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

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

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.

$$\begin{cases} H_0: \beta_2 = 0 \\ H_a: \beta_2 \neq 0 \end{cases}$$

$$SE(b_2) = \sqrt{4} = 2$$

$$T_0 = \frac{b_2 - 0}{2} = \frac{3-0}{2} = 1.5 \sim t_{0.05}(63-3) = 2$$

$\because T_0 = 1.5 < 2$

$\therefore$ do not reject H_0. 在95%信心水準下沒有足夠證據顯示β_2不為0

b.

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

$$SE(b_1 + 2b_2) = \sqrt{3 + 4 \times 4 + 2 \times 2 \times (-2)} = \sqrt{11}$$

$$\because T_0 = \frac{b_1 + 2b_2 - 5}{\sqrt{11}} < 2$$

$\therefore$ do not reject H_0. 在95%信心水準下沒有足夠證據

顯示 $\beta_1+\beta_2$ 不為 5

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

$$SE(b_1-b_2+b_3)=\sqrt{3+4+3-2\times(-2)+2\times1}=4$$

$$\because T_0=\frac{b_1-b_2+b_3-4}{4}=-1.5>-t_{0.05}(63-3)=-2$$

$\therefore$ do not reject H_0，沒有足夠證據顯示在 95% 信心水準下 $\beta_1-\beta_2+\beta_3$ 不為 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?

(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 (截距 = 20.8701)

    - If Bill departs exactly at 6:30 AM (**depart = 0**), encounters **0** red lights (**reds = 0**), and waits for **0** trains (**trains = 0**), then his **predicted** travel time is about **20.8701 minutes**.
    - This is the baseline travel time with all other factors set to zero.

- **depart (0.3681)**

    - For each additional **minute** after 6:30 AM that Bill departs, **holding reds and trains constant**, the travel time **increases by about 0.3681 minutes** on average.
    - For example, departing 10 minutes later than 6:30 AM (i.e., 6:40 AM) would add roughly $10 \times 0.3681 \approx 3.6810$ minutes to the total commute.

- **reds (1.5219)**

    - For each **additional red light** Bill encounters, **holding depart time and trains constant**, the travel time **increases by about 1.5219 minutes** on average.
    - So if Bill encounters 2 more red lights than usual, that would add roughly $2 \times 1.5219 = 3.042$ minutes to his commute.

- **trains (3.0237)**

    - For each **train** Bill must wait for, **holding depart time and reds constant**, the travel time **increases by about 3.0237 minutes** on average.
    - If Bill waits for 2 trains, that adds about $2 \times 3.02 = 6.042$ minutes compared to no trains.

(b)

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

1. **Interpretation of "Precision"**
    - A narrower confidence interval implies a more precise estimate (less uncertainty).
    - A wider interval implies less precision (more uncertainty).
2. **depart** (CI: 0.299 – 0.437)

   - This interval is quite narrow (about 0.14 in width).
   - It does not include zero, which strongly suggests the **depart** effect is both statistically significant and relatively precisely estimated.
3. **reds** (CI: 1.157 – 1.886)
   - Also fairly narrow (about 0.73 in width) and well above zero.
   - Indicates a statistically significant and reasonably precise estimate.
4. **trains** (CI: 1.775 – 4.273)
   - Wider than the intervals for **depart** or **reds** (about 2.50 in width).
   - Still does not include zero, so it is significant.
   - The estimate is less precise than **depart** or **reds**, but we still have a meaningful range (somewhere between roughly 1.8 and 4.3 minutes per train).
5. **(Intercept)** (CI: 17.57 – 24.17)
   - A span of about 6.6 minutes, which is reasonably wide.
   - Intercept estimates often have larger uncertainty, because they represent the baseline (when all other variables are zero).
   - Despite the width, the interval does not approach zero, indicating a clearly positive baseline time.

Conclusion

- All four coefficients are **significantly different from zero** (none of the intervals cross zero).
- **depart** and **reds** have relatively **narrow intervals**, implying more precise estimates.
- **trains** and the **intercept** have somewhat wider intervals, indicating less precision but still providing clear evidence that their effects are positive.

Hence, you do have **reasonably precise estimates** overall, though some coefficients (like **trains** and the **intercept**) have more uncertainty compared to others.

(c)

$$H_0 : \beta_{\text{reds}} \geq 2 \quad \text{vs.} \quad H_1 : \beta_{\text{reds}} < 2$$

Given $\hat{\beta}_{\text{reds}} = 1.5219 \text{ and } \mathrm{SE} = 1.96.$ bc large numbers of sample

So we get t-statistic = -0.2439, one-sided p-value = 0.4037, which is less than alpha, as a result, we do **not** reject the null hypothesis. This indicates that there is insufficient evidence at the 5% significance level to conclude that the expected delay differs from (or is in a particular direction relative to) 2 minutes.

(d)

$$H_0: \beta_{\text{trains}} = 3 \quad \text{vs.} \quad H_1: \beta_{\text{trains}} \neq 3.$$

Explanation

1. **t-Statistic Calculation:**

   The test statistic is calculated as $t = \frac{b4 - \beta_0}{\text{SE}} = \frac{3.0237 - 3}{0.6244}.$

2. **p-Value Calculation:**

   For a two-sided test, the p-value is calculated as $\text{p-value} = 2 \times P(T \leq -|t|).$

3. **Decision Rule:**

   Compare the p-value with α=0.10. If p-value<0.10, reject H0; otherwise, do not reject H0.

t-statistic = 0.03795644. Two-sided p-value = 0.9697533

Result: Do not reject the null hypothesis at the 10% significance level.

(e)

```
Call:
lm(formula = time ~ depart + reds + trains + after730, data = commute5)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.234  -3.706  -0.066   4.512  16.776 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 21.08865    1.71206  12.318  < 2e-16 ***
depart       0.35760    0.03879   9.218  < 2e-16 ***
reds         1.52535    0.18534   8.230 1.13e-14 ***
trains       2.96157    0.64213   4.612 6.44e-06 ***
after730     2.16146    3.37172   0.641    0.522    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.306 on 244 degrees of freedom
Multiple R-squared:  0.5353,	Adjusted R-squared:  0.5277 
F-statistic: 70.28 on 4 and 244 DF,  p-value: < 2.2e-16
```

t_stat = -2.324794

p_value = 0.01045213

Result: p_value < 0.05, we reject H0 and conclude that the extra delay for leaving after 7:30 AM is significantly less than 10 minutes.

(f)

$$H_0 : \beta_{\text{trains}} \geq 3\beta_{\text{reds}} \quad \text{versus} \quad H_1 : \beta_{\text{trains}} < 3\beta_{\text{reds}},$$

Let $L = \beta_{\text{trains}} - 3\,\beta_{\text{reds}}.$

Then

- Under H_0: $L \geq 0$,
- Under H_1: $L < 0$.

Explanation

1. **Linear Combination L:**
   We form $L=\beta_{trains}-3\,\beta_{reds}$. Under the null, we expect $L\geq 0$ (i.e., the train delay is at least 3 times the red light delay).

2. **Standard Error of L:**
   Using the variance–covariance matrix, the variance of L is computed by

$$\text{Var}(L) = \text{Var}(\beta_{\text{trains}}) + 9\ \text{Var}(\beta_{\text{reds}}) - 6\ \text{Cov}(\beta_{\text{trains}}, \beta_{\text{reds}}).$$

Its square root gives the standard error.

3. **t-Statistic and p-Value:**
   The t-statistic is $t = L_{\text{hat}}/se_L$. Because the alternative is $L<0$, the one-sided p-value is computed using the lower tail of the t-distribution.

4. **Decision Rule:**

At $\alpha=0.05$, if the p-value is less than 0.05 we reject H_0; otherwise, we do not reject H_0.

Results:

Estimated L (beta_trains - 3 * beta_reds) = -1.542133

Standard error = 0.844992

t-statistic = -1.825027

One-sided p-value = 0.03460731

Reject H_0 : There is evidence that the expected train delay is less than 3 times the red light delay.

(g)

$$\text{time} = \beta_1 + \beta_2\,\text{depart} + \beta_3\,\text{reds} + \beta_4\,\text{trains} + e$$

已知 depart = 30, reds = 6, trains = 1

The hypothesis we want to test is that leaving at 7:00 AM will result in Bill arriving on or before 7:45 AM. Since 7:45 AM is 45 minutes after 7:00 AM, his travel time should be at most 45 minutes. In terms of the expected travel time E(time|X)E(\text{time}|X)E(time|X) given X=(depart=30, reds=6, trains=1), the hypotheses can be written as:

- H_0: $E(\text{time}|X) \leq 45$ (i.e. 7:00 AM is early enough)
- H_1: $E(\text{time}|X) > 45$ (i.e. leaving at 7:00 AM is not early enough)

Step 1. Obtain the Predicted Mean and Its Standard Error

Step 2. Compute the Standard Error of the Predicted Mean

$$L_{\text{hat}} \pm t_{0.975,\,df} \times SE_{\text{pred}}$$

Step 3. Set Up and Perform the Hypothesis Test

$$t = \frac{L_{\text{hat}} - 45}{SE_L}.$$

Estimated mean travel time = 44.06924

Standard error = 0.5392687

t-statistic = -1.725964

One-sided p-value = 0.9571926

Step 4. Decision at α=0.05

Do not reject H0: There is insufficient evidence to conclude that leaving at 7:00 AM is not early enough (expected travel time <= 45 minutes).

(h)

The original hypothesis setup is correct because it is framed around the primary concern: ensuring Bill is not late.

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
```

查看 summary(model) 輸出中的 Pr(>|t|)，若小於 0.01、0.05、或 0.1，分別代表在 1%、5%、10% 顯著水準下顯著不為 0。

(b)

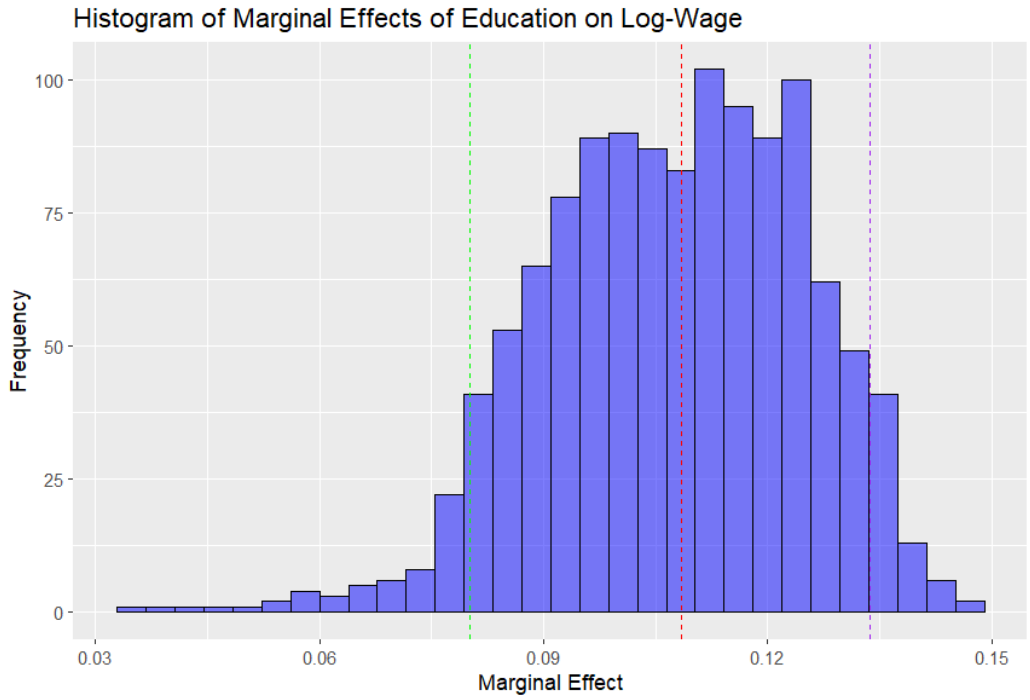
Marginal Effect at EDUC = 12 and EXPER = 5 : 0.1194868

Marginal Effect at EDUC = 12 and EXPER = 20 : 0.1043332

Marginal Effect at EDUC = 16 and EXPER = 5 : 0.1311531

Marginal Effect at EDUC = 16 and EXPER = 20 : 0.1159995

(c)



Median Marginal Effect: 0.1084313

5th Percentile: 0.08008187

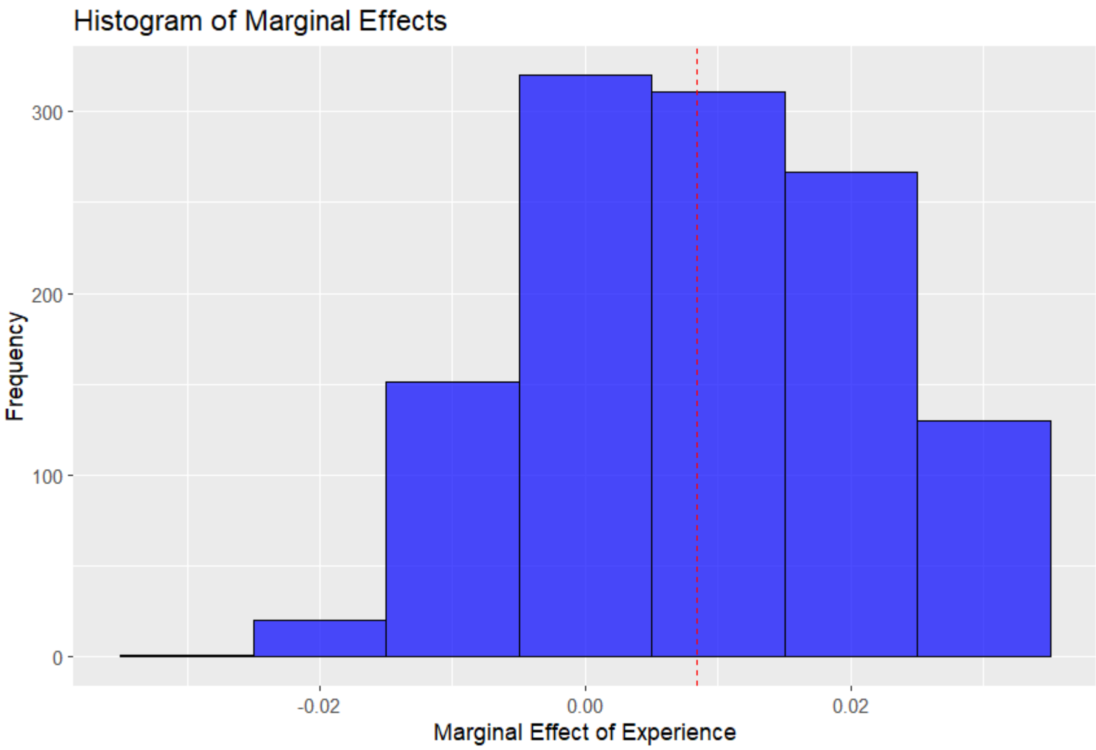
95th Percentile: 0.1336188

(d)

微分後:

$\beta 4+2\beta 5*EXPER+\beta 6*EDUC$

(e)




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

(f)

Difference in log-wages: -0.03588456

Standard error of difference: 0.02148902

t-statistic: -1.669902

p-value: 0.9523996

Fail to reject H0: No significant evidence that David's expected log-wage is greater than Svetlana's.

(g)

New difference in log-wages: 0.03091716

New standard error of difference: 0.01499112

New t-statistic: 2.062365

New p-value: 0.01969445

Reject H0: David's expected log-wage is still significantly greater than Svetlana's.

(h)

Estimated difference in marginal effects: -0.001575327

Standard error of the difference: 0.001533457

t-statistic: -1.027304

p-value: 0.3044854

Fail to reject H0: No significant difference in their marginal effects of experience.

(i)

Jill's estimated years of experience when marginal effect becomes negative: 30.67706

95% Confidence Interval: (15.70815 , 45.64596)