$K=2$, $Y = X\beta + \varepsilon$, $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} 1 & X_{12} \\ 1 & X_{22} \\ \vdots & \vdots \\ 1 & X_{n2} \end{bmatrix}$, $\varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$

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

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

$$= \frac{1}{n\Sigma X^2 - (\Sigma X)^2} \begin{bmatrix} \Sigma X^2 & -\Sigma X \\ -\Sigma X & n \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{n2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \frac{1}{n\left(\Sigma X^2 - \frac{1}{n}(\Sigma X)^2\right)} \begin{bmatrix} \Sigma X^2 & -\Sigma X \\ -\Sigma X & n \end{bmatrix} \begin{bmatrix} \Sigma y \\ \Sigma Xy \end{bmatrix}$$

$$= \frac{1}{n\Sigma(X_i - \bar{X})^2} \begin{bmatrix} \Sigma y \, \Sigma X^2 - \Sigma X \, \Sigma Xy \\ -\Sigma X \Sigma y + n\Sigma Xy \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\bar{y}\,\Sigma X_i^2 - \bar{X}\,\Sigma Xy}{\Sigma(X_i - \bar{X})^2} \\ \dfrac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\bar{y}\left(\Sigma X_i^2 - n\bar{X}^2\right) - \bar{X}\,\Sigma Xy + n\bar{X}^2\bar{Y}}{\Sigma(X_i - \bar{X})^2} \\ \dfrac{\Sigma(X_i - \bar{X})(y_i - \bar{y})}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{y} - \dfrac{\Sigma Xy - n\bar{X}\bar{Y}}{\Sigma(X_i - \bar{X})^2}\bar{X} \\ \dfrac{\Sigma(X_i - \bar{X})(y_i - \bar{y})}{\Sigma(X_i - \bar{X})^2} \end{bmatrix}$$

$K=2$,

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

$$= \sigma^2 \begin{bmatrix} \dfrac{\Sigma X^2}{nS_{XX}} & \dfrac{-\Sigma X}{nS_{XX}} \\ \dfrac{-\Sigma X}{nS_{XX}} & \dfrac{1}{S_{XX}} \end{bmatrix}, \quad S_{XX} = \Sigma(X_i - \bar{X})^2$$

$$Var(b_1|X) = \sigma^2 \cdot \frac{\Sigma X^2}{nS_{XX}}$$

$$Var(b_2|X) = \sigma^2 \frac{1}{S_{XX}}$$

$$cov(b_1, b_2|X) = \sigma^2 \frac{-\Sigma X_i}{nS_{XX}}$$

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. $\hat{\sigma}$.

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)

test statistic : $\dfrac{b_1 - \beta_1}{SE(b_1)} \sim t(1196)$

$t^* = \dfrac{1.4515 - 0}{2.2019} = 0.6592$

(ii)

test statistic = $\dfrac{b_2 - \beta_2}{SE(b_2)} \sim t(1196)$

$t^* = \dfrac{2.7648 - 0}{SE(b_2)} = 5.7103$

$SE(b_2) = \dfrac{2.7648}{5.7103} = 0.4842$

(iii)

test statistic = $\dfrac{b_3 - \beta_3}{SE(b_3)} \sim t(1196)$

$t^* = \dfrac{b_3 - 0}{0.3695} = -3.9376$

$b_3 = -3.9376 \times 0.3695 = -1.4549$

(iv.)

$SST = \sum (Y_i - \bar{Y})^2$

$= 6.39547^2 \times 1199$

$=$

$R^2 = \dfrac{SSR}{SST} = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{46221.62}{49041.54179}$

$= 0.0575$

(V)

$\hat{\sigma} = \sqrt{MSE} = \sqrt{\dfrac{SSE}{1196}} = \sqrt{\dfrac{46221.62}{1196}} = 6.2167$

b.

The value $b_2 = 2.765$ suggest that a 1% increase in total expenditure will increase the share of expenditure going to alcohol by approximately 0.02765 percentage point.

The value $b_3 = -1.4549$ suggest that, if the household has one more child, the share of alcohol expediture of that household decrease by 1.45494 percentage points.

The value $b_4 = -0.1503$ suggest that, if the age of the household head increase by 1 year, the share of alcohol expediture decrease by 0.1503 percentage points.

C.

$b_4 \pm SE(b_4)\, t_{0.025}(1196)$

$-0.1503 \pm 0.0235 \times 1.96$

β_4 95% interval : $[-0.1964, -0.1042]$

This interval tell us that, if the age of the household head increases by 1 year, the share of alcohol expenditure is estimated to decrease by an amount between 0.1042 and 0.1964 percentage point.

d.

With the exception of the intercept, all coefficient estimates are significantly different from zero at 5% level because their p-value are all less than 0.05.

e.

$H_0 : \beta_3 = -2$

$H_1 : \beta_3 \neq -2$

test statistic : $\dfrac{b_3 - 2}{SE(b_3)} \sim t(1196)$

$t^* = \dfrac{-1.4549 - (-2)}{0.3695} = 1.495 < 1.96 = t_{0.975}(1196)$, ∴ reject H_0

There is no evidence to suggest that having an extra child leads to decline in the alcohol budget share that is different from two percentage points.

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.

The expected sign for β_2 is negative, because as the number of grams in a given sale increases, the price per gram should decrease, implying a discount for larger sales. We expect β_3 to be positive, the purer the cocaine, the higher the price. The sign of β_4 will depend on how demand and supply are changing over time. For example, a fixed demand and an increasing supply will lead to a fall in price. A fixed supply and increased demand would lead to a rise in price.

b.

```
Call:
lm(formula = price ~ ., 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
```

$$\widehat{PRICE} = 90.8467 - 0.06\,QUANT + 0.1162\,QUAL - 2.3546\,TREND$$

The estimated values for β_2, β_3 and β_4 are -0.06, 0.1162 and -2.3546 respectively. They imply that as quantity increase by 1 unit, the mean price will go down by 0.06. Also, as the quality increase by 1 unit the mean price goes up by 0.1162. As time increase by 1 year, the mean price decrease by 2.3546. All the signs turn out according to our expectations, with β_4 implying supply has been increasing faster than demand.

c.

R-squared = 0.5097

d.

$H_0: \beta_2 \geq 0$

$H_1: \beta_2 < 0$

$\alpha = 0.05$

test statistic : $\dfrac{b_2 - \beta_2}{SE(\beta_2)} \sim t(52)$

$$t^* = \frac{-0.05997 - 0}{0.01018} = -5.89 < -1.675 = t_{0.05}(52)$$

Reject H_0 and conclude that seller are will to accept a lower price if they can make sales in larger quantities.

e.

$H_0: \beta_3 \leq 0$

$H_1: \beta_3 > 0$

$\alpha = 0.05$

test statistic : $\dfrac{b_3 - 0}{SE(b_3)} \sim t_{0.95}(52)$

$$t^* = \frac{0.11621 - 0}{0.20326} = 0.572 < 1.645 = t_{0.95}(52)$$

$\therefore$ don't reject H_0

We can't say that a premium is paid for better-quality cocaine.

f.

The average annual change in the cocaine price is given by the value $b_4 = -2.3546$. It has a negative sign suggesting that the price decreases over time. A possible reason for a decreasing price is the development of improved technology for producing cocaine, such that suppliers can produce more at the same cost.