Homework 7 - Answers

- From your textbook:

1. From the class notes, $\hat{\beta} = (X^T X)^{-1} X^T Y$. But

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{where} \quad x_i = (1 \;\; x_{t1} \cdots x_{tK}) \text{ a } 1 \times (K+1) \text{ vector.}$$

Hence, $X^T X = [x_1^T \cdots x_n^T] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{t=1}^{n} x_t^T x_t$ and

$X^T Y = [x_1^T \cdots x_n^T] \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{t=1}^{n} x_t^T y_t$. Consequently,

$$\hat{\beta} = \left(\sum_{t=1}^{n} x_t^T x_t \right)^{-1} \left(\sum_{t=1}^{n} x_t^T y_t \right)$$

$$= \left(\frac{1}{n} \sum_{t=1}^{n} x_t^T x_t \right)^{-1} \left(\frac{1}{n} \sum_{t=1}^{n} x_t^T y_t \right)$$

2. (i) $y - Xb = y - Xb + X\hat{\beta} - X\hat{\beta}$

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

$$= \hat{u} + X(\hat{\beta} - b)$$

Then, $(y - Xb)^T(y - Xb) = [\hat{u}^T + (\hat{\beta}-b)^T X^T][\hat{u} + X(\hat{\beta}-b)]$

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

But $X^T\hat{u} = X^T(y - X\hat{\beta}) = X^Ty - X^TX(X^TX)^{-1}X^Ty = 0.$

$$(y - Xb)^T(y - Xb) = \hat{u}^T\hat{u} + (\hat{\beta}-b)^T X^T X(\hat{\beta}-b). \quad (1)$$

(ii) Recall that $\hat{\beta}$ minimizers $S_n(\beta) = (y - X\beta)^T(y - X\beta)$.

Equation (1) shows that for any vector b,

$$S_n(b) = S_n(\hat{\beta}) + (\hat{\beta}-b)^T X^T X(\hat{\beta}-b).$$

$$= S_n(\hat{\beta}) + [X(\hat{\beta}-b)]^T X(\hat{\beta}-b)$$

Letting $v = X(\hat{\beta}-b)$, we have that

$$S_n(b) = S_n(\hat{\beta}) + \sum_{i=1}^{n} v_i^2$$

Since, $\sum_{i=1}^{n} v_i^2 \geq 0$, $S_n(b) \geq S_n(\hat{\beta})$. Since X is of full column rank, equality happens only when $b = \hat{\beta}$. Otherwise, i.e.,

$b \neq \hat{\beta} \Rightarrow S_n(b) > S_n(\hat{\beta})$.

5. (i) $E(\tilde{\beta}|X) = (Z^T X)^{-1} Z^T E(Y|X)$, since Z is a function of X

$= (Z^T X)^{-1} Z^T X\beta = \beta$, since $E(Y|X) = X\beta$.

(ii) $V(\tilde{\beta}|X) = E\{(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)^T | X\}$.

But $\tilde{\beta} = (Z^T X)^{-1} Z^T (X\beta + u) = \beta + (Z^T X)^{-1} Z^T u$. This implies

$\tilde{\beta} - \beta = (Z^T X)^{-1} Z^T u$. Then,

$$V(\tilde{\beta}|X) = E\{(Z^T X)^{-1} Z^T u u^T Z (X^T Z)^{-1} | X\}$$

$$= \sigma^2 (Z^T X)^{-1} Z^T Z (X^T Z)^{-1}$$

We would prefer the estimator with smaller variance, given that both are unbiased. Consider,

$$V(\tilde{\beta}|X) - V(\hat{\beta}|X) = \sigma^2 (Z^T X)^{-1} Z^T Z (X^T Z)^{-1} - \sigma^2 (X^T X)^{-1}$$

$$= \sigma^2 \{(Z^T X)^{-1} Z^T Z (X^T Z)^{-1} - (X^T X)^{-1}\}$$

$$= \sigma^2 \{(Z^T X)^{-1} Z^T Z (X^T Z)^{-1} - \underbrace{(Z^T X)^{-1} Z^T X}_{I} (X^T X)^{-1} \underbrace{X^T Z (X^T Z)^{-1}}_{I}\}$$

$$= \sigma^2 \{(Z^T X)^{-1} Z^T [I - X(X^T X)^{-1} X^T] Z (X^T Z)^{-1}\}$$

$$= \sigma^2 \{A^T M_X A\}, \text{ where } A = Z(X^T Z)^{-1}$$

Because M_X is symmetric and idempotent, $V(\tilde{\beta}|X) - V(\hat{\beta}|X) \geq 0$. The inequality is in the sense that $\sigma^2\{A^T M_X A\}$ is positive semi-definite.

8.

(i) $M^T = \left(I - X(X^TX)^{-1}X^T\right)^T = I - X(X^TX)^{-1}X^T = M.$

$$MM = (I - X(X^TX)^{-1}X^T)(I - X(X^TX)^{-1}X^T)$$
$$= I - X(X^TX)^{-1}X^T - X(X^TX)^{-1}X^T + X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T$$
$$= I - X(X^TX)^{-1}X^T = M$$

(ii) Let $X_{i\cdot}$ be the i^{th} rows of X. Recall that

$\hat{u} = y - X\hat{\beta} = y - X(X^TX)^{-1}X^Ty = My = M(X\beta+u) = Mu$. Then,

$$V(\hat{u}|X) = E\{(\hat{u} - E(\hat{u}|X))(\hat{u} - E(\hat{u}|X))^T|X\}$$
$$= E\{Mu(Mu)^T|X\}, \quad \text{since } E(\hat{u}|x) = E(Mu|X) = 0$$
$$= M\sigma^2 I M = \sigma^2 M. \qquad (2)$$

Since $\sigma^2 > 0$ and M is a variance matrix, all elements of the diagonal of M, i.e., M_{ii} for all $i = 1, 2, \ldots, n$ are such that $M_{ii} \geq 0$.

Also, the i^{th} diagonal element of $X(X^TX)^{-1}X^T$ is given by $\underbrace{X_{i\cdot}}_{1\times(K+1)}(X^TX)^{-1}X_{i\cdot}^T$. But since $(X^TX)^{-1}$ is also a variance matrix (Recall $V(\hat{\beta}|X) = \sigma^2(X^TX)^{-1}$). $X_{i\cdot}(X^TX)^{-1}X_{i\cdot}^T \geq 0$ for all i.

Thus, $-X_{i\cdot}(X^TX)^{-1}X_{i\cdot}^T \leq 0$ and $1 - X_{i\cdot}(X^TX)^{-1}X_{i\cdot}^T \leq 1$. Since

$M_{ii} = 1 - X_{i\cdot}(X^TX)^{-1}X_{i\cdot}^T$, $\quad M_{ii} \leq 1$.

(iii) Proved in part (ii). See (2).

(iv) Follows from the fact that $M_{ii} \neq M_{jj}$ and $M_{ij} \neq 0$.

2. MATLAB question.

See code HOLSWage2.m.

[redacted] - [redacted].

Question 3:

(a) Problem 2:

(i) $H_0: \beta_3 = 0$

$H_A: \beta_3 > 0$

(ii) Note that $\frac{d \log(\text{salary})}{d\text{ros}} = \frac{1}{\text{salary}} \frac{d\,\text{salary}}{d\text{ros}}$

and $\frac{d\text{salary}}{\text{salary}} \approx$ % change in salary. Hence, if

$d\text{ros} = 50$, and given that $\frac{d \log(\text{salary})}{d\text{ros}} = 0.00024$

% salary $= 50 \times 0.00024 = 0.012$.. Arguably a small effect on salary.

(iii) No effect, means $\beta_3 = 0$. The alternative is $H_A: \beta_3 > 0$

$n = 209$, $K+1 = 4$, hence degrees of freedom $= 205$. Choose $\alpha = 0.1$ (10%). This is a one-sided test, ($\beta_3 > 0$). Hence,

$t^*_{1-\alpha, n-(K+1)} = t^*_{.9, 205} = 1.2857$ (Use tinv in MATLAB)

Now, $t = \frac{0.00024}{0.00054} = 0.444 < 1.2857 = t^*_{0.9, 205}$

Hence, we reject $\beta_3 = 0$.

(iv) Yes. The fact that the test in (iii) rejects the H_0 suggests that "ros" should be included in the regression.

Problem 8: (i) $V(\hat\beta_1 - 3\hat\beta_2) = E\left([\hat\beta_1 - 3\hat\beta_2 - (\beta_1 - 3\beta_2)]^2\right)$

$= E\{(\hat\beta_1 - \beta_1) - 3(\hat\beta_2 - \beta_2)\}^2$

$= E(\hat\beta_1 - \beta_1)^2 + 9E(\hat\beta_2 - \beta_2)^2 - 3E(\hat\beta_1 - \beta_1)(\hat\beta_2 - \beta_2)$

$= V(\hat\beta_1) + 9V(\hat\beta_2) - 3\,\text{Cov}(\hat\beta_1, \hat\beta_2)$

The standard-error is $\sqrt{V(\hat\beta_1) + 9V(\hat\beta_2) - 3\,\text{cov}(\hat\beta_1, \hat\beta_2)}$

(ii) Given that $H_0: \beta_1 - 3\beta_2 = 1$, we write

$R = [0\ 1\ \ -3\ \ 0]$ and $r = 1$. Hence,

$$R\beta = r \iff [0\ \ 1\ \ -3\ \ 0]\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\end{bmatrix} = 1. \text{ Then,}$$

$$(R\hat{\beta} - r)^T \frac{\left(R(X^TX)^{-1}R^T\right)^{-1}}{\hat{\sigma}^2}(R\hat{\beta} - r) \sim t_{1-\alpha/2,\, n-4}$$

(iii) Under H_0, $\beta_1 = \theta_1 + 3\beta_2$. Hence,

$$y = \beta_0 + [\theta_1 + 3\beta_2]x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$
$$= \beta_0 + \theta_1 x_1 + (3x_1 + x_2)\beta_2 + \beta_3 x_3 + u \qquad (1)$$

We can run the regression (1) and directly test if $\theta_1 = 1$ using a t-statistic.

Question 4:

1 (b). Problem 2: First, note that if $y^T = (y_1 \cdots y_n)$ we can define $c_0 y$. Let $C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & c_1 & 0 & \cdots & 0 \\ \vdots & & c_2 & & \vdots \\ \vdots & & & \ddots & \vdots \\ 0 & \cdots & & 0 & c_k \end{bmatrix}_{(k+1)\times(k+1)}$; and note that

$$C^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1/c_1 & & & \vdots \\ & & 1/c_2 & & \vdots \\ & & & \ddots & \\ 0 & \cdots & & & 1/c_k \end{bmatrix}. \text{ Also, if } X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{bmatrix}$$

$$\text{Then, } XC = \begin{bmatrix} 1 & c_1 X_{11} & \cdots & c_k X_{1k} \\ 1 & c_1 X_{21} & \cdots & c_k X_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & c_1 X_{n1} & \cdots & c_k X_{nk} \end{bmatrix}.$$

$$\text{Hence, } \tilde{\beta} = \left((XC)^T XC\right)^{-1}(XC)^T c_0 y$$
$$= (C^T X^T X C)^{-1} C^T X^T c_0 y$$
$$= C^{-1}(X^TX)^{-1}(C^T)^{-1} C^T X^T c_0 y$$
$$= c_0 C^{-1}(X^TX)^{-1} X^T y \quad \text{since } c_0 \text{ is a scalar}$$

$$= \begin{bmatrix} c_0 & & & \\ & c_0/c_1 & & \\ & & \ddots & \\ & & & c_0/c_K \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_K \end{bmatrix}$$

$$= \begin{bmatrix} c_0 \hat{\beta}_0 \\ \frac{c_0}{c_1} \hat{\beta}_1 \\ \vdots \\ \frac{c_0}{c_K} \hat{\beta}_K \end{bmatrix}$$

1(b) Problem 4:

(i) $\frac{d \log(\text{wage})}{d \text{Educ}} = \beta_1 + \beta_2 \text{paredu}c$

Here, the percent change on wages due to a change of year of education is a linear function of the parents years of education. If $\beta_2 > 0$, this implies that % wage gains from increasing the years of education is larger for those with more educated parents. It is unclear what might justify such assumption, or for that matter the reverse, i.e., $\beta_2 < 0$.

(ii) $\widehat{\frac{d \log(\text{wages})}{d \text{Educ}}} = 0.047 + 0.00078 \text{ educ} * \text{pareduc}$

For two individuals with the same level of education, and one with college educated parents and, the other with high school educated parents, we have

$\widehat{\frac{d \log(\text{wages})}{d \text{Educ}}} = 0.047 + 0.00078 \text{ Educ} \cdot 24$ (HSE).

$\widehat{\frac{d \log(\text{wages})}{d \text{Educ}}} = 0.047 + 0.00078 \cdot \text{Educ} \cdot 32$ (CE)

$(CE) - (HSE) = [0.02496 - 0.01872] \text{ Educ}$

$= 0.00624 \text{ Educ} \rightarrow$ difference on returns from education

(iii)

$$\frac{\widehat{d \log(\text{wages})}}{d\,Educ} = 0.097 - 0.0016\,pareduc$$

The impact of another year of education depens negatively on parents education.

$H_0: \beta_3 = 0$; $H_A: \beta_3 \neq 0$. Degrees of freedom $= 722 - 6 = 716$
At $\alpha = 0.05$, we have $t = -0.0016/0.0012 = -1.33$ and
$t^*_{716,\,0.975} = 1.9633$. Hence, since $-1.9633 < -1.33$, we accept H_0 that $\beta_3 = 0$.

Question 5:

(a) Here $R = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$, $r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$F^*_{2,522} = 3.013$, $F = 111.8$. Reject H_0.

(b) Here $R = [\,0 \quad 0 \quad 0 \quad 1\,]$, $r = [0]$
$F^*_{1,522} = 3.8593$, $F = 37.99$, Reject H_0.

(c) Here $H_0: \beta_1 - \beta_2 - \beta_3 = 0$, so

$$R = [\,0 \quad 1 \quad -1 \quad -1\,], \quad r = [0].$$

$F^*_{1,522} = 3.8593$, $F = 28.83$. Reject H_0.

Question 6:

Here is a list of the needed assumptions:

If we write the model in matrix form, i.e.,

$Y = X\beta + u$, then:

1. $E(u|X) = 0$. This guarantees that $E(R\hat{\beta}|X) = R\beta$ and $R\beta = r$ under H_0.

2. $E(uu^T|X) = \sigma^2 I_n$. This guarantees that

$$V(R\hat{\beta} - r|X) = \sigma^2 (R(X^TX)^{-1}R^T)$$

3. $U \sim N(0, \sigma^2 I)$, This guarantees that

f the ratio of two (independent) χ^2 random variables divided by their degrees of freedom. m for the numerator and $n-(k+1)$ for the denominator.