Biostat 250A HW6

Elvis Cui

Han Cui

Dept. of Biostat

UCLA

Problem 1-3:

EXERCISES 3a

1. Show that if $\mathbf{X}$ has full rank,
$$(\mathbf{Y}-\mathbf{X}\beta)'(\mathbf{Y}-\mathbf{X}\beta) = (\mathbf{Y}-\mathbf{X}\hat{\beta})'(\mathbf{Y}-\mathbf{X}\hat{\beta}) + (\hat{\beta}-\beta)'\mathbf{X}'\mathbf{X}(\hat{\beta}-\beta),$$
and hence deduce that the left side is minimized uniquely when $\beta = \hat{\beta}$.

2. If $\mathbf{X}$ has full rank, prove that $\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0$. *Hint:* Consider the first column of $\mathbf{X}$.

3. Let
$$\begin{aligned} Y_1 &= \theta + \varepsilon_1 \\ Y_2 &= 2\theta - \phi + \varepsilon_2 \\ Y_3 &= \theta + 2\phi + \varepsilon_3, \end{aligned}$$
where $E[\varepsilon_i] = 0$ $(i = 1, 2, 3)$. Find the least squares estimates of θ and ϕ.

4. Consider the regression model
$$E[Y_i] = \beta_0 + \beta_1 x_i + \beta_2(3x_i^2 - 2) \qquad (i = 1, 2, 3),$$
where $x_1 = -1$, $x_2 = 0$, and $x_3 = +1$. Find the least squares estimates of β_0, β_1, and β_2. Show that the least squares estimates of β_0 and β_1 are unchanged if $\beta_2 = 0$.

Sol. (1): Let $\hat{\beta} = (X^TX)^{-1}X^Ty$, then

$$(Y-X\beta)^T(Y-X\beta)$$
$$= (Y-X\hat{\beta}+X\hat{\beta}-X\beta)^T(Y-X\hat{\beta}+X\hat{\beta}-X\beta)$$
$$= (Y-X\hat{\beta})^T(Y-X\hat{\beta}) + (\hat{\beta}-\beta)^TX^TX(\hat{\beta}-\beta)$$
$$+ \boxed{2(\hat{\beta}-\beta)^TX^T(y-X\hat{\beta})} \quad (\triangle)$$

Since $X^TX\hat{\beta} = X^Ty$ (normal eq.)

we have $(\triangle) = 0$

Thus,

$(Y-X\beta)^T(Y-X\beta)$ equals to

$$\|Y-X\hat{\beta}\|_2^2 + \|X\hat{\beta}-X\beta\|_2^2 \qquad \square$$

(3): Write them in matrix form:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}}\begin{pmatrix} \theta \\ \phi \end{pmatrix}$$

Denote $\begin{pmatrix} 1 & 0 \\ 2 & -1 \\ 1 & 2 \end{pmatrix}$ as X, then

$$\widehat{\begin{pmatrix} \theta \\ \phi \end{pmatrix}} = (X^TX)^{-1}X^Ty, \quad y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$$

which is

$$\hat{\theta}_{LSE} = \frac{1}{6}(Y_1+2Y_2+Y_3); \quad \hat{\phi}_{LSE} = \frac{1}{5}(-Y_1+2Y_3).$$

(4): Let $Z_i := 3X_i^2 - 2$, $y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}$, then

the design matrix is:

$$X = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \\ 1 & 1 & 1 \end{bmatrix}, \quad \text{thus,}$$

$$\widehat{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{LSE} = (X^TX)^{-1}X^Ty = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{pmatrix} y$$

Now let $\beta_2 = 0$, then the design mat:

$$\tilde{X} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad \text{thus}$$

$$\widehat{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{\tilde{X}} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^Ty = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

Therefore, $\hat{\beta}_{0_{LSE}}$ & $\hat{\beta}_{1_{LSE}}$ remain the same.

Problem 4:

4. Let

$$Y_i = \beta_0 + \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \varepsilon_i \qquad (i = 1, 2, \ldots, n),$$

where $\overline{x}_j = \sum_{i=1}^n x_{ij}/n$, $E[\boldsymbol{\varepsilon}] = \mathbf{0}$, and $\mathrm{Var}[\boldsymbol{\varepsilon}] = \sigma^2 \mathbf{I}_n$. If $\hat{\beta}_1$ is the least squares estimate of β_1, show that

$$\mathrm{var}[\hat{\beta}_1] = \frac{\sigma^2}{\sum_i (x_{i1} - \overline{x}_1)^2 (1 - r^2)},$$

where r is the correlation coefficient of the n pairs (x_{i1}, x_{i2}).

Sol. Let $\underset{n\times 3}{X} = \begin{pmatrix} 1 & x_{11}-\bar{x}_1 & x_{12}-\bar{x}_2 \\ \vdots & \vdots & \vdots \\ 1 & x_{n1}-\bar{x}_1 & x_{n2}-\bar{x}_2 \end{pmatrix}$

Then $\widehat{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{LSE} = (X^TX)^{-1}X^Ty$

where $(X^TX)^{-1} = \begin{pmatrix} n & 0 & 0 \\ 0 & a & b \\ 0 & b & c \end{pmatrix}^{-1}$,

$a = \sum_{i=1}^n (x_{i1}-\bar{x}_1)^2$, $b = \sum_{i=1}^n (x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)$,

$c = \sum_{i=1}^n (x_{i2}-\bar{x}_2)^2$

By inverse formula for partition matrices,

$(X^TX)^{-1} = \begin{pmatrix} d & 0 \\ 0 & E \end{pmatrix}$ where

$d = \frac{1}{n}$ & $E = \begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} = \frac{1}{ac-b^2}\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$

$\Rightarrow \mathrm{Var}\widehat{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{LSE} = \sigma^2 (X^TX)^{-1}$

$= \sigma^2 \begin{pmatrix} \frac{1}{n} & 0 \\ 0 & E \end{pmatrix}$

& $\mathrm{Var}(\hat{\beta}_1)_{LSE} = \frac{\sigma^2 c}{ac-b^2}$

However,

$r = \frac{b}{\sqrt{ac}} \Rightarrow$

$1 - r^2 = \frac{ac-b^2}{ac} \Rightarrow$

$\mathrm{Var}(\hat{\beta}_1)_{LSE} = \frac{\sigma^2}{a(1-r^2)}$

where $a = \sum_{i=1}^n (x_{i1}-\bar{x}_1)^2$ $\square$.

Problem 5:

3. Suppose that $\mathbf{X} = (\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(p-1)}, \mathbf{x}^{(p)}) = (\mathbf{W}, \mathbf{x}^{(p)})$ has linearly independent columns.

   (a) Using A.9.5, prove that
   $$\det(\mathbf{X}'\mathbf{X}) = \det(\mathbf{W}'\mathbf{W})\left(\mathbf{x}^{(p)\prime}\mathbf{x}^{(p)} - \mathbf{x}^{(p)\prime}\mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{x}^{(p)}\right).$$

   (b) Deduce that
   $$\frac{\det(\mathbf{W}'\mathbf{W})}{\det(\mathbf{X}'\mathbf{X})} \geq \frac{1}{\mathbf{x}^{(p)\prime}\mathbf{x}^{(p)}},$$
   and hence show that $\mathrm{var}[\hat{\beta}_p] \geq \sigma^2(\mathbf{x}^{(p)\prime}\mathbf{x}^{(p)})^{-1}$ with equality if and only if $\mathbf{x}^{(p)\prime}\mathbf{x}^{(j)} = 0$ $(j = 0, 1, \ldots, p-1)$.

   (Rao [1973: p. 236])

Sol. (a): $X^TX = \begin{pmatrix} W^T \\ x^{(p)T} \end{pmatrix}\begin{pmatrix} W & x^{(p)} \end{pmatrix}$

$$= \begin{pmatrix} W^TW & W^Tx^{(p)} \\ x^{(p)T}W & x^{(p)T}x^{(p)} \end{pmatrix}$$

Thus, by the formula of determinant of Partition matrices, we have

$$\det|X^TX| = \det|W^TW| \det|A_{22\cdot1}|$$

where $A_{22\cdot1} = x^{(p)T}x^{(p)} - x^{(p)T}W(W^TW)^{-1}Wx^{(p)}$.

But this is exactly the desired result Since $A_{22\cdot1}$ is a real number.

(b): WLOG, suppose $\mathrm{Var}\, Y = I$, if not, rescale y to y/σ.

Denote $P_W = W(W^TW)^{-1}W^T$

Then P_W is an orthogonal projection. Thus, $\forall x$, we have

$$x^TP_Wx \geq 0$$

Thus by part (a):

$$\frac{\det|W^TW|}{\det|X^TX|} = \frac{1}{x^{(p)T}x^{(p)} - A_{22\cdot1}} \geq \frac{1}{x^{(p)T}x^{(p)}}.$$

Moreover,

$$\mathrm{Var}\,\hat\beta = (X^TX)^{-1}$$

$\Rightarrow \mathrm{Var}\,\hat\beta_p = (X^TX)^{-1}_{(pp)}$ where (pp) means the p^{th} row p^{th} column of a specified matrix.

Now, by the formula of inverse of partitioned matrices, we have

$(X^TX)^{-1} = \begin{pmatrix} B & C \\ C & A \end{pmatrix}$ where $A = A_{22\cdot1}^{-1} = \frac{1}{A_{22\cdot1}}$, thus,

$\mathrm{Var}\,\hat\beta_p = (x^{(p)T}x^{(p)} - x^{(p)T}W(W^TW)^{-1}W^Tx^{(p)})^{-1} \geq \frac{1}{x^{(p)T}x^{(p)}}$ with "=" iff

$W^Tx^{(p)} = 0$ or in other words, $x^{(j)T}x^{(p)} = 0 \ \forall j \neq p$.

Problem 6:

THEOREM 3.6 *Let* $\mathbf{R}_G = \mathbf{I}_n - \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$, $\mathbf{L} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}$, $\mathbf{M} = (\mathbf{Z}'\mathbf{R}\mathbf{Z})^{-1}$, *and*

$$\hat{\boldsymbol{\delta}}_G = \begin{pmatrix} \hat{\boldsymbol{\beta}}_G \\ \hat{\boldsymbol{\gamma}}_G \end{pmatrix}.$$

Then:

(i) $\hat{\boldsymbol{\gamma}}_G = (\mathbf{Z}'\mathbf{R}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}\mathbf{Y}$.

(ii) $\hat{\boldsymbol{\beta}}_G = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}_G) = \hat{\boldsymbol{\beta}} - \mathbf{L}\hat{\boldsymbol{\gamma}}_G$.

(iii) $\mathbf{Y}'\mathbf{R}_G\mathbf{Y} = (\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}_G)'\mathbf{R}(\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\gamma}}_G) = \mathbf{Y}'\mathbf{R}\mathbf{Y} - \hat{\boldsymbol{\gamma}}_G'\mathbf{Z}'\mathbf{R}\mathbf{Y}$.

(iv)

$$\mathrm{Var}[\hat{\boldsymbol{\delta}}_G] = \sigma^2 \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1} + \mathbf{LML}' & -\mathbf{LM} \\ -\mathbf{ML}' & \mathbf{M} \end{pmatrix}. \quad (3.25)$$

3. If $\hat{\boldsymbol{\beta}}_G = (\hat{\beta}_{G,j})$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_j)$, use Theorem 3.6(iv) to prove that

$$\mathrm{var}[\hat{\beta}_{G,j}] \geq \mathrm{var}[\hat{\beta}_j].$$

Sol. WLOG, assume $\sigma = 1$, then

$\mathrm{Var}\,\hat{\beta}_G = (X^TX)^{-1} + LML^T$ &

$\mathrm{Var}\,\hat{\beta} = (X^TX)^{-1}$.

Since (Z^TRZ) is p.s.d. $\Rightarrow$ $(Z^TRZ)^{-1}$ is also p.s.d. by Spectral thm.

$\Rightarrow$ $M = (Z^TRZ)^{-1}$ is p.s.d. $\Rightarrow$ LML^T is p.s.d. Since $\forall x$

$x^TLML^Tx = (L^Tx)^TM(L^Tx) \geq 0$.

$\Rightarrow$ Diagonal elements of LML^T are non-negative by properties of p.s.d. matrices.

$\Rightarrow$ $\mathrm{Var}\,\hat{\beta}_{G,j} \geq \mathrm{Var}(\hat{\beta}_j)$

j^{th} element of $(X^TX)^{-1} + LML^T$

j^{th} element of $(X^TX)^{-1}$.

Problem 7-10:

EXERCISES 3k

1. Let $Y_i = \beta x_i + \varepsilon_i$ $(i = 1, 2)$, where $\varepsilon_1 \sim N(0, \sigma^2)$, $\varepsilon_2 \sim N(0, 2\sigma^2)$, and ε_1 and ε_2 are statistically independent. If $x_1 = +1$ and $x_2 = -1$, obtain the weighted least squares estimate of β and find the variance of your estimate.

2. Let Y_i $(i = 1, 2, \ldots, n)$ be independent random variables with a common mean θ and variances σ^2/w_i $(i = 1, 2, \ldots, n)$. Find the linear unbiased estimate of θ with minimum variance, and find this minimum variance.

3. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables, and let Y_i have a $N(i\theta, i^2\sigma^2)$ distribution for $i = 1, 2, \ldots, n$. Find the weighted least squares estimate of θ and prove that its variance is σ^2/n.

4. Let $Y_1, Y_2, \ldots, Y_n$ be random variables with common mean θ and with dispersion matrix $\sigma^2\mathbf{V}$, where $v_{ii} = 1$ $(i = 1, 2, \ldots, n)$ and $v_{ij} = \rho$ $(0 < \rho < 1;\ i, j = 1, 2, \ldots, n;\ i \neq j)$. Find the generalized least squares estimate of θ and show that it is the same as the ordinary least squares estimate. *Hint*: $\mathbf{V}^{-1}$ takes the same form as $\mathbf{V}$.

(McElroy [1967])

Sol. 1. if $\varepsilon_1 \perp \varepsilon_2$, then $\mathrm{Cov}\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix} = \begin{pmatrix}\sigma^2 & 0\\ 0 & 2\sigma^2\end{pmatrix} \triangleq V$

Let $x = \begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}+1\\ -1\end{pmatrix}$, then WLSE of β is

$$\hat{\beta}_{WLSE} = (x^T V^{-1} x)^{-1} x^T V^{-1} y$$

$$= \left((1\ -1)\begin{pmatrix}\frac{1}{\sigma^2} & \\ & \frac{1}{2\sigma^2}\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix}\right)^{-1} (1\ -1)^T \begin{pmatrix}\frac{1}{\sigma^2} & \\ & \frac{1}{2\sigma^2}\end{pmatrix}\begin{pmatrix}Y_1\\ Y_2\end{pmatrix}$$

$$= \frac{2}{3}Y_1 - \frac{1}{3}Y_2$$

2. By Gauss-Markov, $a^T\hat{\beta}_{WLSE}$ is the unique BLUE of $a^T\beta$.

Let $V = \sigma^2\begin{pmatrix}\frac{1}{w_1} & & \\ & \ddots & \\ & & \frac{1}{w_n}\end{pmatrix}$, then

$$\hat{\beta}_{WLSE} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$= (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\tilde{Y} \quad \text{where}$$

$$\tilde{X} = \begin{pmatrix}\sqrt{w_1} & & \\ & \ddots & \\ & & \sqrt{w_n}\end{pmatrix} X$$

$$\tilde{Y} = \begin{pmatrix}\sqrt{w_1} & & \\ & \ddots & \\ & & \sqrt{w_n}\end{pmatrix} Y$$

$\mathrm{Var}\,\hat{\beta}_{WLSE} = \sigma^2(\tilde{X}^T\tilde{X})^{-1}$ by the result given by OLSE.

Let the model be

3. $Y_i = i\theta + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, i^2\sigma^2)$

& Let $V = \begin{pmatrix}\sigma^2 & & \\ & 2^2\sigma^2 & \\ & & \ddots \\ & & & n^2\sigma^2\end{pmatrix}$

$$X = \begin{pmatrix}1\\ \vdots\\ n\end{pmatrix}$$

$\Rightarrow$ WLSE of θ is

$$\hat{\theta}_{WLSE} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i}{i}$$

$$\Rightarrow \mathrm{Var}\,\hat{\theta}_{WLSE} = \frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{i^2}\mathrm{Var}\,Y_i$$

$$= \frac{1}{n}\sigma^2.$$

4. First, WLOG, assume $\sigma^2 = 1$, Then

$$V = \begin{bmatrix}1 & \rho & \cdots & \rho\\ \rho & \ddots & & \vdots\\ \vdots & & \ddots & \rho\\ \rho & \cdots & \rho & 1\end{bmatrix} = (1-\rho)I_n + \rho\mathbf{1}\mathbf{1}^T$$

where $\mathbf{1} \in \mathbb{R}^n$ is the vector with all 1's.

Next, $V^{-1} = \left((1-\rho)I_n + \rho\mathbf{1}\mathbf{1}^T\right)^{-1}$ HW 3 Question 4

$$= \frac{1}{1-\rho}I_n - c\mathbf{1}\mathbf{1}^T,$$

$$c = \rho/((1-\rho)^2 + n\rho(1-\rho)).$$

Then since $EY = \theta$ so let $X = \begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$.

$$\Rightarrow \hat{\theta}_{WLSE} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

$$= \left[n\left(\frac{1}{1-\rho} - cn\right)\right]^{-1}\left[\frac{1}{1-\rho} - cn\right]\left[\Sigma_{i=1}^{n} Y_i\right]$$

$$= \frac{1}{n}\Sigma_{i=1}^{n} Y_i$$

But $\hat{\theta}_{OLSE} = (X^TX)^{-1}X^TY = \frac{1}{n}\Sigma_{i=1}^{n} Y_i = \hat{\theta}_{WLSE}$. $\square$

Problem 11:

We note that the centering and subsequent reparameterization of the model do not affect the fitted model $\hat{\mathbf{Y}}$, so that the residuals for both the centered and uncentered models are the same. Hence, from (3.52), the residual sum of squares for both models is given by

$$\begin{aligned}
\text{RSS} &= \mathbf{Y}'(\mathbf{I}_n - \mathbf{P})\mathbf{Y} \\
&= \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n' - \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}\right)\mathbf{Y} \\
&= \mathbf{Y}'\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{Y} - \mathbf{Y}'\tilde{\mathbf{P}}\mathbf{Y} \\
&= \sum_i (Y_i - \overline{Y})^2 - \mathbf{Y}'\tilde{\mathbf{P}}\mathbf{Y}, \qquad (3.54)
\end{aligned}$$

where $\tilde{\mathbf{P}} = \tilde{\mathbf{X}}(\tilde{\mathbf{X}}'\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}'$. We will use this result later.

EXERCISES 3I

1. If $\tilde{Y}_i = Y_i - \overline{Y}$ and $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \ldots, \tilde{Y}_n)'$, prove from (3.54) that RSS $= \tilde{\mathbf{Y}}'(\mathbf{I}_n - \tilde{\mathbf{P}})\tilde{\mathbf{Y}}$.

Sol. $\tilde{Y} = Y - \frac{1}{n}11^T y$, where 1, as usual, is the vector with all ones.

Next, note that

$$1^T\tilde{P} = 1^T \underset{n\times p}{\tilde{X}} (\underset{p\times n}{\tilde{X}^T}\underset{n\times p}{\tilde{X}})^{-1}\underset{p\times n}{\tilde{X}^T} = 0$$

Since $1^T\tilde{X} = 1^T[x_1-\bar{x}_1, \ldots, x_p-\bar{x}_p]$ & $1^T(x_i-\bar{x}_i)=0$

$$\Rightarrow \tilde{Y}^T\tilde{P}\tilde{Y} = (Y-1\bar{y})^T\tilde{P}(Y-1\bar{y})$$

$$= Y^T\tilde{P}Y \underbrace{-2\bar{y}1^T\tilde{P}Y}_{=0} + \underbrace{\bar{y}1^T\tilde{P}1\bar{y}}_{=0}$$

$$= Y^T\tilde{P}Y$$

Thus, $\text{RSS} = \tilde{Y}^T\tilde{Y} - Y^T\tilde{P}Y$

$$= \tilde{Y}^T\tilde{Y} - \tilde{Y}^T\tilde{P}\tilde{Y}$$

$$= \tilde{Y}(I-\tilde{P})Y. \qquad \square$$