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. $\beta_1 - \beta_2 + \beta_3 = 4$

(sol) a. $\begin{cases} H_0: \beta_2 = 0 \\ H_a: \beta_2 \neq 0 \end{cases}$

$\alpha = 0.05$

test statistic: $T = \dfrac{\hat{\beta}_2 - \beta_2}{SE(\hat{\beta}_2)} \overset{H_0}{\sim} t(n-k)$

$RR = \{ |T| \geq t_{0.025}(63-3) \approx 2 \}$

$T_0 = \dfrac{3-0}{\sqrt{4}} = 1.5 \notin RR$, do not reject H_0.

統計上，β_2 並未顯著異於 0。

b. $\begin{cases} H_0: \beta_1 + 2\beta_2 = 5 \\ H_a: \beta_1 + 2\beta_2 \neq 5 \end{cases}$

$\alpha = 0.05$

test statistic: $T = \dfrac{\hat{\beta}_1 + 2\hat{\beta}_2 - 5}{SE(\hat{\beta}_1 + 2\hat{\beta}_2)} \overset{H_0}{\sim} t(n-k)$

$RR = \{ |T| \geq t_{0.025}(63-3) \approx 2 \}$

$Var(\hat{\beta}_1 + 2\hat{\beta}_2) = Var(\hat{\beta}_1) + 4Var(\hat{\beta}_2) + 2 \cdot 2\, Cov(\hat{\beta}_1, \hat{\beta}_2)$
$= 3 + 4 \cdot 4 + 2 \cdot 2 \cdot (-2) = 11$

$\Rightarrow SE(\hat{\beta}_1 + 2\hat{\beta}_2) = \sqrt{Var(\hat{\beta}_1 + 2\hat{\beta}_2)} = \sqrt{11}$

$T_0 = \dfrac{2+6-5}{\sqrt{11}} = 0.9045 \notin RR$, do not reject H_0.

統計上，$\beta_1 + 2\beta_2$ 並未顯著異於 5

c、$\begin{cases} H_0: \beta_1 - \beta_2 + \beta_3 = 4 \\ H_a: \beta_1 - \beta_2 + \beta_3 = 4 \end{cases}$

$\alpha = 0.05$

test statistic: $T = \dfrac{\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3 - 4}{SE(\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3)} \overset{H_0}{\sim} t(n-K)$

$RR = \{ |T| \geq t_{0.025}(63-3) \approx 2 \}$

$$\begin{aligned} Var(\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3) &= Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + Var(\hat{\beta}_3) \\ &\quad - 2Cov(\hat{\beta}_1, \hat{\beta}_2) - 2Cov(\hat{\beta}_2, \hat{\beta}_3) + 2Cov(\hat{\beta}_1, \hat{\beta}_3) \\ &= 3 + 4 + 3 - 2\cdot(-2) - 2\cdot 0 + 2\cdot 1 \\ &= 16 \end{aligned}$$

$\Rightarrow SE(\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3) = \sqrt{Var(\hat{\beta}_1 - \hat{\beta}_2 + \hat{\beta}_3)} = \sqrt{16} = 4$

$T = \dfrac{2-3+(-1)-4}{4} = \dfrac{-6}{4} = -1.5 \notin RR$, do not reject H_0

統計上, $\beta_1 - \beta_2 + \beta_3$ 並未顯著4。

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?

(sol) 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
```

Intercept = 20.8701

The intercept represents the base time (in minutes) it takes Bill to drive to work when all other variables (DEPART, REDS, and TRAINS) are zero.

DEPART = 0.3681

For each minute Bill departs later, the commute time increases by approximately 0.3681 minutes.

REDS = 1.5219

For each red light Bill encounters, the commute time increases by about 1.5219 minutes.

TRAINS = 3.0237

For each train Bill encounters, the commute time increases by approximately 3.0237 minutes.

b.

```
> confint(model1,level=0.95)
                 2.5 %    97.5 %
(Intercept) 17.5694018 24.170871
depart       0.2989851  0.437265
reds         1.1574748  1.886411
trains       1.7748867  4.272505
```

These intervals suggest that we are 95% confident that the true values of the coefficients lie within these ranges.

c. $\begin{cases} H_0: \beta_3 \geq 2 \\ H_a: \beta_3 < 2 \end{cases}$

$\alpha = 0.05$

$RR = \{ T < -1.651097 \}$

test statistic: $-2.583562 \in RR$, reject H_0

The expected delay from each red light is less than 2 min.

d. $\begin{cases} H_0: \beta_4 = 3 \\ H_a: \beta_4 \neq 3 \end{cases}$

$\alpha = 0.1$

$RR = \{ |T| \geq 1.651097 \}$

test statistic: $0.3737 \notin RR$, do not reject H_0

The expected delay from each train is 3 min.

e. $\begin{cases} H_0: \beta_2 \geq \frac{1}{3} \\ H_a: \beta_2 < \frac{1}{3} \end{cases}$

$\alpha = 0.05$

$RR = \{ T < -1.651097 \}$

test statistic: $0.991 \notin$ do not reject H_0

The delaying departure time by 30 min. increase expected travel time by at least 10 min.

f. $\begin{cases} H_0: \beta_4 \geq 3\beta_3 \\ H_a: \beta_4 < 3\beta_3 \end{cases}$

$\alpha = 0.05$

$RR = \{ T < -1.651097 \}$

test statistic: $-1.825 \in RR$, reject H_0

The expected delay from a train is less than three times the delay from a red light.

g、 $\begin{cases} H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \le 45 \\ H_a: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45 \end{cases}$

$\alpha = 0.05$

$RR = \{T > 1.651099\}$

test statistic: -1.725964 ∴ do not reject H_0、

h、 If Bill not late for his meeting, he will wish to build a high probability that his commute time will be less than 45 min. so the alternative hypothesis is commute time less than 45 min、

$\Rightarrow \begin{cases} H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \ge 45 \\ H_a: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 < 45 \end{cases}$

We will reject H_0, ∵ $-1.726 < -1.651$.

Bill's expected commute time is such that he can expected to be on time for the meeting、

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.

<50分> 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
```

All coefficient estimates are significantly different from 0 at 1% level of significance, but I(educ^2) is significant at a 11.49% significance level.

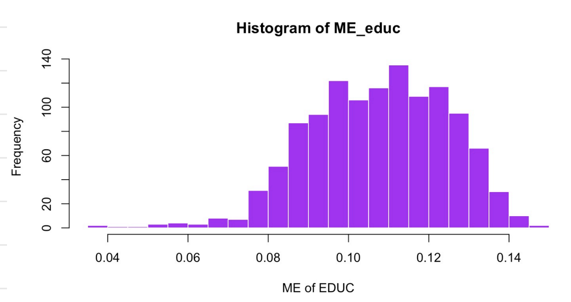
b. $\dfrac{\partial E[\ln(WAGE)|EDUC, EXPER]}{\partial EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$

$= 0.08954 + 0.002916\, EDUC - 0.001010\, EXPER$

EDUC ↑, margin effect ↑

EXPER ↑, margin effect ↓

c.




```
> cat("Median:",median_ME,'\n')
Median: 0.1084313
> cat("5th Percentile:",p5_ME,'\n')
5th Percentile: 0.08008187
> cat("95th Percentile:",p95_ME,'\n')
95th Percentile: 0.1336188
```

most of the marginal effects concentrated between 0.08-0.13.

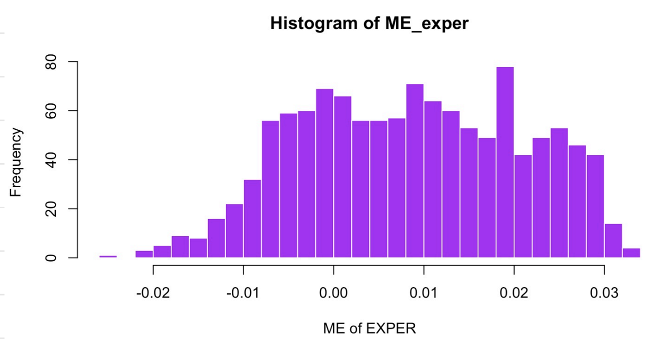
d. $\frac{\partial E[\ln(WAGE)\mid EDUC, EXPER]}{\partial EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$

$= 0.04488 + 2\cdot(-0.000468) EXPER - 0.00101 EDUC$

$= 0.04488 - 0.000936 EXPER - 0.00101 EDUC$

EDUC↑, ME↓ ; EXPER↑, ME↓

e.




```
> cat("Median:",median_ME,'\n')
Median: 0.008418878
> cat("5th Percentile:",p5_ME,'\n')
5th Percentile: -0.01037621
> cat("95th Percentile:",p95_ME,'\n')
95th Percentile: 0.02793115
```

the marginal effects are almost positive, but little of them are negative.

f. David: $\beta_1 + 17\beta_2 + 289\beta_3 + 8\beta_4 + 64\beta_5 + 136\beta_6$

Svetlana: $\beta_1 + 16\beta_2 + 256\beta_3 + 18\beta_4 + 324\beta_5 + 288\beta_6$

$\Rightarrow$ $\beta_1 + 16\beta_2 + 256\beta_3 + 18\beta_4 + 324\beta_5 + 288\beta_6$

$-)\ \beta_1 + 17\beta_2 + 289\beta_3 + 8\beta_4 + 64\beta_5 + 136\beta_6$

$$\begin{cases} H_0: -\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 \geq 0 \\ H_a: -\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 < 0 \end{cases}$$

$\alpha = 5\%$

$RR = \{T < -1.67\}$

test statistic: $-1.6461 \notin RR$, do not reject H_0.

g. $$\begin{cases} H_0: -\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 \geq 0 \\ H_a: -\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 < 0 \end{cases}$$

test statistic: -2.062, t_{cr}: -1.646

$\Rightarrow$ reject H_0.

h. Wendy: $\beta_4 + 2\cdot 17\beta_5 + 12\beta_6 = \beta_4 + 34\beta_5 + 12\beta_6$

Jill: $\beta_4 + 2\cdot 11\beta_5 + 16\beta_6 = \beta_4 + 22\beta_5 + 16\beta_6$

$$\Rightarrow \begin{array}{r} \beta_4 + 34\beta_5 + 12\beta_6 \\ -)\ \beta_4 + 22\beta_5 + 16\beta_6 \\ \hline \end{array}$$

$$\begin{cases} H_0: 12\beta_5 - 4\beta_6 = 0 \\ H_a: 12\beta_5 - 4\beta_6 \neq 0 \end{cases}$$

$\alpha = 0.05$

test statistic: -1.0273, t_{cr}: ± 1.962

$\Rightarrow$ do not reject H_0.

(i) Jill: $\beta_4 + 2\cdot x\beta_5 + 16\beta_6 = 0$

$$\Rightarrow x = \frac{\beta_4 + 16\beta_6}{2\beta_5} - 11 = 19.667$$

C.I. $= [15.96, 23.40]$