1. Since $k=2 \Rightarrow X = \begin{bmatrix} 1 & x_1 \\ 1 & \vdots \\ \vdots & \\ 1 & x_n \end{bmatrix}$, $y_1 = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$$\Rightarrow X' = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \text{ and } X'X = \begin{bmatrix} n & n\cdot\bar{x} \\ n\cdot\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

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

$$\Rightarrow (X'X)^{-1}X' = C\begin{bmatrix} \sum_{i=1}^n x_i^2 - n\bar{x}x_1 & \cdots & \sum_{i=1}^n x_i^2 - n\bar{x}x_n \\ -n(\bar{x}-x_1) & \cdots & -n(\bar{x}-x_n) \end{bmatrix} \text{ where } C = \frac{1}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2}$$

$$\Rightarrow (X'X)^{-1}X'y = C\begin{bmatrix} (y_1+\cdots+y_n)\sum_{i=1}^n x_i^2 - n\bar{x}\left(\sum_{i=1}^n x_i y_i\right) \\ -(y_1+\cdots+y_n)n\bar{x} + n\sum_{i=1}^n x_i y_i \end{bmatrix}$$

$$= \frac{1}{n\sum(x_i-\bar{x})^2} \begin{bmatrix} n\cdot\bar{y}\sum_{i=1}^n x_i^2 - n\bar{x}\sum_{i=1}^n x_i y_i \\ -\bar{y}n^2\bar{x} + n\sum x_i y_i \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{n\left(\bar{y}\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i y_i\right)}{n\sum(x_i-\bar{x})^2} \\ \dfrac{-n\bar{y}\bar{x} + \sum_{i=1}^n x_i y_i}{\sum_{i=1}^n (x_i-\bar{x})^2} \end{bmatrix} = \begin{bmatrix} \dfrac{n\bar{y}\sum x_i^2 - \bar{y}n^2\bar{x}^2 + \bar{y}n^2\bar{x}^2 - n\bar{x}\sum x_i y_i}{n\sum(x_i-\bar{x})^2} \\ \dfrac{\sum x_i y_i - n\bar{x}\bar{y} + n\bar{x}\bar{y} - n\bar{x}\bar{y}}{\sum(x_i-\bar{x})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\bar{y}(n\sum x_i^2 - n^2\bar{x}^2) + n\bar{x}(\bar{y}n\bar{x} - \sum x_i y_i)}{n\sum(x_i-\bar{x})^2} = \bar{y} - \dfrac{n\bar{y}\bar{x} + n\sum x_i y_i}{\sum(x_i-\bar{x})^2}\bar{x} \\ \dfrac{\sum x_i y_i - \bar{y}\sum x_i + \bar{x}\bar{y} - \bar{x}\sum y_i}{\sum(x_i-\bar{x})^2} = \dfrac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \end{bmatrix}$$

$$= \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \text{ since } b_1 = \bar{y} - b_2\bar{x}, \quad b_2 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$$

2. since $var(b) = \sigma^2 (X'X)^{-1}$

$$= \sigma^2 C\begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\cdot\bar{x} & n \end{bmatrix}$$

$$= \sigma^2 \frac{1}{n\sum(x_i-\bar{x})^2} \begin{bmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \dfrac{\sum x_i^2}{n\sum(x_i-\bar{x})^2} & \dfrac{-\bar{x}}{\sum(x_i-\bar{x})^2} \\ \dfrac{-\bar{x}}{\sum(x_i-\bar{x})^2} & \dfrac{1}{\sum(x_i-\bar{x})^2} \end{bmatrix}$$

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

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

$var(b_2|X) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}$

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

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	0.6591907	0.5099
ln(*TOTEXP*)	2.7648	0.4842	5.7103	0.0000
NK	-1.455	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	6.22	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.** $\hat{\sigma}$.

$\frac{\text{coeff}}{\text{se}} = t\text{-statistic}$

$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{SSE}{1199 S_{yy}^2}$

$\hat{\sigma} = \sqrt{\frac{SSE}{1196}}$

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.

b.

$b_2 = 2.7648$, when $\ln(TOTEXP)$ increase 1 unit, the mean WALC will increase 2.7648.

$b_3 = -1.455$, when NK increase (resp. decrease) 1 unit, the mean WALC will decrease (resp. increase) 1.455.

$b_4 = -0.1503$, when AGE increase (resp. decrease) 1 unit, the mean WALC will decrease (resp. increase) 0.1503.

c. $t_{(0.95, 1196)} = 1.962$

$\Rightarrow$ the interval is $[-0.1503 - 1.962 \times 0.0235,\ -0.1503 + 1.962 \times 0.0235]$

$= [-0.1964,\ -0.1042]$

d. since $t_{(0.95, 1196)} = 1.962$, all the absolute value of test-statistic are greater. Thus, They are significant at 5% level.

e. $H_0: \beta_3 = -2$

$H_1: \beta_3 \neq -2$

the test-statistic $= \frac{-1.455 + 2}{0.3695} = 1.474$

Since $|1.474| < t_{(0.05, 1196)}$

$\Rightarrow$ we fail to reject H_0

picture for (b)、(c)

```
Coefficients:
(Intercept)       quant        qual       trend
   90.84669    -0.05997     0.11621    -2.35458

> summary_mod1<-summary(mod1)
> #c
> summary_mod1$r.squared
[1] 0.50965
```

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?

a. I expect that $\beta_2 < 0$, $\beta_3 > 0$, $\beta_4 < 0$.

b. since $\beta_2 = -0.06$, when QUANT increase 1 unit, the price will decrease 0.06.

Since $\beta_3 = 0.1162$, when QUAL increase 1 unit, the price will increase 0.1162.

Since $\beta_4 = -2.35$, when TREND increase 1 unit, the price will decrease 2.35.

All of the sign are same as my expectation.

c.

Approximately 50% variation in cocaine price is explain jointly.

d. $H_0: \beta_2 \geq 0$

$H_1: \beta_2 < 0$

$\Rightarrow$ test-statistic $t = \frac{-0.06}{se\beta_2} = \frac{-0.06}{0.01} = -6 < -t_{(0.95, 52)}$

$\Rightarrow$ We reject H_0.

e. $H_0: \beta_3 \leq 0$

$H_1: \beta_3 > 0$

$\Rightarrow$ test-statistic $t = \frac{0.1162}{0.2032} = 0.5718 < t_{(0.95, 52)}$ $\Rightarrow$ fail to reject H_0

f. β_4. As the time past, if we produce more cocaine than before (higher technology or more people are willing to produce), then the supplement exceed demand $\Rightarrow$ price lower.