$$\hat{\beta} = (X'X)^{-1}(X'Y)$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \end{bmatrix}$$

$$X'X = \begin{bmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} = \begin{bmatrix} N & N\bar{x} \\ N\bar{x} & \Sigma x_i^2 \end{bmatrix}$$

for $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 逆矩陣 $= \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

① 證明 $b_2 = \dfrac{\Sigma(x_i - \bar{x})\Sigma(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$

$$b_1 = \bar{y} - b_2\bar{x}$$

$$\frac{1}{ad-bc} = \frac{1}{N\Sigma x_i^2 - N^2\bar{x}^2} = \frac{1}{N(\Sigma x_i^2 - N\bar{x}^2)}$$

$$\Sigma(x_i - \bar{x})^2 = \Sigma(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \Sigma x_i^2 - 2\bar{x}\Sigma x_i + N\bar{x}^2$$

$$= \Sigma x_i^2 - 2\bar{x}N\bar{x} + N\bar{x}^2 = \Sigma x_i^2 - N\bar{x}^2$$

$$\Rightarrow \frac{1}{N\Sigma(x_i - \bar{x})^2}$$

$$(X'X)^{-1} = \frac{1}{N\sum(X_i-\bar{X})^2}\begin{bmatrix}\sum X_i^2 & -N\bar{X} \\ -N\bar{X} & N\end{bmatrix}$$

$$(XY') = \begin{bmatrix}\sum Y_i \\ \sum X_i Y_i\end{bmatrix} = \begin{bmatrix}N\bar{Y} \\ \sum X_i Y_i\end{bmatrix}$$

$$\begin{bmatrix}\sum X_i^2 & -N\bar{X} \\ -N\bar{X} & N\end{bmatrix}\begin{bmatrix}N\bar{Y} \\ \sum X_i Y_i\end{bmatrix} = \begin{bmatrix}N\bar{Y}\sum X_i^2 - N\bar{X}\sum X_i Y_i \\ -N^2\bar{X}\bar{Y} + N\sum X_i Y_i\end{bmatrix}$$

$$b_2 = \frac{1}{\sum(X_i-\bar{X})^2}\left(\sum X_i Y_i - N\bar{X}\bar{Y}\right)$$

$$\begin{aligned}\sum X_i Y_i - N\bar{X}\bar{Y} &= \sum X_i Y_i - \sum X_i \bar{Y} \\ &= \sum(X_i Y_i - X_i\bar{Y}) = \sum X_i(Y_i - \bar{Y}) \\ &= \sum(X_i-\bar{X})(Y_i-\bar{Y}) + \bar{X}\underbrace{\sum(Y_i-\bar{Y})}_{=0}\end{aligned}$$

$$\Rightarrow \frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2} = b_2$$

#

$$b_1 = \frac{\bar{Y}\sum X_i^2 - \bar{X}\sum X_i Y_i}{\sum (X_i - \bar{X})^2}$$

$$b_2 \sum (X_i - \bar{X})^2 = \sum X_i Y_i - N\bar{X}\bar{Y}$$

$$\sum X_i Y_i = b_2 \sum (X_i - \bar{X})^2 - N\bar{X}\bar{Y}$$

$$\bar{X}\sum X_i Y_i = \bar{X}\left(b_2 \sum (X_i - \bar{X})^2 - N\bar{X}\bar{Y}\right)$$

$$\bar{Y}\sum X_i^2 - \bar{X} b_2 \sum (X_i - \bar{X})^2 - N\bar{X}^2\bar{Y}$$

$$= \bar{Y}\left(\sum X_i^2 - N\bar{X}^2\right) - \bar{X} b_2 \sum (X_i - \bar{X})^2$$

$$= \bar{Y}\sum (X_i - \bar{X})^2 - \bar{X} b_2 \sum (X_i - \bar{X})^2$$

$$= \sum (X_i - \bar{X})^2 (\bar{Y} - b_2\bar{X})$$

$$b_1 = \frac{\sum (X_i - \bar{X})^2 (\bar{Y} - b_2\bar{X})}{\sum (X_i - \bar{X})^2}$$

$$b_1 = \bar{Y} - b_2\bar{X}$$ #

② 證明

$$var(b_1) = \sigma^2 \left[\frac{\Sigma x_i^2}{N\Sigma(x_i-\bar{x})^2} \right]$$

$$var(b_2) = \frac{\sigma^2}{\Sigma(x_i-\bar{x})^2}$$

$$cov(b_1, b_2) = \sigma^2 \left[\frac{-\bar{x}}{\Sigma(x_i-\bar{x})^2} \right]$$

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

↗ 接續上題資訊

$$(X'X)^{-1} = \frac{1}{N\Sigma(x_i-\bar{x})^2} \begin{bmatrix} \Sigma x_i^2 & -N\bar{x} \\ -N\bar{x} & N \end{bmatrix}$$

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \sigma^2 \begin{bmatrix} \frac{\Sigma x_i^2}{N(x_i-x_i)^2} & \frac{-\bar{x}}{\Sigma(x_i-\bar{x})^2} \\ \frac{-\bar{x}}{\Sigma(x_i-\bar{x})^2} & \frac{1}{\Sigma(x_i-\bar{x})^2} \end{bmatrix}$$ #

得証

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.

$$i = \frac{1.4515}{2.2019} = 0.6592$$

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

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

$$iv = R^2 = 1 - 46221.62/(6.39547^2 \times 1199) = 0.0575$$

$$v = \left(\frac{46221.62}{(1200-4)}\right)^{\frac{1}{2}} = 6.2167$$

b、

$\beta_2 \Rightarrow$ 1% increase in ln(TOTEXP)

$\rightarrow$ WALC $\uparrow$ 2.7648%

$\beta_3 \Rightarrow$ 1% increase in NK

$\rightarrow$ WALC $\downarrow$ 1.4549%

$\beta_4 \Rightarrow$ 1% increase in AGE

$\rightarrow$ WALC $\downarrow$ 0.1503%

c、

$-0.1503-1.96\times0.0235, -0.1503+1.96\times0.0235$

$=[-0.1964, -0.1042]$

当AGE↑，WALC有95%會↓ 0.1042%~0.1964%

d、

除了intercept，其它皆顯著，因只有其P值>0.05

e、

$H_0: \beta_4 = -2$

$H_1: \beta_4 \neq -2$

$t = \dfrac{-1.4549-(-2)}{0.3695} = 1.4752$

$1.4752 < 1.96$，不拒絕 H_0