Q 1

Let $k=2$, then $Y_i = \beta_1 + \beta_2 X_i + e_i$

$$\Rightarrow \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

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

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix}_{2\times n} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} \Sigma Y_i \\ \Sigma X_i Y_i \end{bmatrix}_{2\times 1}$$

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

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

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

Set $\bar{X} = \frac{1}{n}\Sigma X_i$, $\bar{Y} = \frac{1}{n}\Sigma Y_i$

then
$$b_2 = \frac{-\Sigma X_i \Sigma Y_i + n\Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2} = \frac{\Sigma X_i Y_i - n(\frac{1}{n}\Sigma X_i)\cdot(\frac{1}{n}\Sigma Y_i)}{\Sigma X_i^2 - n\cdot(\frac{1}{n}\Sigma X_i)^2}$$

$$= \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2} = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} \quad ※$$

and $b_1 = \bar{Y} - b_2 \bar{X}$

※

Q 2

Let $k=2$ by the result of Q1

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

$$\Rightarrow \begin{cases} Var(b_1|x) = \sigma^2 \left[\dfrac{\sum x_i^2}{n\cdot\sum(x_i-\bar{x})^2}\right] \\ Var(b_2|x) = \sigma^2 \dfrac{1}{\sum(x_i-\bar{x})^2} = \dfrac{\sigma^2}{\sum(x_i-\bar{x})^2} \\ Cov(b_1,b_2|x) = \sigma^2 \cdot \left[\dfrac{-\bar{x}}{\sum(x_i-\bar{x})^2}\right] \end{cases}$$

#

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. $\hat{\sigma}$.

i.

$$t = \frac{b_1}{se(b_1)} = \frac{1.4515}{2.2019} \approx 0.6592$$

ii.

$$SE = \frac{b_2}{t} = \frac{2.7648}{5.7103} \approx 0.4842$$

iii.

$$b_3 = t \times SE = -3.9376 \times 0.3695 \approx -1.4549$$

iv.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{46221.62}{(6.39547)^2 \times (1200-1)} \approx 0.0575$$

v.

$$\sqrt{\frac{SSE}{n-4}} = \sqrt{\frac{46221.62}{1196}} \approx 6.2166$$

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

$b_2 = 2.7648$, if the household's total spending goes up by 1%, the alcohol budget share goes up by about 2.76 percentage points. *

$b_3 = 0.3695$, Each extra child in the family increases the alcohol budget share by about 0.37 percentage points. *

$b_4 = -0.1503$, if the head of the household gets one year older, the alcohol budget share goes down by about 0.15 percentage points. *

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

$b_4 \pm 1.96 \times SE = -0.1503 \pm 1.96 \times 0.0235 = -0.1503 \pm 0.0461$

$\Rightarrow [-0.1964, -0.1042]$ *

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

At the 5% significance level, all coefficient estimates except for the intercept are statistically significant. This is because their p-values are all less than 0.05, indicating strong evidence against the null hyphothesis that the coefficients are zero. The intercept, with a p-value of 0.5099, is not statistically significant at the 5% level. *

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.

$H_0: \beta_3 = -2, \quad H_1: \beta_3 \neq -2, \quad \alpha = 0.05$

t-statistic : $\dfrac{\hat{\beta}_3 - (-2)}{SE} = \dfrac{-1.4542+2}{0.3695} \approx 1.477$

$\because \ |t| = 1.477 < 1.96 \approx t_{0.025,1196}$

$\Rightarrow$ not reject H_0 that the effect of an additional child is a 2 percentage point decrease in the alcohol budget share.

#

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?

```
> # 建立多元線性迴歸模型
> model <- lm(price ~ quant + qual + trend, data = cocaine)
>
> # 查看結果摘要
> summary(model)

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
```

$$\Rightarrow \begin{cases} \beta_2 = -0.05997 < 0 \\ \beta_3 = 0.11621 > 0 \\ \beta_4 = -2.35458 < 0 \end{cases}$$

#

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?

The signs of the coefficients are consistent with expectations, though not all are statistically significant.

#

c. What proportion of variation in cocaine price is explained jointly by variation in quantity, quality, and time?

The proportion of variation in cocaine price that is jointly expained by quantity, quality, and time is given by the R-squared-value

$\Rightarrow$ R-squared = 0.5097 by (a)

That means approximately 50.97% of the variation in cocaine prices is explained 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

$H_0 : \beta_2 = 0$ v.s. $H_1 : \beta_2 < 0$

```
> # 輸出檢定結果
> cat("t-value:", t_value, "\n")
t-value: -5.891936
> cat("One-tailed p-value:", p_value_one_tailed, "\n")
One-tailed p-value: 1.42536e-07
```

$\because$ 1.42536e-07 << 0.05

$\Rightarrow$ reject H_0 ※

$\Rightarrow$ There is strong and statistically significant evidence that larger quantities are associated with lower prices. ※

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.

$H_0 : \beta_3 = 0$ v.s. $H_1 : \beta_3 > 0$

```
> # 輸出結果
> cat("t-value:", t_qual, "\n")
t-value: 0.5716946
> cat("One-tailed p-value:", p_value_qual_one_tailed, "\n")
One-tailed p-value: 0.284996
```

$\because$ 0.284996 >> 0.05

$\Rightarrow$ not reject H_0 ※

$\Rightarrow$ There is no statistically evidence that higher-quality cocaine is sold at a higher price

f. What is the average annual change in the cocaine price? Can you suggest why price might be changing in this direction?

From the regression output at (a) $\Rightarrow$ the coefficient on the trend variable is -2.355 ※

There is evidence that the price is declining over time, possibly due to increasing supply, improved distribution efficiency, or shifting market dynamics ※