_1^T j_n j_n^T X_1 = X_1^T X_1 - n\underline{\bar{X}}\,\underline{\bar{X}}^T$$

$$X_c^T y = X_1^T P_{V_0^\perp} y = X_1^T y - X_1^T P_{V_0} y = X_1^T y - \frac{1}{n} X_1^T j_n j_n^T y = X_1^T y - n\underline{\bar{X}}\,\bar{y}$$

$$(4) \Leftrightarrow X_1^T X_1 \beta_1 - n\underline{\bar{x}}\,\underline{\bar{x}}^T \beta_1 = X_1^T y - n\underline{\bar{x}}\,\bar{y}$$

$$\Leftrightarrow X_1^T X_1 \beta_1 - X_1^T y + n\underline{\bar{x}}\,\bar{y} - n\underline{\bar{x}}\,\underline{\bar{x}}^T \beta_1 = 0$$

$$\Leftrightarrow X_1^T X_1 \beta_1 - X_1^T y + n\bar{x}\underbrace{(\bar{y} - \bar{x}^T\beta_1)}_{=\beta_0 \text{ from } (1)} = 0$$

which is equivalent to (2').

b.) $\hat{\beta} = (X^T X)^{-1} X^T y$ in the unconstrained model

where $X = [j_n, X_1]$

$$\Rightarrow (X^T X)^{-1} = \left([j_n, X_1]^T [j_n, X_1]\right)^{-1} = \left(\begin{bmatrix} j_n^T \\ X_1^T \end{bmatrix} [j_n, X_1]\right)^{-1}$$

$$= \begin{pmatrix} j_n^T j_n & j_n^T X_1 \\ X_1^T j_n & X_1^T X_1 \end{pmatrix}^{-1} \xrightarrow{\cancel{\text{using}}} = \begin{pmatrix} n & n\underline{\bar{x}}^T \\ n\underline{\bar{x}} & X_1^T X_1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{n} + \frac{1}{n} n\underline{\bar{x}}^T B^{-1} n\underline{\bar{x}} \frac{1}{n} & -\frac{1}{n} n\underline{\bar{x}}^T B^{-1} \\ -B^{-1} n\underline{\bar{x}} \frac{1}{n} & B^{-1} \end{pmatrix}$$ using formula for inv. of partition Matrix

where $B = X_1^T X_1 - n\underline{\bar{x}} \frac{1}{n} n\underline{\bar{x}}^T$

$$= \begin{pmatrix} \frac{1}{n} + \underline{\bar{x}}^T (X_1^T X_1 - n\underline{\bar{x}}\,\underline{\bar{x}}^T)^{-1} \underline{\bar{x}} & -\bar{x}^T (X_1^T X_1 - n\underline{\bar{x}}\,\underline{\bar{x}}^T)^{-1} \\ -(X_1^T X_1 - n\underline{\bar{x}}\,\underline{\bar{x}}^T)^{-1} \underline{\bar{x}} & (X_1^T X_1 - n\bar{x}\bar{x}^T)^{-1} \end{pmatrix}$$

(8)

$$X_1^T X_1 - n\bar{x}\bar{x}^T = X_1^T\left(I - \tfrac{1}{n} j_n j_n^T\right)X_1$$

$$= X_c^T X_c$$

$$\Rightarrow (X^T X)^{-1} = \begin{pmatrix} \frac{1}{n} + \underline{\bar{x}}^T (X_c^T X_c)^{-1} \underline{\bar{x}} & -\bar{x}^T (X_c^T X_c)^{-1} \\ -(X_c^T X_c)^{-1} \bar{x} & (X_c^T X_c)^{-1} \end{pmatrix}$$

$$\Rightarrow (X^T X)^{-1} X^T y = \begin{pmatrix} \frac{1}{n} + \underline{\bar{x}}^T (X_c^T X_c)^{-1} \underline{\bar{x}} & -\underline{\bar{x}}^T (X_c^T X_c)^{-1} \\ -(X_c^T X_c)^{-1} \underline{\bar{x}} & (X_c^T X_c)^{-1} \end{pmatrix} \begin{pmatrix} j_n^T y \\ X_1^T y \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{1}{n} + \underline{\bar{x}}^T (X_c^T X_c)^{-1} \underline{\bar{x}}\right) n\bar{y} - \underline{\bar{x}}^T (X_c^T X_c)^{-1} X_1^T y \\ -(X_c^T X_c)^{-1} \underline{\bar{x}}\, n\bar{y} + (X_c^T X_c)^{-1} X_1^T y \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} + \underline{\bar{x}}^T (X_c^T X_c)^{-1} \left[\bar{x} n \bar{y} - X_1^T y\right] \\ -(X_c^T X_c)^{-1} X_1 j_n j_n^T \underline{y} / n + (X_c^T X_c)^{-1} X_1^T \underline{y} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} - \underline{\bar{x}}^T (X_c^T X_c)^{-1} \left[X_1^T y - \underline{X}_1^T \underline{j}_n \underline{j}_n^T \underline{y} / n\right] \\ (X_c^T X_c)^{-1} X_1^T \left[I - \frac{1}{n} J_{n,n}\right] y \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} - \bar{x}^T (X_c^T X_c)^{-1} X_1^T \left(I - \frac{1}{n} J_{n,n}\right) \underline{y} \\ (X_c^T X_c)^{-1} X_c^T y \end{pmatrix}$$

$$= \begin{pmatrix} \bar{y} - \underline{\bar{x}}^T \hat{\beta}_1 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \hat{\alpha} - \bar{x}^T \hat{\beta}_1 \\ \hat{\beta}_1 \end{pmatrix} \quad \checkmark$$

7.23 Show $(y-X\beta)^T(y-X\beta) = (y-X\hat{\beta})^T(y-X\hat{\beta}) + (\hat{\beta}-\beta)X^TX(\hat{\beta}-\beta)$

$$(y-X\beta)^T(y-X\beta) = (y-X\hat{\beta}+X\hat{\beta}-X\beta)^T(y-X\hat{\beta}+X\hat{\beta}-X\beta)$$

$$= (y-X\hat{\beta})^T(y-X\hat{\beta}) + (X(\hat{\beta}-\beta))^T(y-X\hat{\beta})$$

$$+ (y-X\hat{\beta})^T(X(\hat{\beta}-\beta)) + (X(\hat{\beta}-\beta))^T X(\hat{\beta}-\beta)$$

$$= (y-X\hat{\beta})^T(y-X\hat{\beta}) + (\hat{\beta}-\beta)X^TX(\hat{\beta}-\beta)$$

$$+ (\hat{\beta}-\beta)^T\underbrace{X^T(I-P_{C(X)})}_{=0}y + y^T\underbrace{(I-P_{C(X)})X}_{=0}(\hat{\beta}-\beta)$$

7.27 Consider two models

(M1) $y = X_1\beta_1 + X_2\beta_2 + e$, ~~e~N~~ $E(e)=0$, $var(e)=\sigma^2 I$

(M2) $y = X_1\beta_1^* + \underline{e}^*$, $E(\underline{e}^*)=0$, $var(\underline{e}^*)=\sigma^2 I$

where X_1 is $n\times k$ of rank k and $X=[X_1, X_2]$ is $n\times(k+1)$ of rank $k+1$

Then the SSEs for the two models are

$$SSE_1 = y^T(I-P_{C(X)})y, \quad SSE_2 = y^T(I-P_{C(X_1)})y$$

$$SSE_2 - SSE_1 = y^T(I - P_{C(X_1)})y - y^T(I - P_{C(X)})y$$

$$= y^T[I - P_{C(X_1)} - I + P_{C(X)}]y$$

$$= y^T[\underbrace{P_{C(X)} - P_{C(X_1)}}_{= P_V}]y \geq 0 \quad \text{(quadratic form}$$

$= P_V$ where $V = C(X_1)^\perp \cap C(X)$ the orthog. complement of $C(X_1)$ w.r.t. $C(X)$

So, $SSE_2 \geq SSE_1$

$$\Rightarrow SST - SSE_2 \leq SST - SSE_1 \Rightarrow SSR_2 \leq SSR_1$$

$$\Rightarrow \frac{SSR_2}{SST} \leq \frac{SSR_1}{SST} \Rightarrow R_2^2 \leq R_1^2$$

7.29 (7.55) $$R^2 = \frac{\hat{\beta}_1^T X_c^T X_c \hat{\beta}_1}{\sum_i (y_i - \bar{y})^2} = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

$$= 1 - \frac{SSE}{\sum (y_i - \bar{y})^2}$$

b.) With the suggested replacements, R^2 becomes

$$1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = 1 - \frac{(n-1)SSE/SST}{n-k-1}$$

$$= \frac{n-k-1-(n-1)SSE/SST}{n-k-1} = \frac{(n-1)[1 - SSE/SST] - k}{n-k-1}$$

$$= \frac{(n-1)R^2 - k}{n-k-1} = R_a^2 \quad (7.58)$$

7.33 a) $$s^2 = \frac{1}{n-k-1}(\underline{y}-\underline{X}\hat{\underline{\beta}})^T \underline{V}^{-1}(\underline{y}-\underline{X}\hat{\underline{\beta}}) \qquad (7.66)$$

$$s^2 = \frac{1}{n-k-1}\,\underline{y}^T[\underline{V}^{-1}-\underline{V}^{-1}\underline{X}(\underline{X}^T\underline{V}^{-1}\underline{X})^{-1}\underline{X}^T\underline{V}^{-1}]\underline{y} \qquad (7.67)$$

show they're equivalent

From 7.66,

$$s^2 = \frac{1}{n-k-1}\left\{\underline{y}^T\underline{V}^{-1}\underline{y} - \underline{y}^T\underline{V}^{-1}\underline{X}\hat{\underline{\beta}} - \hat{\underline{\beta}}^T\underline{X}^T\underline{V}^{-1}\underline{y} + \hat{\underline{\beta}}^T\underline{X}^T\underline{V}^{-1}\underline{X}\hat{\underline{\beta}}\right\}$$

$$= \frac{1}{n-k-1}\left\{y^TV^{-1}y - y^TV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}y\right.$$

$$- y^TV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}y$$

$$\left.+ \underline{y}^T\underline{V}^{-1}X(X^TV^{-1}X)^{-1}\cancel{X^TV^{-1}X(X^TV^{-1}X)^{-1}}X^TV^{-1}y\right\}$$

$$= \frac{1}{n-k-1}\left\{y^TV^{-1}y - y^TV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}y\right\}$$

$$= \frac{1}{n-k-1}\,\underline{y}^T[V^{-1}-V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}]y = s^2 \text{ from } (7.67)$$

7.33b) $E(s^2) = E\left\{\dfrac{y^T[V^{-1} - V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}]y}{n-k-1}\right\}$ from (7.67)

$$= \frac{1}{n-k-1}\left\{E(y^TV^{-1}y) - E(y^TV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}y)\right\}$$

$$= \frac{1}{n-k-1}\Big\{tr(V^{-1}\sigma^2 V) + (X\beta)^TV^{-1}X\beta$$
$$- \left[tr(V^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}\sigma^2 V) + (X\beta)^TV^{-1}X(X^TV^{-1}X)^{-1}X^TV^{-1}X\beta\right]\Big\}$$

$$= \frac{1}{n-k-1}\Big\{\sigma^2 tr(I_n) + \beta^TX^TV^{-1}X\beta - \sigma^2 tr(V^{-1}X(X^TV^{-1}X)^{-1}X^T)$$
$$- \beta^TX^TV^{-1}X\cancel{(X^TV^{-1}X)^{-1}X^TV^{-1}X}\beta\Big\}$$

$$= \frac{1}{n-k-1}\left\{\sigma^2 n - \sigma^2 tr(X^TV^{-1}X(X^TV^{-1}X)^{-1})\right\}$$

(because $tr(AB) = tr(BA)$ if BA is a conformable multiplication)

$$= \frac{1}{n-k-1}\left\{\sigma^2 n - \sigma^2 tr(I_{k+1})\right\} = \frac{\sigma^2(n-k-1)}{n-k-1} = \sigma^2$$

7.34 Complete proof of Thm 7.8B

$$L(\underline{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2} |V|^{1/2}} \exp\left\{\frac{-1}{2\sigma^2}(\underline{y} - \underline{X}\underline{\beta})^T V^{-1}(\underline{y} - \underline{X}\underline{\beta})\right\}$$

max of L w/ respect to $\underline{\beta}, \sigma^2$ same as max of $\ell = \log L$

$$\ell(\underline{\beta}, \sigma^2) = \log L(\underline{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log|V|$$
$$- \frac{1}{2\sigma^2}(\underline{y} - \underline{X}\underline{\beta})^T \underline{V}^{-1}(\underline{y} - \underline{X}\underline{\beta})$$

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}(\underline{y} - \underline{X}\beta)^T \underline{V}^{-1}(\underline{y} - \underline{X}\beta)$$

$$= -\frac{1}{2\sigma^2}\frac{\partial}{\partial \beta}\left[\underline{y}^T \underline{V}^{-1}\underline{y} - y^T V^{-1}\underline{X}\beta - \beta^T \underline{X}^T \underline{V}^{-1} y\right.$$

$$\left. + \beta^T \underline{X}^T V^{-1} X\beta\right] = -\frac{1}{2\sigma^2}\left[-X^T V^{-1} y - X^T V^{-1}\underline{y}\right.$$

$$\left. + 2\underline{X}^T \underline{V}^{-1} X\underline{\beta}\right] = \frac{1}{\sigma^2}\left[X^T \underline{V}^{-1} y - \underline{X}^T \underline{V}^{-1}\underline{X}\underline{\beta}\right] \underline{\text{set}}\ 0$$

$$\Rightarrow \underline{X}^T \underline{V}^{-1}\underline{X}\underline{\beta} = \underline{X}^T \underline{V}^{-1}\underline{y} \Rightarrow \hat{\underline{\beta}} = (\underline{X}^T \underline{V}^{-1}\underline{X})^{-1}\underline{X}^T \underline{V}^{-1}\underline{y}$$

The profile loglikelihood for σ^2 is

$$\ell(\hat{\beta}, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\log|V|$$
$$- \frac{1}{2\sigma^2}\underbrace{(\underline{y} - \underline{X}\hat{\underline{\beta}})^T \underline{V}^{-1}(\underline{y} - \underline{X}\hat{\underline{\beta}})}_{= SSE}$$

7.34 (cont'd)

$$\frac{\partial \ell(\hat{\underline{\beta}}, \sigma^2)}{\partial \sigma^2} = \frac{-n/2}{\sigma^2} + \frac{SSE}{2\sigma^4} \overset{set}{=} 0$$

$$\Rightarrow +(n/2)\sigma^2 = \frac{1}{2} SSE$$

$$\Rightarrow \hat{\sigma}^2 = \frac{SSE}{n}$$

2^{nd} derivative matrices can be checked to verify that these estimators are maximizers of $\ell(\underline{\beta}, \sigma^2)$.

$$\left.\frac{\partial^2 \ell(\underline{\beta}, \sigma^2)}{\partial \underline{\beta}\, \partial \underline{\beta}^T}\right|_{\hat{\underline{\beta}}, \hat{\sigma}^2} = -\frac{1}{\hat{\sigma}^2} \underbrace{X^T V^{-1} X}_{p.d} \Rightarrow \hat{\underline{\beta}} \text{ minimizes } -\ell$$

$$\Rightarrow \hat{\underline{\beta}} \text{ maximizes } \ell$$

$$\left.\frac{\partial^2 \ell(\underline{\beta}, \sigma^2)}{\partial \underline{\beta}\, \partial \sigma^2}\right|_{\hat{\underline{\beta}}, \hat{\sigma}^2} = 0 \quad , \quad \left.\frac{\partial^2 \ell(\underline{\beta}, \sigma^2)}{\partial \sigma^2 \partial \sigma^2}\right|_{\hat{\underline{\beta}}, \hat{\sigma}^2} = -\frac{1}{2}\hat{\sigma}^{-4} n \Rightarrow \hat{\sigma}^2 \text{ maximizes } \ell$$

7.48(7.50) Fitted model: $y_i = \beta_0^* + \beta_1^* x_i + e_i^*$

True model: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i * I(x_i \geq 0) + e_i$

a) $\underline{x} = (-3, -2, -1, 0, 1, 2, 3)^T$

We fit the model $y = X_1\beta_1^* + \underline{e}^*$

where $X_1 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{pmatrix}^T$, $\underline{\beta}_1^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \end{pmatrix}$

When the truth is $y = X_1\beta_1 + X_2\beta_2 + e$ where

$X_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 2 & 3 \end{pmatrix}^T$, $\begin{pmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$

$$E(\hat{\beta}_1^*) = \beta + \underbrace{(X_1^T X_1)^{-1} X_1^T X_2}_{=A} \beta_2$$

Where plugging in $\underline{X}_1, \underline{X}_2$ into the formula for A, we get

$$\underline{A} = \begin{pmatrix} .8571 \\ .5000 \end{pmatrix} \Rightarrow E\begin{pmatrix} \hat{\beta}_0^* \\ \hat{\beta}_1^* \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} .8571 \\ .5000 \end{pmatrix} \beta_2$$

$$= \begin{pmatrix} \beta_0 + .8571\beta_2 \\ \beta_1 + .5000\beta_2 \end{pmatrix}$$

b.) Find $E(S_1^2)$ using the result

$$E(S_1^2) = \sigma^2 + \frac{\beta_2^T X_2^T\left[I - X_1(X_1^T X_1)^{-1} X_1^T\right] X_2 \beta_2}{\cancel{n-k-1}\; n-p-1}$$

Here $n = 7$, $p = 1$, $X_2^T(I - X_1(X_1^T X_1)^{-1} X_1^T) X_2 = 1.8571$

$$\Rightarrow E(S_1^2) = \sigma^2 + \beta_2 (1.8571) \beta_2 / 5 = \sigma^2 + .3714 \beta_2^2$$