Let $K=2$, show that (b1, b2) in p. 29 of slides in Ch 5 reduces to the formula of (b1, b2) in (2.7) - (2.8).

$$X = \begin{pmatrix} 1 & x_{1,2} \\ 1 & x_{2,2} \\ 1 & \vdots \\ 1 & x_{N,2} \end{pmatrix}$$

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

$$= \left[\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{1,2} & x_{2,2} & & x_{N,2} \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & \vdots \\ 1 & x_N \end{pmatrix} \right]^{-1} \left[\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & & x_N \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \right]$$

$$= \begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \end{bmatrix}$$

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

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

$$= \begin{bmatrix} \dfrac{N\bar{y}\Sigma x_i^2 - N\bar{x}\Sigma x_i y_i}{N\Sigma x_i^2 - (N\bar{x})^2} \\ \dfrac{N\Sigma x_i y_i - (N\bar{x})(N\bar{y})}{N\Sigma x_i^2 - (N\bar{x})^2} \end{bmatrix}$$

$$= \bar{y} - b_2 \bar{x}$$

$$= \begin{bmatrix} \dfrac{\bar{y}\Sigma x_i^2 - \bar{x}\Sigma(x_i y_i)}{\Sigma x_i^2 - N\bar{x}^2} \\ \dfrac{\Sigma(x_i y_i) - N\bar{x}\bar{y}}{\Sigma x_i^2 - N\bar{x}^2} \end{bmatrix} = \begin{bmatrix} \dfrac{y\Sigma x_i^2 - \bar{x}\Sigma(x_i y_i)}{\Sigma x_i^2 - N\bar{x}^2} \\ \dfrac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\bar{y} - \frac{[\Sigma(x_i - \bar{x})(y_i - \bar{y})]\bar{x}}{\Sigma(x_i - \bar{x})^2} = \frac{\bar{y}(\overset{-N\cancel{\bar{x}^2}\bar{y}}{\Sigma x_i^2 - 2x_i\bar{x} + \bar{x}^2}) - \bar{x}\Sigma(\overset{-N\bar{x}\bar{y}^{\,2}}{x_i y_i - \bar{x} y_i - x_i \bar{y} + \bar{x}\bar{y}})}{\Sigma(x_i - \bar{x})^2}$$

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

Let $K=2$, show that cov(b1, b2) in p. 30 of slides in Ch 5 reduces to the formula of in (2.14) - (2.16).

$$Var(b) = \sigma^2 (X'X)^{-1}$$

$$= \sigma^2 \left(\frac{1}{N(\Sigma x_i^2) - (\Sigma x_i)^2} \begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & N \end{bmatrix} \right)$$

$$= \begin{bmatrix} \dfrac{\sigma^2 \Sigma x_i^2}{N\Sigma x_i^2 - (N\bar{x})^2} & \dfrac{-N\bar{x}}{N\Sigma x_i^2 - (N\bar{x})^2} \\ \dfrac{-N\bar{x}}{N\Sigma x_i^2 - (N\bar{x})^2} & \dfrac{N}{N\Sigma x_i^2 - (N\bar{x})^2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\sigma^2 \Sigma x_i^2}{N\,\Sigma(x_i - \bar{x})^2} & \dfrac{-\bar{x} \cdot \sigma^2}{\Sigma x_i^2 - N\bar{x}^2} \\ \dfrac{-\bar{x} \cdot \sigma^2}{\Sigma x_i^2 - N\bar{x}^2} & \dfrac{\sigma^2}{\Sigma x_i^2 - N\bar{x}^2} \end{bmatrix}$$

$\rightarrow var(b_1|x)$ $\rightarrow cov(b_1, b_2|x)$

$var(b_2|x)$

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	0.0575	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. $1.4515 / 2.2019$

ii. The standard error for b_2. $2.7648 / 5.7103$

iii. The estimate b_3. $0.3695 \times (-3.9376)$

iv. R^2. $1 - \dfrac{46221.62}{(1200-1) \times (6.39547)^2} \doteq 0.0575$

v. $\hat{\sigma}$. $\sqrt{\dfrac{SSE}{N-K}} = \sqrt{\dfrac{46221.62}{1196}} \doteq 6.2169$

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

b_2: a 1% increase in TOTEXP will increase WALC by 0.02765 percentage points, holding other factors constant.

b_3: If the household has one more child, WALC decreased by 1.4549 percentage points, holding other factors constant.

b_4: If the age of the household head increases by 1 year, WALC decreases by 0.1503 percentage points, holding other factors constant.

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

$-0.1503 \pm 1.96 \times 0.0235$

95% C.I. $= [-0.1964, -0.1042]$

We have 95% confidence that the true population β_4 lies within this range.

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

Except intercept, all coefficient estimates are significantly different from 0 at a 5% level because the p-values are all less than 0.05.

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.

$$\begin{cases} H_0: \beta_3 = -2 \\ H_a: \beta_3 \neq -2 \end{cases} \qquad t = \frac{-1.4549-(-2)}{0.3695} = 1.475 < 1.96 = t_{0.975}$$

There is no evidence to suggest that having an extra child leads to a decline in WALC that is different from 2 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?

β_2 $-$
β_3 $+$
β_4 X

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?

$$\widehat{PRICE} = 90.8467 - 0.0600\ QUANT + 0.1162\ QUAL - 2.3546\ TREND$$
$$(se)\quad (8.5803)\quad (0.0102)\quad (0.2033)\quad (1.3861)$$
$$(t)\quad (10.588)\quad (-5.892)\quad (0.572)\quad (-1.699)$$

$R^2 = 0.5097$

They imply that as QUANT increases by 1 unit, the mean price will go down by 0.1600. Also as QUAL increases by 1 unit, the mean price will go up by 0.1162. As TREND increases by 1 year, the mean price decreases by 2.3546.

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

0.5097

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.

$$\begin{cases} H_0: \beta_2 \geq 0 \\ H_1: \beta_2 < 0 \end{cases}$$

$t = -5.892 < -1.675 = t_{(0.05, 52)}$

We reject H_0 and conclude that sellers are willing to accept a lower price if they can make sales in large quantities.

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.

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

$t = 0.572 < 1.675 = t_{(0.95, 52)}$

We do not reject H_0. We can't conclude 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?

-2.3546, the price decreases over time.

A possible reason for a decreasing price is the development of improved technology for producing cocaine.