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)

$$X = \begin{pmatrix} 1 & x_{1,2} & \cdots & x_{1,K} \\ 1 & x_{2,2} & \cdots & x_{2,K} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N,2} & \cdots & x_{N,K} \end{pmatrix}, \Rightarrow X = \begin{bmatrix} 1 & x_{1,2} \\ 1 & x_{2,2} \\ \vdots & \vdots \\ 1 & x_{N,2} \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}$$

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

$$b = (X'X)^{-1}(X'Y) = \left[\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{pmatrix}\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}\right]^{-1}\left[\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}\right]$$

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

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

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

$$= \begin{bmatrix} \dfrac{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i - \sum_{i=1}^N x_i (\sum_{i=1}^N x_i y_i)}{N(\sum_{i=1}^N x_i^2) - (\sum_{i=1}^N x_i)^2} \\ \\ \dfrac{N(\sum_{i=1}^N x_i y_i) - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{N(\sum_{i=1}^N x_i^2) - (\sum_{i=1}^N x_i)^2} \end{bmatrix} = \begin{bmatrix} \dfrac{\sum_{i=1}^N x_i^2 \cancel{N}\bar{y} - \cancel{N}\bar{x}(\sum_{i=1}^N x_i y_i)}{\cancel{N}(\sum_{i=1}^N x_i^2) - N^{\cancel{2}}\bar{x}^2} \\ \\ \dfrac{\cancel{N}(\sum_{i=1}^N x_i y_i) - \cancel{N}\bar{x}N\bar{y}}{\cancel{N}\sum_{i=1}^N x_i^2 - N^{\cancel{2}}\bar{x}^2} \end{bmatrix} = \begin{bmatrix} \dfrac{\bar{y}\sum_{i=1}^N x_i^2 - \bar{x}(\sum_{i=1}^N x_i y_i)}{\sum_{i=1}^N x_i^2 - N\bar{x}^2} \\ \\ \dfrac{\sum_{i=1}^N (x_i y_i) - N\bar{x}\bar{y}}{\sum_{i=1}^N x_i^2 - N\bar{x}^2} \end{bmatrix}$$

$$= \begin{bmatrix} \bar{y} - b_2\bar{x} \\ \\ \dfrac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \end{bmatrix} = \begin{bmatrix} b_1 \\ \\ b_2 \end{bmatrix}$$

$$b_2 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad (2.7)$$

$$b_1 = \bar{y} - b_2\bar{x} \qquad (2.8)$$

$$b_2 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i y_i - \bar{x}y_i - \bar{y}x_i + \bar{x}\bar{y})}{\sum(x_i^2 - 2\bar{x}x_i + \bar{x}^2)} = \frac{\sum x_i y_i - \bar{x}N\bar{y} - \bar{y}N\bar{x} + N\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x}N\bar{x} + N\bar{x}^2} = \frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^2 - N\bar{x}^2}$$

$$b_1 = \bar{y} - b_2\bar{x} = \bar{y} - \left(\frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^2 - N\bar{x}^2}\right)\bar{x} = \frac{(\sum x_i^2)\bar{y} - \cancel{N\bar{x}^2\bar{y}} - (\sum x_i y_i)\bar{x} + \cancel{N\bar{x}^2\bar{y}}}{\sum x_i^2 - N\bar{x}^2} = \frac{\bar{y}\sum x_i^2 - \bar{x}\sum x_i y_i}{\sum x_i^2 - N\bar{x}^2}$$

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).

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

$$= \sigma^2 \left[\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix} \right]^{-1}$$

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

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

Var(b1|X) cov(b1,b2|X)

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

cov(b1,b2|X) var(b2|X)

$$\left(\sum (x_i-\bar{x})^2 = \sum (x_i^2 - 2\bar{x}x_i + \bar{x}^2) = \sum x_i^2 - 2\bar{x}N\bar{x} + N\bar{x}^2 = \sum x_i^2 - N\bar{x}^2 \right)$$

$$\operatorname{var}(b_1|\mathbf{x}) = \sigma^2 \left[\frac{\sum x_i^2}{N\sum (x_i-\bar{x})^2} \right] \quad (2.14)$$

$$\operatorname{var}(b_2|\mathbf{x}) = \frac{\sigma^2}{\sum (x_i-\bar{x})^2} \quad (2.15)$$

$$\operatorname{cov}(b_1, b_2|\mathbf{x}) = \sigma^2 \left[\frac{-\bar{x}}{\sum (x_i-\bar{x})^2} \right] \quad (2.16)$$

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.
- **ii.** The standard error for b_2.
- **iii.** The estimate b_3.
- **iv.** R^2.
- **v.** $\hat{\sigma}$.

a. (i) $t_{b_1} = \dfrac{1.4515}{2.2019} \doteq 0.6592$

(ii) $se(b_2) = \dfrac{2.7648}{5.7103} \doteq 0.4842$

(iii) $\hat{b}_3 = -3.9376 \times 0.3695 \doteq -1.4549$

(iv) $R^2 = 1 - \dfrac{46{,}221.62}{(6.39547)^2 \times (1200-1)} \doteq 0.0575$

(v.) $\hat{\sigma} = \sqrt{\dfrac{\sum \hat{e}_i^2}{N-K}} = \sqrt{\dfrac{46{,}221.62}{1200-4}} \doteq 6.2167$

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

b_2：在其他變數保持不變下，當家庭總支出增加1%時，酒類支出佔總支出的百分比將約增加0.027648個百分點。

b_3：在其他變數保持不變下，當家庭多一名孩童，酒類支出佔總支出的百分比將降低1.4549個百分點。

b_4：在其他變數保持不變下，當戶主年齡增加一歲時，酒類支出佔總支出的百分比將降低0.1503個百分點。

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

$$-0.1503 - 1.96 \times 0.0235 = -0.1964$$

$$-0.1503 + 1.96 \times 0.0235 = -0.1042$$

$\Rightarrow$ The 95% interval estimate $= [-0.1964, -0.1042]$

有95%的信心水準，真正母體的β_4會落在這個區間，且這個區間意味著，若戶主年齡增加1歲，則估計酒類支出在總支出中所占的比率會下降0.1042至0.1964個百分點。

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

除了截距項外，所有的迴歸係數估計在5%的顯著水準下都和零顯著不同，因為它們的p值都小於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.

$H_0: \beta_3 = -2$

$H_1: \beta_3 \neq -2$

$$t = \frac{-1.4549-(-2)}{0.3695} = 1.4752 < \underset{(t_{0.975,1196})}{1.96} \Rightarrow \text{無法拒絕 } H_0$$

因此沒有足夠證據顯示，家庭多一名小孩對酒類支出佔比所造成的降低幅度和2個百分點有顯著差異。(沒有證據接受H_1)

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：預期為負號，因隨著每筆交易克數增加，每克價格應該會降低，因為量大會有折扣的現象。

β_3：預期為正，隨cocaine的純度愈高，價格就愈高。

β_4：符號取決於需求與供給在時間上的變動情況。
ex 當需求固定，但供給增加，則價格會下降。

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?

```
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 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
```

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

$\beta_2 = -0.06$ ⇒ 表示在其他條件不變下，當數量增加1單位，平均價格會下降0.06

$\beta_3 = 0.1162$ ⇒ 表示在其他條件不變下，當純度提高1單位時，平均價格會上升0.1162

$\beta_4 = -2.3546$ ⇒ 表示在其他條件不變下，時間每增加1年，平均價格就會降低2.3546，
係供給的成長速度快於需求的成長速度。

⇒ All係數都符合預期。

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

$R^2 = 0.5097$

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.

$H_0: \beta_2 \geq 0 \quad H_1: \beta_2 < 0$

$t = -5.892 \quad |-5.892| > 1.675 \; (t_{(0.95, 52)})$

拒絕 H_0，接受 H_1，代表若能一次成交較大數量的交易，賣方願意接受較低的價格。

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.

$H_0: \beta_3 \leq 0 \quad H_1: \beta_3 > 0$

$t = 0.572 < t_{(0.95, 52)} = 1.675 \Rightarrow$ 無法拒絕 H_0，代表沒有充份證據顯示較高的cocaine純度會得到額外的價格溢價.

f. What is the average annual change in the cocaine price? Can you suggest why price might be changing in this direction?

$\beta_4 = -2.3546 \Rightarrow$ 表示在其他條件不變下，時間每增加1年，平均價格就會降低2.3546 (From 1984 - 1991)

因此代表 The average annual change in the cocaine price is -2.3546.

$\Rightarrow$ 價格下降的可能原因之一，可能是因隨著時間技術進步，因此供給者能在相同時間內產出更多cocaine，導致供給大於需求，故平均價格隨時間下降。