Q1

$$Y_i = b_1 + b_2 x_i + e_i \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

matrix form : $Y = X\beta + e$

$$\beta = (X'X)^{-1}(X'Y)$$

$$(X'X) = \begin{bmatrix} 1, 1 \cdots 1 \\ x_1, x_2, \cdots x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n (x_i)^2 \end{bmatrix}, \quad X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1, & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$(X'X)^{-1} \cdot (X'Y) = \frac{1}{n\sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \cdot \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$= \frac{1}{n\sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i \\ -\sum_{i=1}^n x_i \sum_{i=1}^n y_i + n\sum_{i=1}^n x_i y_i \end{bmatrix}$$

$\sum_{i=1}^n x_i = n \cdot \bar{x}$

$$b_1 = \frac{\sum_{i=1}^n x_i^2 \cdot n\bar{y} - n\bar{x} \cdot \sum_{i=1}^n x_i y_i}{n\sum_{i=1}^n x_i^2 - (n\bar{x})^2} = \frac{\bar{y}\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \quad ①$$

$$b_2 = \frac{-n^2\bar{x}\bar{y} + n\sum_{i=1}^n x_i y_i}{n\sum_{i=1}^n x_i^2 - (n\bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \quad ②$$

The Ordinary Least Squares (OLS) Estimators

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

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

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.

$*\sum x_i \bar{y} = \bar{y}\sum x_i = \bar{y} \cdot n \cdot \bar{x}$

(2.7)式展開, $b_2 = \dfrac{\sum (x_i y_i - \overset{n\bar{x}\bar{y}}{x_i\bar{y}} - \overset{n\bar{x}\bar{y}}{\bar{x}y_i} + \overset{n\bar{x}\bar{y}}{\bar{x}\bar{y}})}{\sum x_i^2 - \underset{2n\bar{x}^2}{2\sum x_i\bar{x}} + \underset{n\bar{x}^2}{\sum \bar{x}^2}} = \dfrac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$ 與②相等

(2.8)展開

$$b_1 = \bar{y} - \left(\frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}\right) \cdot \bar{x} = \frac{\bar{y}\sum x_i^2 - \cancel{\bar{y} \cdot n\bar{x}^2} - \bar{x}\sum x_i y_i + \cancel{n\bar{x}^2\bar{y}}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\bar{y}\sum x_i^2 - \bar{x}\sum x_i y_i}{\sum x_i^2 - n\bar{x}^2}$$ 與①相等

Q2

$$Var(b) = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2}\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & -\sum_{i=1}^{n} x_i \\ -\sum_{i=1}^{n} x_i & n \end{bmatrix} = \begin{array}{cc} Var(b_1|x) & cov(b_1,b_2|x) \\ \multicolumn{2}{c}{\begin{bmatrix} \dfrac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} & \dfrac{-\sigma^2 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} \\ \dfrac{-\sigma^2 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} & \dfrac{\sigma^2 n}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} \end{bmatrix}} \\ Cov(b_2,b_1|x) & Var(b_2|x) \end{array}$$

$$Var(b_1|x) = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} = \sigma^2 \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n} x_i^2 - (n\bar{x})^2}$$

$$Var(b_2|x) = \frac{\sigma^2 n}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} = \frac{\sigma^2 n}{n\sum_{i=1}^{n} x_i^2 - (n\bar{x})^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

和下式

皆相同

$$Cov(b_1,b_2|x) = \frac{-\sigma^2 \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i^2} = \frac{-\sigma^2 \cdot n\bar{x}}{n\sum_{i=1}^{n} x_i^2 - n^2\bar{x}^2} = \frac{-\sigma^2 \bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

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

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

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

分母展開

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - 2\underbrace{\sum x_i \bar{x}}_{2n\bar{x}^2} + \underbrace{\sum \bar{x}^2}_{n\bar{x}^2} = \underline{\sum x_i^2 - n\bar{x}^2}$$

$$N\sum (x_i - \bar{x})^2 = N\left(\sum x_i^2 - 2\sum x_i \bar{x} + \sum \bar{x}^2\right)$$

$$= N\left(\sum x_i^2 - n\bar{x}^2\right)$$

$$= \underline{n\sum x_i^2 - (n\bar{x})^2}$$

(2.14)分母展開

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	0.6592	0.5099
ln(*TOTEXP*)	2.7648	0.4842	5.7103	0.0000
NK	−1.4549	0.3695	−3.9376	0.0001
AGE	−0.1503	0.0235	−6.4019	0.0000
R-squared	0.0575	Mean dependent var		6.19434
S.E. of regression	6.2161	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) $\frac{\text{Coeff.}(b_1)}{SE} = \frac{1.4515}{2.2019} = 0.6529$

(ii) $SE = \frac{b_2}{t} = \frac{2.7648}{5.7103} = 0.4842$

(iii) $b_3 = SE \cdot t = -3.9376 \times 0.3695 = -1.4549$

(iv) $R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{46221.62}{SST} = 1 - \frac{46221.62}{49041.54} = 0.0575$

$\hookrightarrow (N-1)s_y^2 = (1200-1) \times 6.39547^2 = 49041.54$

(v) $\hat{\sigma}^2 = \frac{SSR}{N-K} = \frac{46221.62}{1200-4} = 38.64$

$\hat{\sigma} = \sqrt{38.64} = 6.2161$

(b)

$b_2 = 2.7648$ (TOTEXP有加log)

其他條件不變之下，TOTEXP 每增加 1%，WALC平均上升2.7648%

$b_3 = -1.4549$

其他條件不變之下，家中每多一個孩子(NK)，WALC平均下降1.4549%

$b_4 = -0.1503$

其他條件不變之下，AGE每增一單位，WALC平均下降0.1503%

(C)

CI for b_4 :

$-0.1503 \pm 1.9619 \times 0.0235$

$t_{1196, 0.025} = 1.9619$

$\Rightarrow -0.1964 \le b_4 \le -0.1042$

在95%的信心水準下，真實的 b_4 會被涵蓋在 interval 中

(d) 除了 intercept 外，其他 beta 的 p-value 都小於 0.05，他們都顯著

(e)

$H_0 : b_3 = -2$, $H_1 : b_3 \neq -2$

$$\frac{-1.4549 - (-2)}{0.3695} = 1.4752 < t_{1196, 0.025} = 1.9619$$

not reject H_0，表示統計上沒有顯著證明支持額外NK
會對 WALC造成的effect
不是 −2%

Q23

(a)

b2 預期符號：負

販賣數量增加時，通常會有數量折扣，降低價格

b3 預期符號：正

純度較高的可卡因應該能賣得更高的價格

b4 預期符號：可能為負

市場供應增加或競爭變激烈導致價格下降

(b)

```
Call:
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
```

b2 數量每增加 1 克，價格會降低 0.05997 美元，符合數量折扣假設

b3 純度每提高 1%，價格上升 0.11621 美元，符合預期

b4 每年價格下降 2.35458 美元，可能反映市場供應增加

大致符合預期

(c)

R^2 = 0.5097，代表該三個變數可以解釋 50.97%的價格的變異

(d)

H0 : b2 = 0

H1 : b2 < 0

$t_{0.05, 52}$= -1.6745

critical value = -0.05997/0.01018 = 5.89 < -1.6745。拒絕 H0，接受 H1

(e)

H0 : b3 = 0

H1 : b3 > 0

$t_{0.05, 52}$= 1.6745

critical value = 0.11621/0.20326 = 0.57< 1.6745。不拒絕 H0，沒有充足證據顯示 qual 對 price 有正面影響力

(f)

價格每年平均下降 2.35458，在其他條件不變下。可能代表市場逐年的供給增加，讓價格下降