Q_1 : $K=2$, $Y = X\beta + e$

$Y = [y_1, y_2, \dots, y_n]^T$

$\beta = [\beta_1, \beta_2]^T$, $e = [e_1, e_2, \dots, e_n]^T$

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix}$$

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

$$(X^T X)^{-1} X^T Y = \frac{1}{n\Sigma x_i^2 - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

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

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

$$b_2 = \frac{n\Sigma x_i y_i - n^2 \bar{x}\bar{y}}{n\Sigma x_i^2 - n^2\bar{x}^2} = \frac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2} = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} \#$$

$$b_1 = \frac{\Sigma x_i^2 n\bar{y} - n\bar{x}(\Sigma x_i y_i)}{n\Sigma x_i^2 - (\Sigma x_i)^2} = \frac{[n\Sigma x_i^2 \bar{y} - (\Sigma x_i)^2 \bar{y}] + [(\Sigma x_i)^2 \bar{y} - n\bar{x}(\Sigma x_i y_i)]}{n\Sigma x_i^2 - (\Sigma x_i)^2}$$

$$= \bar{y} - \frac{\cancel{n^2}\bar{x}^2 \bar{y} - \cancel{n}\bar{x}(\Sigma x_i y_i)}{\cancel{n}\Sigma x_i^2 - \cancel{n^2}\bar{x}^2} = \bar{y} - b_2 \bar{x} \#$$

Q_2 : $Var(b)$

$$\sigma^2 (X^T X)^{-1} = \sigma^2 \frac{1}{n\Sigma x_i - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n \end{bmatrix} \Rightarrow$$ variance covariance matrix

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

$$Var(b_2|X) = \sigma^2 \frac{n}{n\Sigma x_i - \Sigma x_i^2} = \sigma^2 \frac{\cancel{n}}{\cancel{n}\Sigma x_i - n^{\cancel{2}}\bar{x}^2} = \sigma^2 \frac{1}{\Sigma x_i - n\bar{x}^2}$$

$$= \sigma^2 \cdot \frac{1}{\Sigma (x_i - \bar{x})^2}$$

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

$$= \sigma^2 \frac{-\bar{x}}{\Sigma (x_i - \bar{x})^2}$$

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		0.5099
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.

	Coefficient	Std. Error	t-Statistic	Prob.
C	1.4515	2.2019	0.6592	0.5099
ln(TOTEXP)	2.7648	0.4842	5.7103	0
NX	-1.4549	0.3695	-3.9376	0.0001
AGE	-0.1503	0.0235	-6.4019	0

R^2	0.0575	Mean dp. v.	6.19434
S.E. of regression	6.2167	S.D. dp v.	6.39547
Sum squared resid	46221.62		

$$1-\frac{46221.62}{6.39547^2\cdot(1200-1)} = 0.0575$$

$$\sqrt{\frac{46221.62}{(1200-4)}} = 6.2167$$

(a) $\frac{1.4515-0}{2.2019} = 0.6592$

$\frac{2.7648}{5.7103} = 0.4842$

$-3.9376\times 0.3695 = -1.4549$

(b) b_2：當 ln(TOTEXP) increase 1單位，其餘變數不變，WALC increase 2.7648 單位(%)

b_3：當 NX increase 1單位，其餘變數不變，WALC decrease 1.4549 單位(%)

b_4：當 AGE increase 1單位，其餘變數不變，WALC decrease 0.1503 單位(%)

(c) $b_4 \pm 0.0235\, t_{0.025,1196} = -0.1503 \pm 0.0235\cdot 1.96$

$= -0.1503 \pm 0.046$

$= (-0.1963, -0.1043)$

(d) yes，因為 p-value 都小於 0.05

顯著拒絕 $H_0: \beta_K = 0$

(e) $H_0: \beta_3 = -2$

$H_a: \beta_3 \neq -2$

$$\frac{-1.4549-(-2)}{0.3695} = 1.475$$

$1.475 < t_{0.025,1196} = 1.96$

不拒絕 H_0

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. 依據我的看法，我認為當賣的數量變多時，價格會下降，因為供給可能超過需求。
而品質上升時，價格應該上升，因為產品更具競爭力。
隨著年份的變化，需要視法規與人民使用需求，而沒辦法輕易只透過年份判斷。
因此，我認為 **β2 和 β3 分別是負值、正值**。而 **β4 則需進一步觀察**。

b.

```
> summary(lr)$coefficients
               Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 90.84668753 8.58025368 10.5878790 1.393119e-14
quant       -0.05996979 0.01017828 -5.8919359 2.850720e-07
qual         0.11620520 0.20326448  0.5716946 5.699920e-01
trend       -2.35457895 1.38612032 -1.6986829 9.535543e-02
```

b2 和 b3 與預期結果相似，而透過模型可以觀察 b4（年份）增加時，價格呈現下降的趨勢。

c.

```
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
```

可透過獨立變數解釋的比例為 R-squared，約 51%

d. 拒絕 H0, **β2 顯著小於 0**

```
#H0:Beta2>=0, Ha:Beta<0
estBeta2 = coef(sum_lr)[2, 1]
seBeta2 = coef(sum_lr)[2, 2]
test_statistic = (estBeta2-0)/seBeta2
df = df.residual(lr)
tc = qt(0.05, df, lower.tail = FALSE)
abs(test_statistic) > tc #TRUE: reject H0
```

Test statistic = -5.891936

Tc = 1.67

e. 不拒絕 H0, 無法去明 **β3 顯著大於 0**

```
#H0:Beta3=0, Ha:Beta3>=0
estBeta3 = coef(sum_lr)[3, 1]
seBeta3 = coef(sum_lr)[3, 2]
test_statistic = (estBeta3-0)/seBeta3
df = df.residual(lr)
tc = qt(0.025, df, lower.tail = FALSE)
abs(test_statistic) > tc #FALSE: non-reject H0
```

Test statistic = 0.5716946

Tc = 2

f. 收先依照 trend 數值將 data 依照年份的價格相加取得平均，代表當年度的 cocaine price，再計算每年成長的 percentage of price，並取得 price 年平均百分比幾何平均變化為 -5.3%，於於 data 缺乏 trend 為 6, 7 的資料，僅能計算前四年的變化計算幾何平均。

我推測在製造 cocaine 的品質上應該會越來越好，但由於受到製造成本下降以及生產速度提升，導致生產數量增加，價格下降。