1. 令模型為 $y_i = b_1 + b_2 x_i + \varepsilon_i$，矩陣形式為：$Y = X\beta + e$

其中 $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$，$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$，$\beta = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$，

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

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

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

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

$$= \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \# \cdots ①$$

Note $\sum x_i = n \cdot \bar{x}$

$\sum y_i = n \cdot \bar{Y}$

$$b_2 = \frac{-\sum x_i \cdot \sum y_i + n\sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{\cancel{n}(\sum x_i \cdot y_i - n \cdot \bar{X}\bar{Y})}{\cancel{n}(\sum x_i^2 - n\bar{X}^2)} = \frac{\sum x_i y_i - n\bar{X}\bar{Y}}{\sum x_i^2 - n\bar{X}^2} \cdots \text{Equation 2.7} \ \#$$

$b_1 = \bar{y} - b_2\bar{x} \Rightarrow$ Equation 2.8，將 b_2 代入，改寫 b_1

$$b_1 = \frac{1}{\sum x_i^2 - n\bar{x}^2} \cdot \left[\bar{y} \cdot \sum x_i^2 - \cancel{n\bar{x}^2 \cdot \bar{Y}} - \bar{x} \cdot \sum x_i y_i + \cancel{n \cdot \bar{x}^2 \bar{Y}}\right] = \frac{\bar{y} \cdot \sum x_i^2 - \bar{x} \cdot \sum x_i y_i}{\sum x_i^2 - n\bar{x}^2} \ ②$$

從①提出 $b_1 = \dfrac{\sum x_i^2 \cdot \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2} = \dfrac{n[\bar{y} \cdot \sum x_i^2 - \bar{x} \cdot \sum x_i y_i]}{n(\sum x_i^2 - n\bar{x}^2)}$

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

2. 令所有參數設定如 1. 小題，則 $(X'X)^{-1} = \dfrac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$

$$Var(b) = \sigma^2 \cdot \frac{1}{n \cdot \sum x_i^2 - (\sum x_i)^2} \cdot \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} = \begin{bmatrix} Var(b_1|X) & Cov(b_1,b_2|X) \\ Cov(b_1,b_2|X) & Var(b_2|X) \end{bmatrix}$$

Note $\sum(x_i - \bar{x})^2 = \sum(x_i^2 - 2x_i \cdot \bar{x} + \bar{x}^2) = \sum x_i^2 - n\bar{x}^2$

$$Var(b_1|X) = \sigma^2 \cdot \frac{\sum x_i^2}{n\sum x_i^2 - (\sum x_i)^2} = \sigma^2 \cdot \frac{\sum x_i^2}{n[\sum x_i^2 - n\bar{x}^2]} = \sigma^2 \cdot \frac{\sum x_i^2}{n[\sum(x_i - \bar{x})^2]}$$

…… Equation 2.14

$$Var(b_2|X) = \sigma^2 \cdot \frac{n}{n\sum x_i^2 - (\sum x_i)^2} = \sigma^2 \frac{n}{n[\sum x_i^2 - n\bar{x}^2]} = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

…… Equation 2.15

$$Cov(b_1,b_2|X) = \sigma^2 \frac{\sum x_i}{n\sum x_i - (\sum x_i)^2} = \sigma^2 \frac{n \cdot \bar{x}}{n \cdot [\sum x_i^2 - n\bar{x})^2]} = \sigma^2 \cdot \frac{\bar{x}}{\sum(x_i - \bar{x}^2)}$$

…… Equation 2.16

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. $\hat{\sigma}$.

a.

i. $t_{b_1} = \frac{1.4515}{2.2019} = 0.6592$, ii. $SE_{b_2} = \frac{2.7648}{5.7103} = 0.4842$

iii. $\hat{\beta}_3 = -3.9376 \times 0.3695 = -1.4552$, iv. $TSS = \sigma^2 \cdot (N-1) = 6.39547^2 \times 1199 = 48934.74$

$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{46221.62}{48934.74} = 0.0555$

V. $S.E. = \sqrt{\frac{46221.62}{1196}} = \sqrt{38.65} \doteqdot 6.22$

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

b. ln(TOTEXP)每上升1，WALC增加2.7648%

家庭的孩子數增加1，WALC降低 −1.4552%

一家之主的年齡上升1，WALC降低 −0.1503%

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

$t_{0.025,1196} \approx 1.96 \Rightarrow -0.1503 \pm 1.96 \times 0.0235 = -0.1503 \pm 0.0461 = [-0.1964, -0.1042]$

每當一家之主年齡上升1，在95%的信心水準之下，WALC會下降(因區間皆<0)

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

否，常數在95%信心水準下可能與0無異，因其p-value = 0.5099 > 0.05

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.

$H_0: \beta_3 = -2,\ H_a: \beta_3 \neq -2$

$$t = \frac{-1.4552-(-2)}{0.3695} = 1.474,\quad t_{0.025,1196} \approx 1.96,\quad |t| \leq 1.96 \Rightarrow \text{Do not Reject } H_0$$

在95%信心水準下無充足證據說明 $\beta_3 \neq -2$，

Q23

(a)

beta2、beta3為正，因為價格應該會根據品質(QUAT)以及需求(QUANT)上漲

beta4可能為負因人們開始逐漸意識到毒品的危險導致購買需求減少，價格也降低或市政府以及警方開始加強取締的力度。

(b)QUANT之係數為負，可能是因為交易量越多被抓的風險越高，使價格降低。

```
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 .
```

(c)約有51%價格的變化可以被此回歸解釋

```
> summary_model <- summary(model)
> r_squared <- summary_model$r.squared
> r_squared
[1] 0.50965
```

(d)H0:銷售量與價格為正相關，Ha:銷售量與價格為負相關

```
[1] -0.05996979
> quant_p_value
[1] 2.85072e-07
>
> if (quant_p_value / 2 < 0.05 && quant_coef < 0) {
+   cat("拒絕 H0: 存在負相關關係，即銷售量越大，價格越低。\n")
+ } else {
+   cat("未能拒絕 H0: 沒有足夠的證據支持銷售量與價格之間存在負相關關係。\n")
+ }
拒絕 H0: 存在負相關關係，即銷售量越大，價格越低。
```

(e)H0:品質與價格無關;Ha:品質與價格有相關性

```
> if (qual_p_value < 0.05 && qual_coef > 0) {
+   cat("拒絕 H0: 可卡因品質對價格有顯著影響，且品質越好價格越高。\n")
+ } else {
+   cat("未能拒絕 H0: 沒有足夠證據表明品質對價格有顯著影響。\n")
+ }
未能拒絕 H0: 沒有足夠證據表明品質對價格有顯著影響。
```

(f)每年價格變化為-2.3546，可能是因為政府的鄭彩改變，增加毒品的管制或者是民眾意識到毒品所帶來的嚴重損害。造成毒品交易較低、價格降低。

```
> trend_coef <- summary(model)$coefficients["trend", "Estimate"]
>
> trend_coef
[1] -2.354579
```