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. H0: β2 = 0, Ha: β2 ≠ 0
   Test_statistic = 1.5
   Rejection region: {|t |> 2.0003}
   Non-reject H0, insufficient to say β2 ≠ 0
b. H0: β1 + 2β2 = 5, Ha: β1 + 2β2 ≠ 5
   Test_statistic = 0.9045
   Rejection region: {|t |> 2.0003}
   Non-reject H0, insufficient to say β1 + 2β2 ≠ 5
c. H0: β1 − β2 + β3 = 4, Ha: β1 − β2 + β3 ≠ 4
   Test_statistic = -1.5
   Rejection region: {|t |> 2.0003}
   Non-reject H0, insufficient to say β1 − β2 + β3 ≠ 4

Degree of freedom = 63 – 3 = 60

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.

```
(Intercept)      depart        reds      trains
 20.8701362   0.3681251   1.5219429   3.0236958
```

b1 = 20.87013, b2 = 0.3681, b3 = 1.5219, b4 = 3.0237

b1: 如果早上六點半出門，沒有遇到紅燈而且也沒有火車經過，預計抵達公司所需的時間約 21 分鐘。

b2: 每晚一分鐘出門，預計所需抵達公司時間增加約 24 秒。

b3: 每多遇到一個紅燈，預計所需抵達公司時間增加約 90 秒。

b4: 每多遇到一個台火車，預計所需抵達公司時間增加約 3 分鐘。

b. 下圖為各個估計值的 95%區間估計

	lb	ub
intercept	17.5694018	24.170871
depart	0.2989851	0.437265
reds	1.1574748	1.886411
trains	1.7748867	4.272505

c. H0: β3 >= 2, Ha: β3 < 2

Test_statistic = -2.583562

Rejection region: {t < -1.651097}

Non-reject H0, insufficient to say β3 < 2.

無法顯著說明多遇到一個紅綠燈，抵達公司增加的時間小於二

```
#c. H0: Beta3 >= 2
t = (coef(lr)[3]-2)/summary(lr)$coefficients[3, 2]
tc = qt(0.05, df, lower.tail = TRUE)
t < tc #reject H0
```

d. H0: β4 = 3, Ha: β4 ≠ 3

Test_statistic = 0.03737444

Rejection region: {|t| > 1.651097}

Non-reject H0, insufficient to say β4 ≠ 3.

無法顯著說明多遇到一輛火車，抵達公司增加的時間不等於三

```
#d. H0: Beta4 = 3
t = (coef(lr)[4]-3)/summary(lr)$coefficients[4, 2]
tc = qt(0.05, df, lower.tail = FALSE)
t > tc #non-reject H0
```

e. H0: β2 >= 1/3, Ha: β2 < 1/3

Test_statistic = 0.9911646

Rejection region: {t < -1.651097}

Non-reject H0, insufficient to say β2 < 1/3.

無法顯著說明晚一分鐘出門，抵達公司增加的時間小於 **20** 秒

```
#e. H0: 30Beta2 >= 10
#H0: Beta2 >= 1/3
t = (coef(lr)[2]-1/3)/summary(lr)$coefficients[2, 2]
tc = qt(0.05, df, lower.tail = TRUE)
t < tc #non-reject H0
```

f. $H_0: \beta_4 >= 3\beta_3$, $H_a: \beta_4 < 3\beta_3$

Test_statistic = -1.825027

Rejection region: $\{t < -1.651097\}$

Reject H0, sufficient to say $\beta_4 - \beta_3 >= 0$.

```
#f. H0: Beta4 >= 3Beta3
#H0: Beta4 - 3Beta3 >= 0
coeff = c(0, 0, -3, 1)
estL = sum(coeff * est)
varL = t(coeff)%*%vcovar%*%coeff
stdL = varL^0.5
t = estL/stdL
tc = qt(0.05, df, lower.tail = TRUE)
t < tc #reject H0
```

顯著說明遇到一班火車比起遇到三個紅綠燈，抵達公司所需額外的時間較短

g. $H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 <= 45$, $H_a: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45$

Test_statistic = -1.725964

Rejection region: $\{t > 1.651097\}$

Non-reject H0, insufficient to say $\beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45$.

無法顯著說明會遲到

```
#g. H0: B1 + 30B2 + 6B3 + B4 <= 45
coeff = c(1, 30, 6, 1)
estL = sum(coeff * est)
varL = t(coeff)%*%vcovar%*%coeff
stdL = varL^0.5
t = (estL - 45)/stdL
tc = qt(0.05, df, lower.tail = FALSE)
t > tc #non-reject H0 證據不夠說會遲到
```

h. 要將 g 小題的 H0 跟 Ha 調整

$H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 >= 45$, $H_a: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 < 45$

Test_statistic = -1.725964

Rejection region: $\{t < -1.651097\}$

Reject H0, sufficient to say $\beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 < 45$.

顯著說明不會遲到

```r
#h. H0: B1 + 30B2 + 6B3 + B4 >= 45
#Ha: B1 + 30B2 + 6B3 + B4 < 45
coeff = c(1, 30, 6, 1)
estL = sum(coeff * est)
varL = t(coeff)%*%vcovar%*%coeff
stdL = varL^0.5
t = (estL - 45)/stdL
tc = qt(0.05, df, lower.tail = TRUE)
t < tc #reject H0 可以準時
```

Degree of freedom = 249 – 4 = 245

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.

```
            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 ***
educ:exper  -1.010e-03  3.791e-04  -2.665 0.007803 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

當significant level 大於 p-value時，coefficient estimates “significantly different from zero”

b.

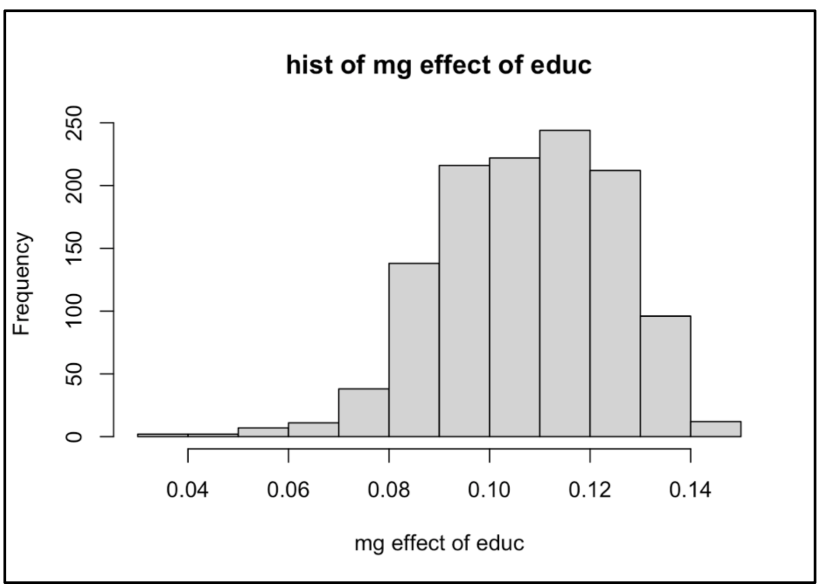
```
#b. marginal effect ∂E[ln(WAGE)|EDUC, EXPER]/∂EDUC
#β2 + 2*β3*EDUC + β6*EXPER
```

= 0.0894 + 2 * 0.001458 * EDUC – 0.00101EXPER

當教育年數（EDUC）上升，marginal effect 上升

當經驗（EXPER）上升，marginal effect 下降

c. Marginal effect 都是正的，且稍微左偏，表示一半以上的人如果接受多一年的教育能夠讓 ln(wage) 增加超過 0.1 單位。




```
        P5     Median        P95 
0.08008187 0.1084313 0.1336188 
```

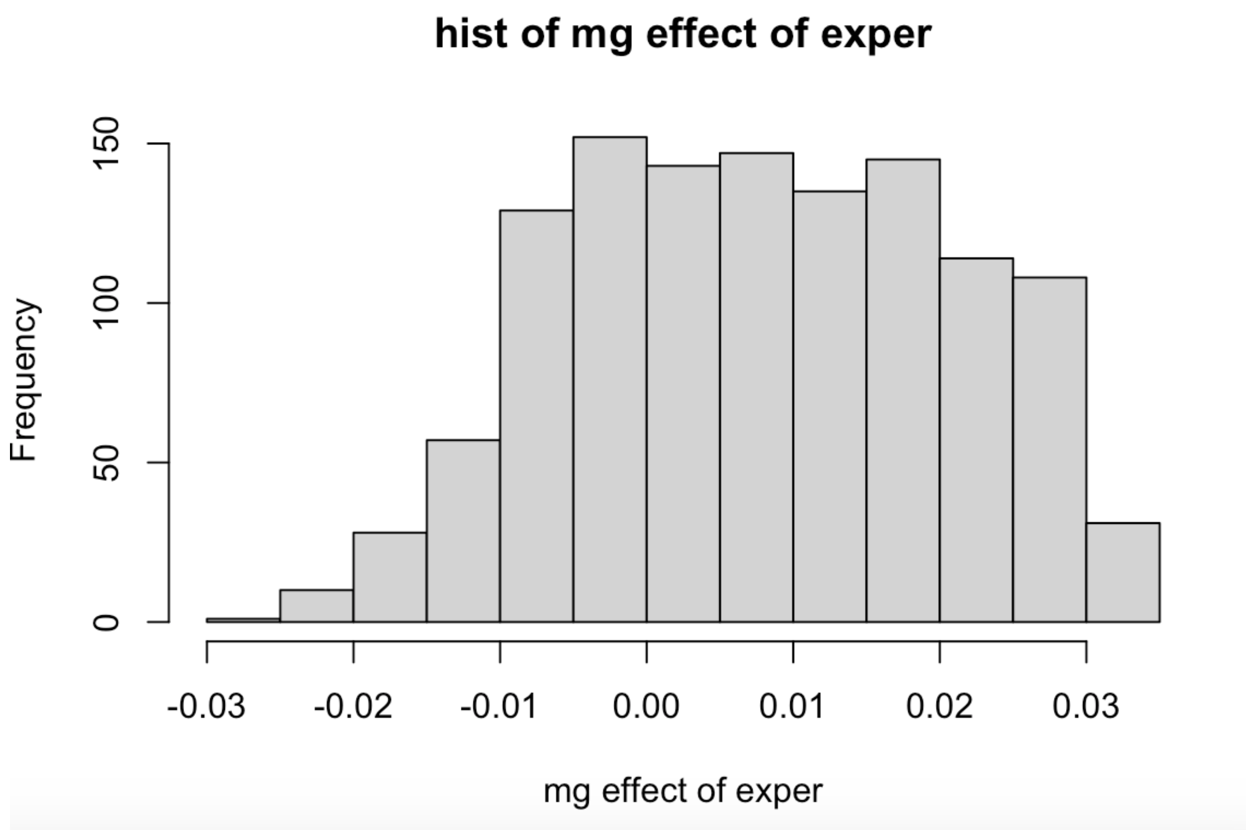
d.

```
#d. ∂E[ln(WAGE)|EDUC, EXPER]/∂EXPER
#β4 + 2*β5*EEPER + β6*EDUC
```

= 0.04487888 + 2 * (-0.0004680225) * EXPER - 0.001010236*EDUC

EXPER 或是 EDUC 上升，marginal of EXPER 都下降。

e. Marginal effect 普遍是正的，但有少部分人會因為經驗增加卻導致 ln(wage)下降。



P5	Median	P95
-0.0123942	0.007751161	0.02874354

f. H0: $-\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 >= 0$

Ha: $-\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 < 0$

Test_statistic = 1.669902

Rejection region: $\{t < -1.646131\}$

Non-reject H0, insufficient to say $-\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 < 0$.

無法顯著說明 **David's expected log-wage is greater**

g. H0: $-\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 >= 0$

Ha: $-\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 < 0$

Test_statistic = -2.062365

Rejection region: $\{t < -1.646131\}$

Reject H0, sufficient to say $-\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 < 0$.

顯著說明 **David's expected log-wage is greater**

h. H0: $12\beta_5 - 4\beta_6 = 0$

Ha: $12\beta_5 - 4\beta_6 \neq 0$

Test_statistic = -1.027304

Rejection region: $\{|t| > 1.646131\}$

Non-reject H0, insufficient to say $12\beta_5 - 4\beta_6 \neq 0$.

無法顯著說明 **their marginal effects of extra experience are not equal**

i.

```
      lb       ub
15.95776 23.39636
```

```
#i. x = (-β4 - β6)/ 2β5 -11
est = (-b4-16*b6)/(2*b5) - 11
g4 = -1/(2*b5)
g5 = (b4+16*b6)/(2*b5^2)
g6 = -8/b5
g = c(g4, g5, g6)
covv = vcov(mod)[4:6, 4:6]
varr = t(g)%*%covv%*%g
sd = varr^0.5
t = qt(0.025, df, lower.tail = FALSE)
lb = est - t * sd
ub = est + t * sd
interval = data.frame(lb = lb, ub = ub)
rownames(interval) = ""
```