HW Q1 / Q2 :

* 試證當 $k=2$ 時，b 和 CH2 提到關於 b_1, b_2 的公式及抽樣分配結果相同

$Y = X\beta + e$

其中 Y 為 $N\times 1$ 之 vector ，且 $X = \begin{pmatrix} 1 & X_{12} \\ 1 & X_{22} \\ \vdots & \vdots \\ 1 & X_{N2} \end{pmatrix}$; $k=2$

X $N \times K$

β $K \times 1$

e $N \times 1$

$$X'X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{22} & \cdots & X_{N2} \end{pmatrix} \begin{pmatrix} 1 & X_{12} \\ 1 & X_{22} \\ \vdots & \vdots \\ 1 & X_{N2} \end{pmatrix}$$

$$= \begin{pmatrix} N & \sum_{i=1}^{N} X_{i2} \\ \sum_{i=1}^{N} X_{i2} & \sum_{i=1}^{N} X_{i2}^2 \end{pmatrix}$$

$$(X'X)^{-1} = \frac{1}{N\sum_{i=1}^{N} X_{i2}^2 - \left(\sum_{i=1}^{N} X_{i2}\right)^2} \begin{pmatrix} \sum_{i=1}^{N} X_{i2}^2 & -\sum_{i=1}^{N} X_{i2} \\ -\sum_{i=1}^{N} X_{i2} & N \end{pmatrix}$$

即行列式 = Δ

$$X'Y = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{N2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_{i2} y_i \end{pmatrix}$$

$$\tilde{b} = (X'X)^{-1}(X'Y) = \frac{1}{\Delta} \begin{pmatrix} \sum_{i=1}^{N} x_{i2}^2 & -\sum_{i=1}^{N} x_{i2} \\ \sum_{i=1}^{N} x_{i2} & N \end{pmatrix} \begin{pmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_{i2} y_i \end{pmatrix}$$

$$= \frac{1}{\Delta} \begin{pmatrix} \sum_{i=1}^{N} x_{i2}^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_{i2} \sum_{i=1}^{N} x_{i2} y_i \\ -\sum_{i=1}^{N} x_{i2} \sum_{i=1}^{N} y_i + N \sum_{i=1}^{N} x_i y_i \end{pmatrix}$$

且 $Cov(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \dfrac{\sigma^2}{\Delta} \begin{pmatrix} \sum_{i=1}^{N} x_{i2}^2 & -\sum_{i=1}^{N} x_{i2} \\ \sum_{i=1}^{N} x_{i2} & N \end{pmatrix}$

因此 $\begin{cases} b_1 = \dfrac{1}{\Delta} \left(\sum_{i=1}^{N} x_{i2}^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_{i2} \sum_{i=1}^{N} x_{i2} y_i \right) \\ b_2 = \dfrac{1}{\Delta} \left(-\sum_{i=1}^{N} x_{i2} \sum_{i=1}^{N} y_i + N \sum_{i=1}^{N} x_i y_i \right) \end{cases}$

$$\begin{cases} Var(b_1) = \dfrac{\sigma^2}{\Delta} \sum_{i=1}^{N} x_{i2}^2 \\ Var(b_2) = \dfrac{\sigma^2}{\Delta} N \end{cases}$$

$$\hat{b} \sim N\left(\hat{\beta},\ \sigma^2 \begin{pmatrix} \dfrac{\sum x_{i2}^2}{\Delta} \\ \dfrac{N}{\Delta} \end{pmatrix}\right)$$

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.

$= N$

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 $= SSE$		

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) $t_1 = \dfrac{1.4515}{2.2019} = 0.6592$

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

(iii) $\hat{b}_3 = -3.9376 \times 0.3695 = -1.4558$

(iv) 先求 $TSS = (6.39547)^2 \times (1{,}200 - 1) = 49{,}024.9$

且已知 $SSE = 46{,}221.62$

$\Rightarrow R^2 = 1 - \dfrac{46{,}221.62}{49{,}024.9} = 0.0573$

(a) $\hat{\sigma} = \sqrt{MSE} = \sqrt{\frac{SSE}{n-K}} = \sqrt{\frac{46{,}221.62}{1{,}200-4}} = 6.216$

b.

$b_2 = 2.7648$：表示家庭總支出每增加1%，酒精支出占比會增加2.7648百分點

$b_3 = -1.4558$：表示每多一個小孩，家庭酒精支出占比少1.46百分點

$b_4 = -0.1503$：表示家庭負責人每年年齡增加，酒精支出占比減少0.1503百分點

c.

$$\begin{cases} \hat{\beta}_4 = -0.1503 \\ SE(\hat{\beta}_4) = 0.0235 \end{cases}$$

在 $N=1{,}200$ 下，適用大樣本，採用Z尾分

$$C.I. = \hat{\beta}_4 \pm Z_{0.025} \times SE(\hat{\beta}_4)$$

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

$$= (-0.1963, -0.1043)$$

<u>解釋</u>：我們有95%的信心認為，家庭負責人年齡

對酒精支出占比的影響係數落在

-0.1963 到 -0.1043 之間

d、

在 $\alpha = 0.05$ 下，

$$\begin{cases} \hat{\beta}_1 \text{ 之 p-value} = 0.5099 > 0.05 \Rightarrow \text{不顯著} \\ \hat{\beta}_2 \text{ 之 p-value} = 0.0000 < 0.05 \Rightarrow \text{顯著} \\ \hat{\beta}_3 \text{ 之 p-value} = 0.0001 < 0.05 \Rightarrow \text{顯著} \\ \hat{\beta}_4 \text{ 之 p-value} = 0.0000 < 0.5 \Rightarrow \text{顯著} \end{cases}$$

⇒ 除常數項以外，皆顯著

e、

即 ①
$$\begin{cases} H_0: \beta_3 = -2 \\ H_a: \beta_3 \neq -2 \end{cases}$$

② $\alpha = 0.05$

③ $$z = \frac{\hat{\beta}_3 - (-2)}{SE(\hat{\beta}_3)}$$ （∵大樣本下）

④ $$z_1 = \frac{-1.4558 - (-2)}{0.3695} = 1.4725$$

⑤ $Z_1 = 1.4725 < Z_{0.025} = 1.96$

$\therefore$ do not reject H_0

$\Rightarrow$ 即在 $\alpha = 0.05$ 下，無法否定 NK係數 $= 2$，即

無法否定每多一位孩子會使支出減少 2 個百分點

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?

已知迴歸模型為:

$$PRICE = \beta_1 + \beta_2 QUANT + \beta_3 QUAL + \beta_4 TREND + e$$

a.

- β_2（QUANT）：預期為**負**。根據經濟學常理，**大宗購買有折扣**（quantity discount）。

- β_3（QUAL）：預期為**正**。品質越好，價格越高（quality premium）。

- β_4（TREND）：不確定。可能隨時間上升（通膨/需求上升）或下降（執法更嚴、技術進步），視情況而定。

b.

```
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
```

模型在 R 跑出來的結果:

- β_2（QUANT）：和預期相同為**負且顯著**，表示**大宗購買有折扣**。

- β_3（QUAL）：和預期相同為**正但不顯著**（p-value 大於預設的顯著水準 0.05）

- β_4（TREND）：模型結果係數為負但不顯著。表示隨時間上升下通膨/需求下降（執法更嚴、技術進步）。

c.

即求 R^2=0.50965

代表「由數量、品質、趨勢」共同解釋的價格變異比例。

d.

左尾檢定：是否販售數量越多，價格越低

假設設定：

- H_0:β_2=0（數量對價格無影響）
- H_1:β_2<0（數量越多，價格越低）

得到 p-value= 1.42536e-07<<顯著水準 0.05

拒絕 H_0，表示數量越多，價格越低。

e.

雙尾檢定：品質對價格是否有影響

假設設定：

- $H_0: \beta_3=0$（品質無影響）
- $H_1: \beta_3 \neq 0$（品質對價格有影響）

得到 p-value= 0.569992>顯著水準 0.05

不拒絕 H_0，表示無法否定品質無影響。

f.

計算 β_4（TREND 的係數）= -2.354579

該係數表示每年（TREND 單位 = 1 年）價格變化的金額。如果為正，則是每年價格平均上升的金額；反之則是下降。

價格下降可能有以下幾個原因：

1. **供給增加：** **生產者或走私增加，市場供給變多，導致價格下降。**
2. **執法更嚴：** **增加交易風險，轉向規模化、低價銷售以減少曝光風險。**
3. **品質下降：** **為了應對執法與成本，部分交易品質變差，拉低整體售價。**
4. **競爭加劇：** **更多供應者進入市場，價格競爭加劇。**
5. **需求結構改變：** **若消費者轉向其他藥物，對可卡因的需求下降，價格下滑。**

The coefficient of trend is approximately **-1.10**, indicating that the average cocaine price **decreased by $1.10 per gram each year** during 1984–1991.
This downward trend may be explained by increased market competition, greater law enforcement pressure, or increased supply from trafficking sources.