Q1:

$$X=\begin{bmatrix}1 & X_1\\ \vdots & \vdots\\ n & X_n\end{bmatrix},\quad Y=\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

$k=2$

$$X'X=\begin{bmatrix}1 & 1\\ X_1 & X_2\end{bmatrix}\cdot\begin{bmatrix}1 & X_1\\ 1 & X_2\end{bmatrix}=\begin{bmatrix}n & \Sigma X_i\\ \Sigma X_i & \Sigma X_i^2\end{bmatrix}$$

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

from p29

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

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

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

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

$$\Rightarrow \hat{\beta}_2^{OLS}=\frac{n\Sigma X_i y_i-\Sigma X_i y_i}{n\Sigma X_i^2-(\Sigma X_i)^2}=\frac{\Sigma X_i y_i-\frac{\Sigma X_i y_i}{n}}{\Sigma X_i^2-\frac{(\Sigma X_i)^2}{n}}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(y_i-\bar{y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$\hat{\beta}_1^{OLS}=\frac{\Sigma X_i^2 y_i-\Sigma X_i\cdot\Sigma X_i y_i}{n\Sigma X_i^2-(\Sigma X_i)^2}=\frac{\bar{y}\cdot n\Sigma X_i^2-n\bar{X}\cdot\Sigma X_i y_i}{n\Sigma X_i^2-n\cdot\bar{X}^2}$$

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

$$=\frac{\bar{y}(\Sigma X_i-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}-\bar{X}\cdot\frac{\Sigma(X_i-\bar{X})\Sigma(y_i-\bar{y})}{\Sigma(X_i-\bar{X})^2}=\bar{y}-\bar{X}\cdot\hat{\beta}_2$$

Q2.

$K=2$

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

$$Var(\hat{\beta}_1 | X) = \frac{\sigma^2 \cdot \Sigma X_i^2}{n \cdot \Sigma X_i^2 - (\Sigma X_i)^2} = \frac{\sigma^2 \cdot \Sigma X_i^2}{n\left(\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}\right)} = \frac{\sigma^2 \cdot \Sigma X_i^2}{n \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$Var(\hat{\beta}_2 | X) = \frac{\sigma^2 \cdot n}{n \Sigma X_i^2 - (\Sigma X_i)^2} = \frac{\sigma^2 \cdot n}{n \cdot \Sigma (X_i - \bar{X})^2} = \frac{\sigma^2}{\Sigma (X_i - \bar{X})^2}$$

$$Cov(\hat{\beta}_1, \hat{\beta}_2 | X) = \frac{\sigma^2 \cdot \Sigma X_i}{n \cdot \Sigma X_i^2 \cdot (\Sigma X_i)^2}$$

$$= \frac{-\sigma^2 \cdot n \cdot \bar{X}}{n \cdot \Sigma (X_i - \bar{X})^2} = \frac{-\sigma^2 \cdot \bar{X}}{\Sigma (X_i - \bar{X})^2} = \frac{-\bar{X}}{\Sigma (X_i - \bar{X})^2} \cdot \sigma^2$$

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		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.

<sol> a. (i) $t_1 = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \dfrac{1.4515}{2.2019} = 0.6592$

(ii) $5.7103 = t_2 = \dfrac{\hat{\beta}_2}{SE(\hat{\beta}_2)} = \dfrac{2.7648}{SE(\hat{\beta}_2)}$ ∴ $SE(\hat{\beta}_2) = 0.4842$

(iii) $-3.9376 = t_3 = \dfrac{\hat{\beta}_3}{SE(\hat{\beta}_3)} = \dfrac{\hat{\beta}_3}{0.3695}$ ∴ $\hat{\beta}_3 = -1.4549$

(iv) $R^2 = 1 - \dfrac{SSE}{S_{YY}} = 1 - \dfrac{SSE}{(n-1)S_Y^2} = 1 - \dfrac{46221.62}{(1200-1)6.39547^2} = 0.0575$

(v.) $\hat{\sigma}^2 = \dfrac{SSE}{n-K} = \dfrac{46221.62}{1200-4} = 38.6468 \Rightarrow \hat{\sigma} = 6.2167$

b. b2 (2.7648): This coefficient indicates that a 1% increase in total expenditure (TOTEXP) leads to a 2.7648% increase in the percentage of the budget spent on alcohol (WALC).

b3 (-1.4549): The negative coefficient suggests that an increase in the number of children in the household (NK) is associated with a decrease in the percentage of the budget spent on alcohol (WALC).

b4 (-0.1503): The negative coefficient indicates that as the age of the household head (AGE) increases, the percentage of the budget spent on alcohol (WALC) decreases.

c、β_4 的 95% C.I. 為 $\hat{\beta}_4 \pm t_{0.025}(1200-4)\,SE(\hat{\beta}_4) = -0.1503 \pm 1.96 \cdot 0.0235$

$\Rightarrow [-0.19636, -1.0424]$

The 95% confidence interval for b4 is [-0.19636, -1.0424]. This means we are 95% confident that the true value of b4, the effect of the age of the household head on the percentage of the budget spent on alcohol, lies within this interval.

d、The t-values for each estimate are significant because their p-values are all less than the 5% significance level (for example, the p-value for b2 is 0.0000, indicating the variable is significant). Therefore, we can reject the null hypothesis and conclude that these estimates are significant at the 5% level.

e、$\begin{cases} H_0: \beta_3 = -2 \\ H_a: \beta_3 \neq -2 \end{cases}$

$\alpha = 0.05$

$t_3 = \dfrac{\hat{\beta}_3 - \beta_3}{SE(\hat{\beta}_3)} \overset{H_0}{\sim} t(n-K)$

$RR = \{|t_3| \geq t_{0.025}(1200-4)\} = \{|t_3| \geq 1.96\}$

$t_3 = \dfrac{-1.4549-(-2)}{0.3695} = 1.4752 \notin RR$, do not reject H_0、

統計上，β_3 並未顯著異於 2，表 there is no evidence to suggest that having an extra child leads to a decline in the alcohol budget share that is different from 2%、

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?

(sol) a.

b2: The expected sign is positive because as the quantity of cocaine sold (QUANT) increases, the price (PRICE) should rise.

b3: The expected sign is positive because as the quality of cocaine (QUAL) increases, the price (PRICE) will also increase.

b4: The expected sign is positive because, over time, the price (PRICE) is likely to rise, reflecting inflation or changes in market demand.

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
(Intercept) 90.84669    8.58025  10.588
quant       -0.05997    0.01018  -5.892
qual         0.11621    0.20326   0.572
trend       -2.35458    1.38612  -1.699
            Pr(>|t|)    
(Intercept) 1.39e-14 ***
quant       2.85e-07 ***
qual          0.5700    
trend         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: The coefficient for (QUANT) is -0.05997, indicating that as the quantity sold increases, the price decreases, and this variable is significant (p-value close to 0).

b3: The coefficient for (QUAL) is 0.11621, suggesting that as the quality of cocaine increases, the price rises, but this variable is not significant (p-value is 0.5700).

b4: The coefficient for (TREND) is -2.35458, showing that over time, the price decreases, but the significance of this variable is lower (p-value is 0.0954).

c、 According to the regression results, the multiple R^2 of 0.5097 indicates that approximately 50.97% of the variation in cocaine price is jointly explained by the quantity, quality, and time variables.

d. $\begin{cases} H_0: \beta_2 \geq 0 \\ H_a: \beta_2 < 0 \end{cases}$

$\alpha = 0.05$

test statistic: $T = \dfrac{\hat{\beta}_2}{SE(\hat{\beta}_2)} \overset{H_0}{\sim} t(n-k)$

$RR = \{ T < t_{0.05}(56-4) = -1.675 \}$

$T_0 = \dfrac{-0.05997}{0.01018} = -5.892 \in RR$, reject H_0

統計上，β_2 顯著異於 0，表 sellers are willing to accept a lower price if they can make sales in larger quantity.

e、 $\begin{cases} H_0: \beta_3 \leq 0 \\ H_a: \beta_3 > 0 \end{cases}$

$\alpha = 0.05$

test statistic: $T = \dfrac{\hat{\beta}_3}{SE(\hat{\beta}_3)} \overset{H_0}{\sim} t(n-k)$

$RR = \{ T > t_{0.05}(56-4) \} = \{ T > 1.675 \}$

$T_0 = \dfrac{0.11621}{0.20326} = 0.5717 \notin RR$, do not reject H_0.

統計上，β_3 顯著 ≤ 0，we can't conclude that a premium is paid for better quality cocaine.

f、 According to the regression results, the coefficient for the TREND variable is -2.35458, indicating that the price of cocaine decreases by approximately 2.35 USD per year. Therefore, the average annual price change is a decrease of 2.35 USD. This may reflect market competition or other economic factors.