1.Let K=2, show that(b1,b2) in p.29 of slides in CH5 reduces to the formula of (b1,b2) in (2.7)-(2.8)

$K=2$, $Y=X\beta+e$ 其中 $Y=(y_1, y_2, y_3 \cdots y_N)'$, $\beta=(\beta_1, \beta_2)'$

$e=(e_1, e_2, \cdots e_N)'$ $\quad X=\begin{bmatrix} 1 & X_{1,2} \\ 1 & X_{2,2} \\ \vdots & \vdots \\ 1 & X_{N,2} \end{bmatrix}$

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

$$(X'X)=\begin{bmatrix} 1, 1, \cdots 1 \\ X_{1,2} X_{2,2} \cdots X_{N,2} \end{bmatrix}\begin{bmatrix} 1 & X_{1,2} \\ 1 & X_{2,2} \\ 1 & X_{3,2} \\ \vdots & \vdots \\ 1 & X_{N,2} \end{bmatrix}=\begin{bmatrix} N & \sum_{i=1}^{N} X_{i,2} \\ \sum_{i=1}^{N} X_{i,2} & \sum_{i=1}^{N} X_{i,2}^2 \end{bmatrix}$$

$$\Rightarrow (X'X)^{-1}=\frac{1}{N\sum_{i=1}^{N} X_{i,2}^2-\left(\sum_{i=1}^{N} X_{i,2}\right)^2}\begin{bmatrix} \sum_{i=1}^{N} X_{i,2}^2 & -\sum_{i=1}^{N} X_{i,2} \\ -\sum_{i=1}^{N} X_{i,2} & N \end{bmatrix}$$

$$X'Y=\begin{bmatrix} 1, 1, \cdots 1 \\ X_{1,2} X_{2,2} \cdots X_{N,2} \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}=\begin{pmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} X_{i,2} y_i \end{pmatrix}$$

$$(X'X)^{-1}(X'Y)=\frac{1}{N\sum_{i=1}^{N} X_{i,2}^2-\left(\sum_{i=1}^{N} X_{i,2}\right)^2}\begin{bmatrix} \sum_{i=1}^{N} X_{i,2}^2 & -\sum_{i=1}^{N} X_{i,2} \\ -\sum_{i=1}^{N} X_{i,2} & N \end{bmatrix}\begin{pmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} X_{i,2} y_i \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}=\begin{bmatrix} \dfrac{\sum_{i=1}^{N} X_{i,2}^2 \sum_{i=1}^{N} y_i-\sum_{i=1}^{N} X_{i,2}\sum_{i=1}^{N} X_{i,2} y_i}{N\sum_{i=1}^{N} X_{i,2}^2-\left(\sum_{i=1}^{N} X_{i,2}\right)^2} \\ \dfrac{-\sum_{i=1}^{N} X_{i,2}\sum_{i=1}^{N} y_i+N\sum_{i=1}^{N} X_{i,2} y_i}{N\sum_{i=1}^{N} X_{i,2}^2-\left(\sum_{i=1}^{N} X_{i,2}\right)^2} \end{bmatrix}$$

$$\Rightarrow b_2=\frac{-N^2\bar{x}\bar{y}+N\sum_{i=1}^{N} X_{i,2} y_i}{N\sum_{i=1}^{N} X_{i,2}^2-(N\bar{x})^2}=\frac{\sum_{i=1}^{N} X_{i,2} y_i-N\bar{x}\bar{y}}{\sum_{i=1}^{N} X_{i,2}^2-N\bar{x}^2}=\frac{\sum_{i=1}^{N}(X_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{N}(X_i-\bar{x})^2} \#$$

$$\Rightarrow b_1=\frac{\sum_{i=1}^{N} X_{i,2}^2 N\bar{y}-N\bar{x}\sum_{i=1}^{N} X_{i,2} y_i-N^2\bar{x}^2\bar{y}+N^2\bar{x}^2\bar{y}}{N\sum_{i=1}^{N} X_{i,2}^2-N^2\bar{x}^2}$$

$$=\frac{\sum_{i=1}^{N} X_{i,2}^2\bar{y}-N\bar{x}^2\bar{y}}{\sum_{i=1}^{N}(X_i-\bar{x})^2}-\bar{x}\left(\frac{\bar{x}\sum_{i=1}^{N} X_{i,2} y_i-N\bar{x}^2\bar{y}}{\sum_{i=1}^{N}(X_i-\bar{x})^2}\right)=\bar{y}-\bar{x}b_2 \#$$

2.Let K=2, show that cov(b1,b2) in p.30 of slides in CH5 reduces to the formula of (2.14)-(2.16)

$$Var(b)=\sigma^2(X'X)^{-1}=\frac{\sigma^2}{N\sum_{i=1}^{N}X_{i,2}^2-\left(\sum_{i=1}^{N}X_{i,2}\right)^2}\begin{bmatrix}\sum_{i=1}^{N}X_{i,2}^2 & -\sum_{i=1}^{N}X_{i,2}\\ -\sum_{i=1}^{N}X_{i,2} & N\end{bmatrix}$$

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

$$Cov(b_1,b_2)=\frac{-\sigma^2\sum_{i=1}^{N}X_{i,2}}{N\sum_{i=1}^{N}X_{i,2}^2-\left(\sum_{i=1}^{N}X_{i,2}\right)^2}=\frac{-\sigma^2 N\bar{X}}{N\sum_{i=1}^{N}X_{i,2}^2-N^2\bar{X}^2}=\frac{-\sigma^2\bar{X}}{\left(\sum_{i=1}^{N}X_{i,2}-\bar{X}\right)^2}$$

$$Var(b_1)=\frac{\sigma^2\sum_{i=1}^{N}X_{i,2}^2}{N\sum_{i=1}^{N}X_{i,2}^2-\left(\sum_{i=1}^{N}X_{i,2}\right)^2}=\frac{\sigma^2\sum_{i=1}^{N}X_{i,2}^2}{N\sum_{i=1}^{N}(X_i-\bar{X})^2}$$

$$Var(b_2)=\frac{N\sigma^2}{N\sum_{i=1}^{N}X_{i,2}^2-\left(\sum_{i=1}^{N}X_{i,2}\right)^2}=\frac{\sigma^2}{\sum_{i=1}^{N}(X_i-\bar{X})^2}$$

#

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	0.6592	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	1.0575	Mean dependent var		6.19434
S.E. of regression	6.2167	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. $\hat{\sigma}$.

$$i: t\ statistic\ for\ b_1 = \frac{1.4515}{2.2019} = 0.6592$$

$$ii: std\ error\ for\ b_2 = \frac{2.7648}{5.7103} = 0.4842$$

$$iii: b_3 = 0.3695 \times (-3.9376) = -1.4549$$

$$iv: R^2 = 1 - \frac{46221.62}{1199 \times (6.39547)^2} = 0.0575$$

$$v: \hat{\sigma} = \left(\frac{46221.62}{1196}\right)^{0.5} = 6.2167$$

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

β_2:LN(TOTEXP)每上升 1，WALC 上升 2.7648%

β_3:number of kids 每上升 1 個，WALC 減少 1.4549%

β_4:age 每上升 1 year，WALC 減少 0.1503%

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

$\beta_4 \pm t_{1196,0.025} \times std\ error$

➔$(-0.1503 - 1.96 \times 0.0235, -0.1503 + 1.96 \times 0.0235)$

➔$(-0.1964, -0.1042)$

95%的信心水準下，age 每上升 1 year，WALC 減少的幅度介於0.1042 、0.1964 之間

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

除了 constant 以外，其他都顯著，因為其 p-value 皆小於 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.

$H_0: \beta_3 = -2$ 、$H_1: \beta_3 \neq -2$

$t = \frac{\beta_3 - (-2)}{std\ error} = \frac{-1.4549+2}{0.3695} = 1.4752 < 1.96$ ➔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?

β_2:預期是負的，若量大可能可壓低每克的購買價格
β_3:預期是正的，純度越高表示其生產成本越高，因此出售價格亦較高
β_4:不確定，一段時間內可能有政策禁令或需求增減狀況，無法推論正負

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
```

β_2:-0.05997 預期為負，符合預期
β_3:0.11621 預期為正，符合預期
β_4:-2.35458 預期為不確定，不符合預期

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

```
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
```

Multiple R-squared: 0.5097，表示有約 51%可被 variation in quantity, quality, time 解釋

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.

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

$\beta\_2: -0.05997$ 、$std.Error: 0.01018$ 、$t\ value: -5.892$

$t_{0.95,52} = 1.675 < 5.892$➔拒絕$H_0$

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.

$H_0: \beta_3 = 0$ 、$H_1: \beta_3 > 0$

$\beta\_3: 0.11624$ 、$std.Error: 0.20326$ 、$t\ value: 0.572$

$t_{0.95,52} = 1.675 > 0.572$➔無法拒絕$H_0$➔表示品質不見得能有較高價格

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

平均年變化為$\beta_4 = -2.35458$，表示平均價格每過一年下降約 2.35 美元/每克

原因可能是 cocaine 的生產技術進步而使產量大幅增加(供給過多)導致價格下跌