The Ordinary Least Squares (OLS) Estimators

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

$$b_1 = \bar{y} - b_2\bar{x} \tag{2.8}$$

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.

1.Let K=2, show that (b1,b2) in p.29 of slides in Ch5 reduces to the formula of (b1,b2) in (2.7)-(2.8)

當 $K=2$，令 $Y_i = \beta_1 + \beta_2 X_i + e_i$, $i = 1, 2, \ldots, N$

令 $X = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \Rightarrow X^{-1}X = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \Sigma X_i \\ \Sigma X_i & \Sigma X_i^2 \end{bmatrix}$

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

$X^{-1}Y = \begin{bmatrix} 1 & \cdots & 1 \\ X_1 & X_2 \cdots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}$

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

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

而 $b_2 = \frac{-\Sigma X_i \Sigma Y_i + n\Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} = \frac{-n^2\bar{X}\bar{Y} + n\Sigma X_i Y_i}{n\Sigma X_i^2 - (n\bar{X})^2} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2}$

$b_1 = \frac{\Sigma X_i^2 \Sigma Y_i - \Sigma X_i \Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} = \frac{n\Sigma X_i^2 \bar{Y} - n\bar{X}\Sigma X_i Y_i}{n\Sigma X_i^2 - (n\bar{X})^2} = \frac{\bar{Y}[\Sigma X_i^2 - n\bar{X}^2] - \bar{X}[\Sigma X_i Y_i - n\bar{X}\bar{Y}]}{\Sigma(X_i - \bar{X})^2} = \bar{Y} - b_2\bar{X}$

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

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

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

2. **Let K=2, show that cov(b1,b2) in p.30 of slides in Ch5 reduces to the formula of (b1,b2) in (2.14)-(2.16)**

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

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

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

$$= \begin{bmatrix} Var(b_1|X) & Cov(b_1, b_2|X) \\ Cov(b_1, b_2|X) & Var(b_2|X) \end{bmatrix}$$

5.3 Consider the following model that relates the percentage of a household's budget spent on alcohol *WALC* to total expenditure *TOTEXP*, age of the household head *AGE*, and the number of children in the household *NK*.

$$WALC = \beta_1 + \beta_2 \ln(TOTEXP) + \beta_3 NK + \beta_4 AGE + e$$

This model was estimated using 1200 observations from London. An incomplete version of this output is provided in Table 5.6.

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.

ai. b_1的 t 值: $\frac{\hat{b}_1}{SE(b_1)} = \frac{1.4515}{2.2019} \doteqdot 0.6594$

aii $SE(b_2)$: $\frac{\hat{b}_2}{t} = \frac{2.7648}{5.7103} \doteqdot 0.4841$

aiii. $b_3 = t \cdot SE = -3.9376 \cdot 0.3695 \doteqdot -1.4549$

aiv. $R^2 = 1 - \frac{SSE}{SSR} \doteqdot 1 - \left(\frac{4.6221.62}{(1200-1) \times (6.39547)^2}\right) \doteqdot 0.0575$

av. $\hat{\sigma} \doteqdot \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{4.6221.62}{1200-4}} \doteqdot 0.6242$

b. b_2: 表示TOTEXP対数增加1%，酒精支出佔比增約 2.76%
b_3: 表示每增一孩子，酒精支出佔比平均減約 1.45%
b_4: 表示家庭主每增1歲，酒精支出佔比平均減約 0.15%

c. $CI_{95\%} = \hat{b}_4 \pm t_{0.025, n-k-1} = -0.1503 \pm 1.96 \times 0.0235 = [-0.1964, -0.1042]$

d. $\left\{\begin{array}{l} b_2\ (0) < 0.05 \\ b_3\ (0.001) < 0.05 \\ b_4\ (0) < 0.05 \end{array}\right)$ 全显著

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

$t = \frac{\hat{b}_3 - (-2)}{SE(\hat{b}_3)} = \frac{-1.4549 + 2}{0.3695} \approx 1.4752$

而 $t^* = t_{0.025, 1196} \approx 1.96$

因 $1.4752 < 1.96$. no reject H_0

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?

a. β_2 預期負，量大压低購買價
β_3 預期正，純度高成本高.
β_4 不確定，無法推論

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?

```
lm(formula = price ~ quant + qual + trend, data = cocaine)

Residuals:
    Min      1Q  Median      3Q     Max
-43.479 -12.014  -3.743  13.969  43.753

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
```

b. $\beta_2 = -0.05997$ 符合預期
$\beta_3 = 0.11621$ 符合預期
$\beta_4 = -2.35458$ 不符合預期

c. Multiple $R^2 = 0.5097$
表約 51% 可被解釋

```
"t-value for H0: beta2 = 0 is -5.892"
"Critical t-value at 5% significance level: -1.675"
```

d $H_0: \beta_2 \geq 0$ / $H_1: \beta_2 < 0$

$t = -5.892 < t_{0.95,52} = -1.675$

Reject H_0

e.
```
t-value for testing beta3 > 0: 0.5717
Critical t-value (5% level): 1.6462
```

$H_0: \beta_3 = 0$, $H_1: \beta_3 > 0$

$t = 0.5717 < t_{0.95,52} = 1.675$, not reject H_0

f. 平均年變化: $\beta_4 = -2.35458$
表示平均價格每年下降約 2.35 美元/每克
可能是因生產技術↑, S 過多 P 下降