HW0324

1. 建立矩陣形式：

$y = Xb + \varepsilon$

構建正規方程：

最小化殘差平方和：

$S(b_1,b_2) = \Sigma(y_i - b_1x_{i1} - b_2x_{i2})^2$

正規方程組：

$\partial S/\partial b_1 = -2\Sigma x_{i1}(y_i - b_1x_{i1} - b_2x_{i2}) = 0$

$\partial S/\partial b_2 = -2\Sigma x_{i2}(y_i - b_1x_{i1} - b_2x_{i2}) = 0$

整理後：

$b_1\Sigma x_{i1}^2 + b_2\Sigma x_{i1}x_{i2} = \Sigma x_{i1}y_i$

$b_1\Sigma x_{i1}x_{i2} + b_2\Sigma x_{i2}^2 = \Sigma x_{i2}y_i$

矩陣表示：

$[\ \Sigma x_{i1}^2\ \Sigma x_{i1}x_{i2}\ ]\ [b_1] = [\Sigma x_{i1}y_i]$

$[\ \Sigma x_{i1}x_{i2}\ \Sigma x_{i2}^2\ ]\ [b_2]\ [\Sigma x_{i2}y_i]$

記作：$X'Xb = X'y$

求解矩陣方程：

令：

$a = \Sigma x_{i1}^2$

$b = \Sigma x_{i1}x_{i2}$

$d = \Sigma x_{i2}^2$

逆矩陣：

$(X'X)^{-1} = 1/(ad-b^2)\ [\ d\ -b\ ]$

$[\ -b\ a\ ]$

解：

$[b_1] = 1/(ad-b^2)\ [\ d\ -b\ ][\Sigma x_{i1}y_i]$

$[b_2]\ [\ -b\ a\ ][\Sigma x_{i2}y_i]$

最終解：

$b_1 = (d\Sigma x_{i1}y_i - b\Sigma x_{i2}y_i)/(ad-b^2)$

$b_2 = (a\Sigma x_{i2}y_i - b\Sigma x_{i1}y_i)/(ad-b^2)$

完整公式：

$b_1 = [(\Sigma x_{i2}^2)(\Sigma x_{i1}y_i) - (\Sigma x_{i1}x_{i2})(\Sigma x_{i2}y_i)] / [(\Sigma x_{i1}^2)(\Sigma x_{i2}^2) - (\Sigma x_{i1}x_{i2})^2]$

$b_2 = [(\Sigma x_{i1}^2)(\Sigma x_{i2}y_i) - (\Sigma x_{i1}x_{i2})(\Sigma x_{i1}y_i)] / [(\Sigma x_{i1}^2)(\Sigma x_{i2}^2) - (\Sigma x_{i1}x_{i2})^2]$

2. 在普通最小平方法(OLS)中，係數向量 b 的變異數-共變異數矩陣為：
   $Var(b) = \sigma^2(X'X)^{-1}$
   其中 σ^2是誤差變異數。
   K=2 情況
   當設計矩陣 X 是 n×2 矩陣時，定義：
   $a = \Sigma x_{i1}^2$ (i=1 到 n)
   $b = \Sigma x_{i1}x_{i2}$
   $d = \Sigma x_{i2}^2$
   矩陣推導
   a. X'X 矩陣：
      $[X'X] = [\ a\ b\ ]$
      $[\ b\ d\ ]$
   b. 逆矩陣計算：
      $(X'X)^{-1} = 1/(ad-b^2) \times [\ d\ -b\ ]$
      $[\ -b\ a\ ]$
   c. 變異數-共變異數矩陣：
      $Var(b) = \sigma^2/(ad-b^2) \times [\ d\ -b\ ]$
      $[\ -b\ a\ ]$
   共變異數推導
   b_1和 b_2的共變異數為矩陣的非對角線元素：
   $cov(b_1,b_2) = -\sigma^2 b/(ad-b^2)$
   結論
   兩個迴歸係數估計量 b_1和 b_2的共變異數為：
   $cov(b_1,b_2) = -\sigma^2(\Sigma x_{i1}x_{i2})/[(\Sigma x_{i1}^2)(\Sigma x_{i2}^2)-(\Sigma x_{i1}x_{i2})^2]$

3.

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.

變數	係數	標準誤差	t 統計量	概率
C (截距)	1.4515	2.2019	0.6592	0.5099
ln(TOTEXP)	2.7648	0.4841	5.7103	0.0000
NK	-1.4545	0.3695	-3.9376	0.0001
AGE	-0.1503	0.0235	-6.4019	0.0000
R^2	0.1132			

b.

b_2 (ln(TOTEXP))：

2.7648 表示總支出每增加 1%（對數形式），家庭酒精預算占比平均增加 2.76 個百分點。

b_3 (NK)：

-1.4545 表示每增加一個孩子，家庭酒精預算占比平均減少 1.45 個百分點，反映孩子數量增加會降低酒精支出比例。

b_4 (AGE)：

-0.1503 表示戶主年齡每增加 1 歲，酒精預算占比平均減少 0.15 個百分點，表示年齡越大，酒精消費占比越低。

c. $b4 \pm t0.025,1196 \times \text{Std. Error}$

$-0.1503 \pm 1.962 \times 0.0235 - 0.1503 \pm 1.962 \times 0.0235$

$=[-0.1964,-0.1042] = [-0.1964,-0.1042]$

d. 所有變數對於估計酒精預算占比皆有顯著影響。

e. 1

假設設定：

虛無假設 H0:β3=−2$H0:\beta3=-2$

對立假設 H1:β3≠−2$H1:\beta3!=-2$

檢驗統計量：

t=−1.4545−(−2)0.3695≈1.476$t=0.3695-1.4545-(-2)\approx 1.476$

臨界值：

t0.025,1196≈1.962$t0.025,1196\approx 1.962$

結論：

由於 ∣1.476∣<1.962∣1.476∣<1.962，無法拒絕 H0$H0$。

→ 增加一個孩子使酒精預算占比減少 2 個百分點的假設未被推翻。

4.

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?

a.

β_2 (QUANT)：預期為負（數量越大，單價可能越低，因量大折扣或風險考量）。
β_3 (QUAL)：預期為正（純度越高，價格越高，反映品質溢價）。
β_4 (TREND)：預期為正（隨時間推移，通脹或市場需求可能推升價格）。

b.

```
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 .
```

係數意義為變數每增加 1 單位，古柯鹼價格會變動多少單位，其中時間因素係數方向與預期不符，可能是因為古柯鹼會隨擺放時間變成而跌價。

c. `Multiple R-squared:  0.5097`，表示有 50.97%古柯鹼價格變動的比例可以被這三個變數所解釋。

d. $H0 : \beta2 \geq 0$（數量不影響或提高價格）

H1 : β2<0$H1:\beta2<0$（數量增加降低價格）

```
> t_stat <- summary(model)$coefficients["quant", "t value"]
> p_value <- pt(t_stat, df = 52)
> p_value
[1] 1.42536e-07
```

由 p-value 可知，有足夠證據拒絕數量會使價格增加的虛無假設。

e. H0 : β3=0$H0:\beta3=0$（品質不影響價格）

H1 : β3>0$H1:\beta3>0$（高品質有溢價）

```
> t_stat_qual <- summary(model)$coefficients["qual", "t value"]
> p_value_qual <- pt(t_stat_qual, df = 52, lower.tail = FALSE)
> p_value_qual
[1] 0.284996
```

由 p-value 可知，沒有證據拒絕高品質的古柯鹼會對價格帶來溢價。

f. 根據係數 `trend        -2.35458`，每過一年，古柯鹼價格約會下降 2.35%，我認為這是由於古柯鹼本身可能會因為潮濕或其他因素而變質，進而影響價格所致。