5.31

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.

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

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.

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.

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.)

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.

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 $\mathbf{X}$ represents the observations on all explanatory variables.]

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?

a.

```
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
```

β_1= 20.8701, 在6:30出發，且沒有遇到紅燈與等待火車通過，抵達大學的時間為20.8701
β_2= 0.3681, 在其他條件都不變之下，他晚一單位的時間出門，抵達時間會增加0.3681單位
β_3= 1.5219, 在其他條件都不變之下，他每多遇到一個紅綠燈，抵達時間會增加1.5219單位
β_4= 3.0237,在其他條件都不變之下，他每多等待一個火車，抵達時間會增加3.0237單位

b.

```
> confints <- confint(mod1,  level=0.95)
> confints
                2.5 %    97.5 %
(Intercept) 17.5694018 24.170871
depart       0.2989851  0.437265
reds         1.1574748  1.886411
trains       1.7748867  4.272505
```

Yes, we obtained precise estimator of each of coefficients.

c.

```
     reds
-2.583562
> tcr <- qt(alpha, df)
> tcr
[1] -1.651097
```

$H_0 : \beta_3 > = 2$
$H_1 : \beta_3 < 2$

$-2.583562 < -1.651097$ ，所以拒絕H_0

d.

```
   trains
0.03737444
```

```
> tcr_10
[1] -1.651097
```

$H_0 : \beta_4 = 3$
$H_1 : \beta_4 \neq 3$

0.03737444 < 1.651097 ，所以不拒絕 H_0

e.

```
   depart
0.9911646
```

```
> tcr
[1] -1.651097
```

$H_0 : 60 * \beta_2 - 30 * \beta_2 > = 10$
$H_1 : 60 * \beta_2 - 30 * \beta_2 < 10$

$0.9911646 > -1.651097$ ，所以不拒絕H_0

f.

```
-1.825027
```

```
[1] -1.651097
```

$H_0 : \beta_4 >= 3 * \beta_3$
$H_1 : \beta_4 < 3 * \beta_3$

$-1.825027 < -1.651097$ ，所以我們拒絕H_0

g.

```
[1,] -1.725964
```

```
[1] 1.651097
```

$H_0 : \beta_1 + 30 * \beta_2 + 6 * \beta_3 + \beta_4 <= 45$
$H_1 : \beta_1 + 30 * \beta_2 + 6 * \beta_3 + \beta_4 > 45$

$-1.725964 < 1.651097$ ，所以我們不拒絕H_0

f.

從g小題來看，我們沒有證據說明通勤時間會大於45分鐘，因此如果要檢驗是否會遲到應該要用以下的假設

$H_0 : \beta_1 + 30 * \beta_2 + 6 * \beta_3 + \beta_4 >= 45$
$H_1 : \beta_1 + 30 * \beta_2 + 6 * \beta_3 + \beta_4 < 45$

$-1.725964 < -1.651097$，所以我們拒絕虛無假設，接受替代假設
因此可以說 Bill 不會的通勤時間不會超過 45 分鐘

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"?

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.

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.

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

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.

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.

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?

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.

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.

a.

```
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
```

從p-value 來看，每個係數都是顯著異於0，除了educ^2這個係數外

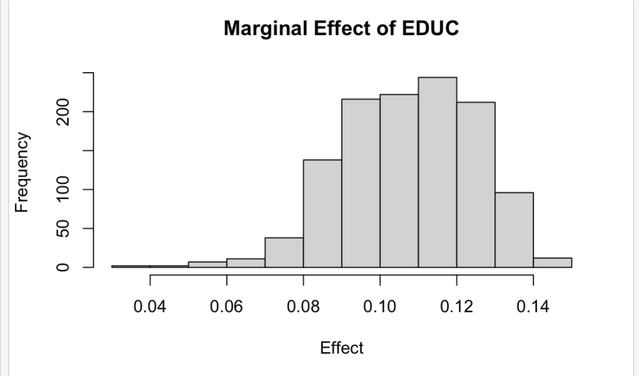
b.

Marginal effect :

$\beta_2 + 2 * \beta_3 * EDUC + \beta_6 * EXPER$ = 0.08954+2*(0.001458)*EDUC−0.00101*EXPER

The marginal effect of education increases as EDUC increase, but decreases as EXPER increase.

c.



從圖中可以發現大部分的數據都集中在5%–95%的區間之中

```
        5%        50%        95%
0.08008187 0.10843125 0.13361880
```

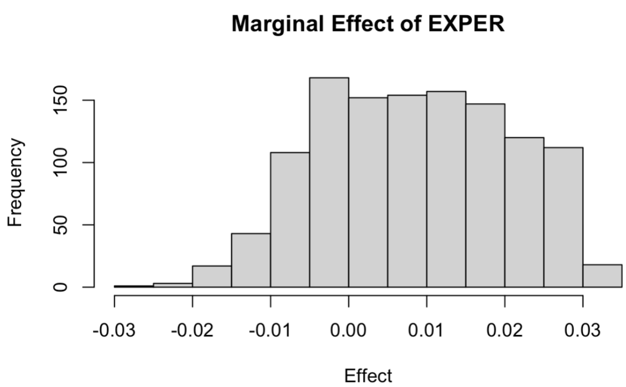
```
> #c.
> summary(marginal_educ)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
0.03565 0.09513 0.10843 0.10735 0.12050 0.14787 
```

d.

Marginal effect：
$\beta_4 + 2*\beta_5*EXPER + \beta_6*EDUC$ = 0.04488 – 2*(0.000468)*EXPER–0.00101*EDUC

The marginal effect of experience decreases as EXPER increase, and decreases as EDUC increase.

e.




```
          5%          50%          95% 
-0.010376212  0.008418878  0.027931151 
```

從圖中可以發現大部分的是正的，但仍有少部分的 marginal effect 是負的

```
> #e.
> summary(marginal_exper)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
-0.025279 -0.001034  0.008419  0.008652  0.018586  0.033989 
```

f.

$H_0 : \beta_1 + 16 * \beta_2 + 16^2 * \beta_3 + 18 * \beta_4 + 18^2 * \beta_5 + (16 * 18) * \beta_6 > = \beta_1 + 17 * \beta_2 + 17^2 * \beta_3 + 8 * \beta_4 + 8^2 * \beta_5 + (17 * 8) * \beta_6$

$H_0 : \beta_1 + 16 * \beta_2 + 16^2 * \beta_3 + 18 * \beta_4 + 18^2 * \beta_5 + (16 * 18) * \beta_6 < \beta_1 + 17 * \beta_2 + 17^2 * \beta_3 + 8 * \beta_4 + 8^2 * \beta_5 + (17 * 8) * \beta_6$

t 檢定

```
         [,1]
[1,] 1.669902
```

$\alpha = 0.05$

```
[1] -1.646131
```

1.669902 > −1.646131 ，所以我們不拒絕虛無假設

g.

$H_0 : \beta_1 + 16 * \beta_2 + 16^2 * \beta_3 + 26 * \beta_4 + 26^2 * \beta_5 + (16 * 26) * \beta_6 > = \beta_1 + 17 * \beta_2 + 17^2 * \beta_3 + 16 * \beta_4 + 16^2 * \beta_5 + (17 * 16) * \beta_6$

$H_0 : \beta_1 + 16 * \beta_2 + 16^2 * \beta_3 + 26 * \beta_4 + 26^2 * \beta_5 + (16 * 26) * \beta_6 < \beta_1 + 17 * \beta_2 + 17^2 * \beta_3 + 16 * \beta_4 + 16^2 * \beta_5 + (17 * 16) * \beta_6$

t 檢定

```
          [,1]
[1,] -2.062365
```

−2.062365 < − 1.669902，所以我們拒絕虛無假設，接受 David 的 log−wage 較高

h.

$H_0 : \beta_4 + 2 * \beta_5 * 17 + \beta_6 * 12 = \beta_4 + 2 * \beta_5 * 11 + \beta_6 * 16$

$H_1 : \beta_4 + 2 * \beta_5 * 17 + \beta_6 * 12 \neq \beta_4 + 2 * \beta_5 * 11 + \beta_6 * 16$

```
          [,1]
[1,] -1.027304
```

```
> tcr
[1] -1.961953
```

−1.027304 > −1.961953 ，所以我們不拒絕虛無假設

i.

```
> exper_zero
I(educ * exper) 
       19.67706 
> 
```

```
> lower
I(educ * exper) 
       15.95776 
> upper
I(educ * exper) 
       23.39636 
> 
```