HW0331

5.6 Suppose that, from a sample of 63 observations, the least squares estimates and the corresponding estimated covariance matrix are given by

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} \qquad \widehat{\text{cov}}(b_1, b_2, b_3) = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Using a 5% significance level, and an alternative hypothesis that the equality does not hold, test each of the following null hypotheses:

a. $\beta_2 = 0$
b. $\beta_1 + 2\beta_2 = 5$
c. $\beta_1 - \beta_2 + \beta_3 = 4$

a. $H_0: \beta_2 = 0$, $H_1: \beta_2 \neq 0 \Rightarrow t = \frac{b_2 - 0}{se(b_2)} = \frac{3}{\sqrt{4}} = 1.5 < 2.0003 = t_{(0.975, 60)}$

$\Rightarrow$ 不拒絕 H_0 #

b. $H_0: \beta_1 + 2\beta_2 = 5$, $H_1: \beta_1 + 2\beta_2 \neq 5 \Rightarrow t = \frac{2+2\times3-5}{\sqrt{3+4\times4+2\times1\times2\times(-2)}} = \frac{3}{\sqrt{11}} \doteq 0.9045 < 2.003$

$\Rightarrow$ 不拒絕 H_0 #

c. $H_0: \beta_1 - \beta_2 + \beta_3 = 4$, $H_1: \beta_1 - \beta_2 + \beta_3 \neq 4 \Rightarrow t = \frac{2-3-1-4}{\sqrt{3+4+3-2\times(-2)+2\times1-2\times0}} = \frac{-6}{4}$

$= -1.5 \Rightarrow |t| = 1.5 < 2.0003$

$\Rightarrow$ 不拒絕 H_0 #

5.31 Each morning between 6:30 AM and 8:00 AM Bill leaves the Melbourne suburb of Carnegie to drive to work at the University of Melbourne. The time it takes Bill to drive to work (*TIME*), depends on the departure time (*DEPART*), the number of red lights that he encounters (*REDS*), and the number of trains that he has to wait for at the Murrumbeena level crossing (*TRAINS*). Observations on these variables for the 249 working days in 2015 appear in the file *commute5*. *TIME* is measured in minutes. *DEPART* is the number of minutes after 6:30 AM that Bill departs.

a. Estimate the equation

$$TIME = \beta_1 + \beta_2 DEPART + \beta_3 REDS + \beta_4 TRAINS + e$$

Report the results and interpret each of the coefficient estimates, including the intercept β_1.

$$TIME = 20.8701 + 0.3681 \cdot DEPART + 1.5219 \cdot REDS + 3.0237 \cdot TRAINS$$

```
Call:
lm(formula = time ~ depart + reds + trains, data = commute5)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.4389  -3.6774  -0.1188   4.5863  16.4986 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  20.8701     1.6758  12.454  < 2e-16 ***
depart        0.3681     0.0351  10.487  < 2e-16 ***
reds          1.5219     0.1850   8.225 1.15e-14 ***
trains        3.0237     0.6340   4.769 3.18e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.299 on 245 degrees of freedom
Multiple R-squared:  0.5346,	Adjusted R-squared:  0.5289 
F-statistic: 93.79 on 3 and 245 DF,  p-value: < 2.2e-16
```

$\beta_1 = 20.8701$，當其他變數為 0 時，基本通勤時間為 20.8701 分鐘

$\beta_2 = 0.3681$，固定其他變數下，出發時間晚一分鐘，通勤時間增加 0.3681 分鐘

$\beta_3 = 1.5219$，每多遇到一個紅燈，時長增加 1.5219 分鐘

$\beta_4 = 3.0237$，每多遇到一列火車，時長增加 3.0237 分鐘

b. Find 95% interval estimates for each of the coefficients. Have you obtained precise estimates of each of the coefficients?

```
> print(conf_intervals)
                 2.5 %    97.5 %
(Intercept) 17.5694018 24.170871
depart       0.2989851  0.437265
reds         1.1574748  1.886411
trains       1.7748867  4.272505
```

trains 的信賴區間較寬，估計的不確定性較高

c. Using a 5% significance level, test the null hypothesis that Bill's expected delay from each red light is 2 minutes or more against the alternative that it is less than 2 minutes.

$$H_0: \beta_3 \geq 2, H_1: \beta_3 < 2$$

$t = \frac{1.5219-2}{0.185} = -2.584 < -1.65$ ➔拒絕H_0

d. Using a 10% significance level, test the null hypothesis that the expected delay from each train is 3 minutes against the alternative that it is not 3 minutes.

$$H_0\colon \beta_4 = 3, H_1\colon \beta_4 \neq 3$$

$t = \frac{3.0237-3}{0.634} = 0.0374 < 1.65$➔不拒絕$H_0$

e. Using a 5% significance level, test the null hypothesis that Bill can expect a trip to be at least 10 minutes longer if he leaves at 7:30 AM instead of 7:00 AM, against the alternative that it will not be 10 minutes longer. (Assume other things are equal.)

$$H_0\colon (60-30)\beta_2 \geq 10, H_1\colon (60-30)\beta_2 < 10$$

$t = \frac{0.3681-1/3}{0.0351} = 0.9905 > -1.65$➔不拒絕$H_0$

f. Using a 5% significance level, test the null hypothesis that the expected delay from a train is at least three times greater than the expected delay from a red light against the alternative that it is less than three times greater.

$$H_0\colon \beta_4 \geq 3\beta_3, H_1\colon \beta_4 < 3\beta_3$$

$t = \frac{3.0237-3*1.5219}{(0.40197+9*0.03423+6*0.000648)^{0.5}} = -1.825 < -1.65$➔拒絕$H_0$

g. Suppose that Bill encounters six red lights and one train. Using a 5% significance level, test the null hypothesis that leaving Carnegie at 7:00 AM is early enough to get him to the university on or before 7:45 AM against the alternative that it is not. [Carry out the test in terms of the expected time $E(TIME|\mathbf{X})$ where **X** represents the observations on all explanatory variables.]

$$H_0\colon \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \leq 45, H_1\colon \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45$$

$t = \frac{20.8701+30*0.3681+6*1.5219+3.0237-45}{se(\beta_1+30\beta_2+6\beta_3+\beta_4)} = -1.726 < 1.65$➔不拒絕$H_0$

➔7.出門時間足夠

h. Suppose that, in part (g), it is imperative that Bill is not late for his 7:45 AM meeting. Have the null and alternative hypotheses been set up correctly? What happens if these hypotheses are reversed?

$$H_0\colon \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45, H_1\colon \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \leq 45$$

$t = \frac{20.8701+30*0.3681+6*1.5219+3.0237-45}{se(\beta_1+30\beta_2+6\beta_3+\beta_4)} = -1.726 < -1.65$➔拒絕$H_0$

➔7.出門時間足夠

5.33 Use the observations in the data file *cps5_small* to estimate the following model:

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EDUC^2 + \beta_4 EXPER + \beta_5 EXPER^2 + \beta_6 (EDUC \times EXPER) + e$$

a. At what levels of significance are each of the coefficient estimates "significantly different from zero"?

除了 educ^2 以外，所有係數都在 1%水準下顯著異於 0

```
Call:
lm(formula = log(wage) ~ educ + I(educ^2) + exper + I(exper^2) +
    I(educ * exper), data = cps5_small)

Residuals:
    Min      1Q  Median      3Q     Max
-1.6628 -0.3138 -0.0276  0.3140  2.1394

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.038e+00  2.757e-01   3.764 0.000175 ***
educ             8.954e-02  3.108e-02   2.881 0.004038 **
I(educ^2)        1.458e-03  9.242e-04   1.578 0.114855
exper            4.488e-02  7.297e-03   6.150 1.06e-09 ***
I(exper^2)      -4.680e-04  7.601e-05  -6.157 1.01e-09 ***
I(educ * exper) -1.010e-03  3.791e-04  -2.665 0.007803 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4638 on 1194 degrees of freedom
Multiple R-squared:  0.3227,    Adjusted R-squared:  0.3198
F-statistic: 113.8 on 5 and 1194 DF,  p-value: < 2.2e-16
```

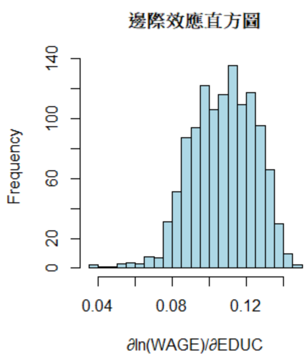
b. Obtain an expression for the marginal effect $\partial E[\ln(WAGE)|EDUC, EXPER]/\partial EDUC$. Comment on how the estimate of this marginal effect changes as *EDUC* and *EXPER* increase.

$$\frac{\partial E[\,ln(WAGE) \mid EDUC, EXPER\,]}{\partial EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$$

$$= 0.08954 + 0.001458 * 2 * EDUC - 0.00101 * EXPER$$

$EDUC$ 上升➔marginal effect 上升; $EXPER$ 上升➔ marginal effect 下降

c. Evaluate the marginal effect in part (b) for all observations in the sample and construct a histogram of these effects. What have you discovered? Find the median, 5th percentile, and 95th percentile of the marginal effects.




```
> cat("第 5 百分位數:", quantiles[1], "\n")
第 5 百分位數: 0.08008187
> cat("中位數:", quantiles[2], "\n")
中位數: 0.1084313
> cat("第 95 百分位數:", quantiles[3], "\n")
第 95 百分位數: 0.1336188
```

➔可見 EDUC 上升會提升 WAGE(邊際效應都是正的)，直方圖呈現鐘型

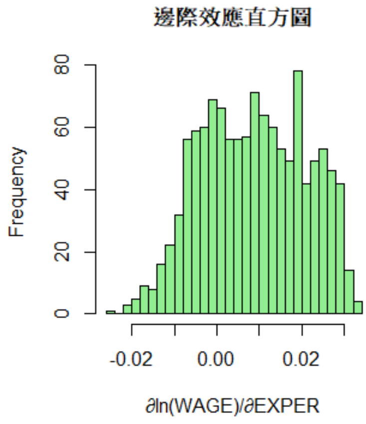
d. Obtain an expression for the marginal effect $\partial E\left[\ln(WAGE)|EDUC, EXPER\right]/\partial EXPER$. Comment on how the estimate of this marginal effect changes as *EDUC* and *EXPER* increase.

$$\frac{\partial E[ln(WAGE)]}{\partial EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$$

$$= 0.04488 - 2 * 0.000468 * EXPER - 0.00101 * EDUC$$

$EDUC$ 上升➔marginal effect 下降; $EXPER$ 上升➔ marginal effect 下降

e. Evaluate the marginal effect in part (d) for all observations in the sample and construct a histogram of these effects. What have you discovered? Find the median, 5th percentile, and 95th percentile of the marginal effects.




```
> quantiles_exper <- quantile(marginal_effect_exper_e, probs = c(0.0
5, 0.5, 0.95))
> cat("第 5 百分位數:", quantiles_exper[1], "\n")
第 5 百分位數: -0.01037621
> cat("中位數:", quantiles_exper[2], "\n")
中位數: 0.008418878
> cat("第 95 百分位數:", quantiles_exper[3], "\n")
第 95 百分位數: 0.02793115
```

有部分邊際效應是負的

f. David has 17 years of education and 8 years of experience, while Svetlana has 16 years of education and 18 years of experience. Using a 5% significance level, test the null hypothesis that Svetlana's expected log-wage is equal to or greater than David's expected log-wage, against the alternative that David's expected log-wage is greater. State the null and alternative hypotheses in terms of the model parameters.

$$\left(\beta_1 + \beta_2(17) + \beta_3(289) + \beta_4(8) + \beta_5(64) + \beta_6(136)\right)$$
$$- \left(\beta_1 + \beta_2(16) + \beta_3(256) + \beta_4(18) + \beta_5(324) + \beta_6(288)\right)$$
$$= \beta_2 + 33\beta_3 - 10\beta_4 - 260\beta 5 - 152\beta_6$$

$$H_0: \beta_2 + 33\beta_3 - 10\beta_4 - 260\beta_5 - 152\beta_6 \le 0$$
$$H_1: \beta_2 + 33\beta_3 - 10\beta_4 - 260\beta_5 - 152\beta_6 > 0$$

```
> # 輸出結果
> cat("差值:", delta, "\n")
差值: -0.03588456
> cat("標準誤:", se_delta, "\n")
標準誤: 0.02148902
> cat("t 統計量:", t_stat, "\n")
t 統計量: -1.669902
> cat("p 值:", p_value, "\n")
p 值: 0.9525307
```

P 值>0.05➔不拒絕H_0

g. After eight years have passed, when David and Svetlana have had eight more years of experience, but no more education, will the test result in (f) be the same? Explain this outcome?

$$\begin{aligned}&\big(\beta_1+\beta_2(17)+\beta_3(289)+\beta_4(16)+\beta_5(16*16)+\beta_6(17*16)\big)\\&\qquad-\big(\beta_1+\beta_2(16)+\beta_3(256)+\beta_4(26)+\beta_5(26*26)+\beta_6(16*26)\big)\\&\qquad=\beta_2+33\beta_3-10\beta_4-420\beta_5-144\beta_6\end{aligned}$$

$$H_0:\beta_2+33\beta_3-10\beta_4-420\beta_5-144\beta_6\le 0$$
$$H_1:\beta_2+33\beta_3-10\beta_4-420\beta_5-144\beta_6> 0$$

```
> # 輸出結果
> cat("差值:", delta_g, "\n")
差值: 0.03091716
> cat("標準誤:", se_delta_g, "\n")
標準誤: 0.01499112
> cat("t 統計量:", t_stat_g, "\n")
t 統計量: 2.062365
> cat("p 值:", p_value_g, "\n")
p 值: 0.01958648
>
```

P 值<0.05➔拒絕H_0

h. Wendy has 12 years of education and 17 years of experience, while Jill has 16 years of education and 11 years of experience. Using a 5% significance level, test the null hypothesis that their marginal effects of extra experience are equal against the alternative that they are not. State the null and alternative hypotheses in terms of the model parameters.

$$(\beta_4+34\beta_5+12\beta_6)-(\beta_4+22\beta_5+16\beta_6)=12\beta_5-4\beta_6$$

$$H_0:12\beta_5-4\beta_6=0$$
$$H_1:12\beta_5-4\beta_6\neq 0$$

```
> # 輸出結果
> cat("差值:", delta_h, "\n")
差值: -0.001575327
> cat("標準誤:", se_delta_h, "\n")
標準誤: 0.001533457
> cat("t 統計量:", t_stat_h, "\n")
t 統計量: -1.027304
> cat("p 值:", p_value_h, "\n")
p 值: 0.8478613
>
```

P 值>0.05➔拒絕H_0

i. How much longer will it be before the marginal effect of experience for Jill becomes negative? Find a 95% interval estimate for this quantity.

```
> # 輸出結果
> cat("再過多久經驗邊際效應變為負值:", theta, "\n")
再過多久經驗邊際效應變為負值: 19.67706
> cat("95% 置信區間: [", ci_lower, ", ", ci_upper, "]\n")
95% 置信區間: [ 15.96146 ,  23.39265 ]
>
```