Chapter 05

1.

The Ordinary Least Squares (OLS) Estimators

$$b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (2.7)$$

$$b_1 = \bar{y} - b_2\bar{x} \quad (2.8)$$

where $\bar{y} = \sum y_i/N$ and $\bar{x} = \sum x_i/N$ are the sample means of the observations on y and x.

Let $k=2$. Thus, $Y_i = b_1 + b_2 X_i + e_i$.

Besides, $X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}$, $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$.

And, $X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}$

$$(X'X)^{-1} = \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} \Sigma y_i \\ \Sigma X_i y_i \end{bmatrix}$$

Therefore, $b = (X'X)^{-1}(X'Y) = \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix} \begin{bmatrix} \Sigma y_i \\ \Sigma X_i y_i \end{bmatrix}$

$$= \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 \Sigma y_i - \Sigma X_i \Sigma X_i y_i \\ -\Sigma X_i \Sigma y_i + n\Sigma X_i y_i \end{bmatrix}$$

Hence, $b_2 = \frac{-\Sigma X_i \Sigma y_i + n\Sigma X_i y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} = \frac{-n^2 \bar{x}\bar{y} + n\Sigma X_i y_i}{n\Sigma X_i^2 - n^2\bar{x}^2} = \frac{-n\bar{x}\bar{y} + \Sigma X_i y_i}{\Sigma X_i^2 - n\bar{x}^2}$

$$= \frac{(-2n\bar{x}\bar{y} + n\bar{x}\bar{y}) + \Sigma X_i y_i}{\Sigma X_i^2 + (-2n\bar{x}^2 + n\bar{x}^2)} = \frac{(-\Sigma X_i \bar{y} - \bar{x}\Sigma y_i) + n\bar{x}\bar{y} + \Sigma X_i y_i}{\Sigma X_i^2 - 2\bar{x}\Sigma X_i + n\bar{x}^2}$$

$$= \frac{\Sigma X_i y_i - \Sigma X_i \bar{y} - \bar{x}\Sigma y_i + n\bar{x}\bar{y}}{\Sigma X_i^2 - 2\bar{x}\Sigma X_i + n\bar{x}^2} = \frac{\Sigma(X_i - \bar{x})(y_i - \bar{y})}{\Sigma(X_i - \bar{x})^2}$$

$$b_1 = \frac{\Sigma X_i^2 \Sigma y_i - \Sigma X_i \Sigma X_i y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} = \frac{\Sigma X_i^2 \Sigma y_i - n^2\bar{x}^2\bar{y} - \Sigma X_i \Sigma X_i y_i + n^2\bar{x}^2\bar{y}}{n\Sigma(X_i - \bar{x})^2}$$

$$= \frac{\Sigma X_i^2 \bar{y} - \Sigma X_i \bar{x}\bar{y} - \bar{x}\Sigma X_i y_i + \bar{x}^2 \Sigma y_i}{\Sigma(X_i - \bar{x})^2}$$

$$= \frac{\Sigma X_i^2 \bar{y} - 2\Sigma X_i \bar{x}\bar{y} + n\bar{y}\bar{x}^2 - \bar{x}\Sigma X_i y_i + \Sigma X_i \bar{x}\bar{y} + \bar{x}^2\Sigma y_i - n\bar{x}^2\bar{y}}{\Sigma(X_i - \bar{x})^2}$$

$$= \frac{\bar{y}\Sigma(X_i^2 - 2X_i\bar{x} + \bar{x}^2) - \bar{x}\Sigma(X_i y_i - X_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{\Sigma(X_i - \bar{x})^2}$$

$$= \frac{\bar{y}\Sigma(X_i - \bar{x})^2 - \bar{x}\Sigma(X_i - \bar{x})(y_i - \bar{x})}{\Sigma(X_i - \bar{x})^2} = \bar{y} - b_2\bar{x}$$

2.

$$\operatorname{var}(b_1|\mathbf{x}) = \sigma^2 \left[\frac{\sum x_i^2}{N\sum(x_i - \bar{x})^2} \right] \quad (2.14)$$

$$\operatorname{var}(b_2|\mathbf{x}) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \quad (2.15)$$

$$\operatorname{cov}(b_1, b_2|\mathbf{x}) = \sigma^2 \left[\frac{-\bar{x}}{\sum(x_i - \bar{x})^2} \right] \quad (2.16)$$

$$\operatorname{var}(b) = \sigma^2 (X'X)^{-1}$$

$$= \sigma^2 \frac{1}{n\Sigma X_i^2 - (\Sigma X_i)^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$$

$$= \frac{\sigma^2}{n\Sigma(X_i - \bar{X})^2} \begin{bmatrix} \Sigma X_i^2 & -\Sigma X_i \\ -\Sigma X_i & n \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma^2 \Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2} & \frac{\sigma^2(-\bar{X})}{\Sigma(X_i - \bar{X})^2} \\ \frac{\sigma^2(-\bar{X})}{\Sigma(X_i - \bar{X})^2} & \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

Therefore, $\operatorname{var}(b_1|X) = \frac{\sigma^2 \Sigma X_i^2}{n\Sigma(X_i - \bar{X})^2}$

$$\operatorname{var}(b_2|X) = \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2}$$

$$\operatorname{cov}(b_1, b_2|X) = \frac{\sigma^2(-\bar{X})}{\Sigma(X_i - \bar{X})^2}$$

3.

TABLE 5.6 Output for Exercise 5.3

Dependent Variable: *WALC*				
Included observations: 1200				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.4515	2.2019		0.5099
ln(*TOTEXP*)	2.7648		5.7103	0.0000
NK		0.3695	−3.9376	0.0001
AGE	−0.1503	0.0235	−6.4019	0.0000
R-squared		Mean dependent var		6.19434
S.E. of regression		S.D. dependent var		6.39547
Sum squared resid	46221.62			

a. Fill in the following blank spaces that appear in this table.
  i. The t-statistic for b_1.
  ii. The standard error for b_2.
  iii. The estimate b_3.
  iv. R^2.
  v. $\hat{\sigma}$.

b. Interpret each of the estimates b_2, b_3, and b_4.

c. Compute a 95% interval estimate for β_4. What does this interval tell you?

d. Are each of the coefficient estimates significant at a 5% level? Why?

e. Test the hypothesis that the addition of an extra child decreases the mean budget share of alcohol by 2 percentage points against the alternative that the decrease is not equal to 2 percentage points. Use a 5% significance level.

a.

(i) $t_{stat} = \frac{b_1}{se(b_1)} = \frac{1.4515}{2.2019} = 0.6592$

(ii) $se(b_2) = \frac{t_{stat}}{b_2} = \frac{2.7648}{5.7103} = 0.4842$

(iii) $b_3 = t_{stat} \cdot se(b_3) = 0.3695 \times (-3.9376) = -1.4549$

(iv) Since $S_Y = \sqrt{\frac{1}{n-1}\Sigma(y_i - \bar{y})^2}$, $SST = (n-1)S_Y^2 = (1200-1) \times 6.39547 = 49041.5418$

$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{46221.62}{49041.5418} \approx 0.0575$

(V) $\hat{\sigma} = \sqrt{\frac{SSE}{n-k}} = \sqrt{\frac{46221.62}{1200-4}} \approx 6.2170$

b.

→ $b_2 = 2.7648$
A household's budget spent on alcohol will increase 2.7649 units when total expenditure is increased by 1 unit, and other factors are held constant.

→ $b_3 = -1.45494$
A household's budget spent on alcohol will decrease 1.45494 units when the number of children in the household is increased by 1 unit, and other factors are held constant.

→ $b_4 = -0.1503$
A household's budget spent on alcohol will decrease 0.1503 units when the age of the household head is increased by 1 unit, and other factors are held constant.

c. 95% CI for β_4 :

$b_4 \pm t_{0.975,1196}\, se(b_4) = [-0.1503 - 1.96 \times 0.0235,\ -0.1503 + 1.96 \times 0.0235] = [-0.1964,\ -0.1042]$

→ A household's budget spent on alcohol will decrease between 0.1042 and 0.1964 units when the age of the household head is increased by 1 unit.

d. Except for the intercept (p = 0.5099 > 0.5),
all the other coefficient estimates significant at a 5% level. the p-value are all less than 0.5

e. $H_0: \beta_3 = -2$, $H_1: \beta_3 \neq -2$

$t = \dfrac{b_1 - \hat{\beta}_3}{se(b_3)} = \dfrac{-1.4515 + 2}{0.3695} = 1.475 < 1.96 = t_{0.975,1196}$

→ We fail to reject H_0. There's no sufficient evidence to state the decrease is different from 2%.

4. **5.23** The file *cocaine* contains 56 observations on variables related to sales of cocaine powder in northeastern California over the period 1984–1991. The data are a subset of those used in the study Caulkins, J. P. and R. Padman (1993), "Quantity Discounts and Quality Premia for Illicit Drugs," *Journal of the American Statistical Association*, 88, 748–757. The variables are

PRICE = price per gram in dollars for a cocaine sale
QUANT = number of grams of cocaine in a given sale
QUAL = quality of the cocaine expressed as percentage purity
TREND = a time variable with 1984 = 1 up to 1991 = 8
Consider the regression model

$$PRICE = \beta_1 + \beta_2 QUANT + \beta_3 QUAL + \beta_4 TREND + e$$

a. What signs would you expect on the coefficients β_2, β_3, and β_4?
b. Use your computer software to estimate the equation. Report the results and interpret the coefficient estimates. Have the signs turned out as you expected?
c. What proportion of variation in cocaine price is explained jointly by variation in quantity, quality, and time?
d. It is claimed that the greater the number of sales, the higher the risk of getting caught. Thus, sellers are willing to accept a lower price if they can make sales in larger quantities. Set up H_0 and H_1 that would be appropriate to test this hypothesis. Carry out the hypothesis test.
e. Test the hypothesis that the quality of cocaine has no influence on expected price against the alternative that a premium is paid for better-quality cocaine.
f. What is the average annual change in the cocaine price? Can you suggest why price might be changing in this direction?

a. β_2 → negative, larger quantity often implies bulk discount
β_3 → positive, higher quality with higher price
β_4 → uncertain, increasing supply with fixed demand leads fall in price and vice versa.

b.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 90.84669    8.58025  10.588 1.39e-14 ***
quant       -0.05997    0.01018  -5.892 2.85e-07 ***
qual         0.11621    0.20326   0.572   0.5700    
trend       -2.35458    1.38612  -1.699   0.0954 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 20.06 on 52 degrees of freedom
Multiple R-squared:  0.5097,	Adjusted R-squared:  0.4814 
F-statistic: 18.02 on 3 and 52 DF,  p-value: 3.806e-08
```

$\widehat{PRICE} = 90.84669 - 0.05997\,QUANT + 0.11621\,QUAL - 2.35458\,TREND + e$

All the signs are same as expected.

→ The price will decrease by 0.05997 unit when quantity increase 1 unit with other factors held constant.
The price will increase by 0.11621 unit when quality increase 1 unit with other factors held constant.
The price will decrease by 2.35458 unit when time increase 1 unit with other factors held constant.

c. $R^2 = 0.5097$

d. $H_0: \beta_2 \geq 0$, $H_1: \beta_2 < 0$

$t = -5.892 < t_{0.05,54} = -1.6736$

$\longrightarrow$ We reject H_0. There's sufficient evidence to state that sellers are willing to accept lower price with larger quantity.

e. $H_0: \beta_3 \leq 0$, $H_1: \beta_3 > 0$

$t = 0.572 < t_{0.05,54} = 1.6736$

$\longrightarrow$ We fail to reject H_0. There's no sufficient evidence to state that quality affects price.

f. average $= -2.3548$

It may be due to the technologial improvement or market saturation.