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

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

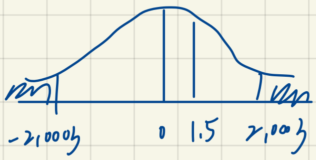
Using a 5% significance level, and an alternative hypothesis that the equality does not hold, test each of the following null hypotheses:

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

a. $H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$$t = \frac{b_k - \beta_k}{se(b_k)} \sim t(N-K)$$



$$t = \frac{3-0}{\sqrt{4}} = \frac{3}{2} \sim t(60) = 1.5 < 2.0003$$

$t(60, 0.05) = 2.0003$ ⇒ 不拒絕 H_0

b. $H_0: \beta_1 + 2\beta_2 = 5$

$H_1: \beta_1 + 2\beta_2 \neq 5$

$$t = \frac{(2+2\times 3)-5}{\sqrt{11}} \approx 0.9045 < 2.0003$$

⇒ 不拒絕 H_0

$$se(b_1 + 2b_2) = \sqrt{var(b_1 + 2b_2)}$$

$$\widehat{var}(b_1 + 2b_2 \mid X) = \widehat{var}(b_1) + 4\,\widehat{var}(b_2) + 2\cdot 2\,\widehat{cov}(b_1, b_2)$$

$$= 3 + 4\times 4 + 2\times 2\times(-2)$$

$$= 19 - 8 = 11$$

c. $H_0: \beta_1 - \beta_2 + \beta_3 = 4$

$H_1: \beta_1 - \beta_2 + \beta_3 \neq 4$

$$t = \frac{(2-3-1)-4}{\sqrt{16}} = \frac{-6}{4} = -1.5 < -2.0003$$

$$se(b_1 - b_2 + b_3) = \sqrt{var(b_1 - b_2 + b_3)}$$

⇒ 不拒絕 H_0

$$\widehat{var}(b_1 - b_2 + b_3) = \widehat{var}(b_1) + \widehat{var}(b_2) + \widehat{var}(b_3) - 2\,cov(b_1, b_2) + 2\,cov(b_1, b_3) - 2\,cov(b_2, b_3)$$

$$= 3 + 4 + 3 - 2\cdot(-2) + 2\cdot 1 - 2\cdot 0$$

$$= 16$$

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?

(adj) 精确的

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.) 6:30→7:00 →30min 7:00→7:30 →30min，会晚10min

至少

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. $\beta_4 \geq 3\beta_3$

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.

```
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
```

Holding other variables constant, for every minute Bill departs later than 6:30, Time increase by 0.3681 mintes.

Holding other variables constant, one more red light that he encounters, Time increase by 1.5219 minutes

Holding other variables constant, one more train he hase to wait, Time increase by 3.0237 minutes

If he depart at 6:30, doesn't encounter any red light and doesn't have to wait any train, he will spend 20.8701 minutes to drive to work.

b.

```
> #5.31.b
> confint(table, 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
```

Yes, we have obtained precise estimates of each of the coefficients, however the trains has a relatively wide confidence interval, suggesting greater uncertain in its estimate.

c. $\alpha = 0.05$

左尾 $H_0: \beta_3 \geq 2$

$H_1: \beta_3 < 2$

$$t = \frac{1.5219 - 2}{0.185} = -2.5893 < -1.6511$$

⇒ 不拒絕 H_0.

d. $\alpha = 0.01$ 双尾

$H_0: \beta_4 = 3$

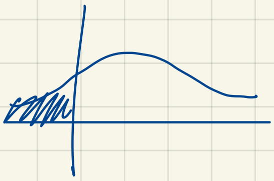
$H_1: \beta_4 \neq 3$

$$t = \frac{3.0237 - 3}{0.634} = 0.0374 < 1.6511$$

⇒ 不拒絕 H_0

e. β_2 ⇒ 6:30A.M ⇒ DEPART 左尾

$\alpha = 0.05$ 60-30 { 7:00 ⇒ 30min ; 7:30 ⇒ 60

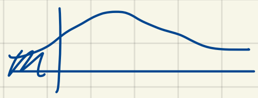


$H_0: 60\beta_2 - 30\beta_2 \geq 10 \Rightarrow \beta_2 \geq \frac{1}{3}$

$H_0: 60\beta_2 - 30\beta_2 < 10 \Rightarrow \beta_2 < \frac{1}{3}$

$$t = \frac{0.3681 - \frac{1}{3}}{0.0351} = 0.9905 > -1.6511$$

⇒ 不拒絕 H_0.



-1.6511

f. $H_0: \beta_4 \geq 3\beta_3 \Rightarrow \beta_4 - 3\beta_3 \geq 0$ 左尾

$H_1: \beta_4 < 3\beta_3 \Rightarrow \beta_4 - 3\beta_3 < 0$

$$t = \frac{3.0237 - 3 \times 1.5219}{0.844992} = -1.8249 < -1.6511$$

⇒ 拒絕 H_0, H_1 為真

7:00 Depart 左尾.

g. $H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 = 45$

$H_1: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 < 45$

$$t = \frac{44.062924 - 45}{0.5392687} = -1.726 < -1.6511$$

⇒ reject H_0, H_1 為真

h. $H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 = 45$ $t = -1.726 < -1.6511$

$H_1: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \le 45$

reject H_0 => so Bill will arrive university on time.

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
```

All coefficient estimates are significantly different from zero.

at the 1% level of significance, expect from $EDUC^2$ event at 10% level

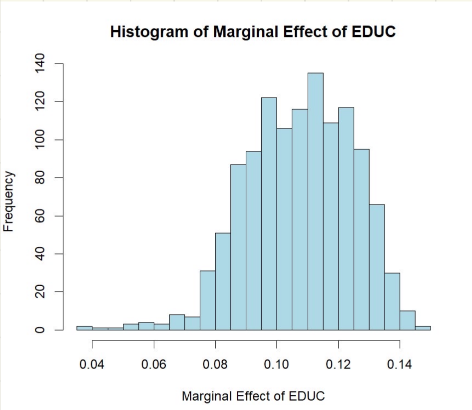
b. $\dfrac{\partial E[\ln(WAGE) \mid EDUC, EXPER]}{\partial EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$

$= 0.08954 + 2 \times 0.001458\, EDUC + (-0.00101) \times EXPER$

EDUC↑ → marginal effect ↑

EXPER↑ → marginal effect ↓

(c)



5%	50%	95%
0.08008187	0.10843125	0.13361880

The distribution is approximately symmetric and bell-shaped

The effects are positive throughout, meaning more education consistently increases expected log wage for all individuals

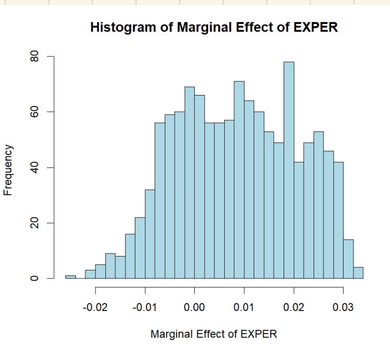
(d) $\dfrac{\partial E[\ln(WAGE) \mid EDUC, EXPER]}{\partial EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$

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

EXER↑ → marginal effect ↓

EDUC↑ → marginal effect ↓

(e)



5%	50%	95%
-0.010376212	0.008418878	0.027931151

Most marginal effects are positive, only a small portion may experience negative

(f) $H_0: \beta_1+17\beta_2+17^2\beta_3+8\beta_4+8^2\beta_5+8\times17\beta_6 \le \beta_1+16\beta_2+16^2\beta_3+18\beta_4+18^2\beta_5+18\times16\beta_6$

$H_1: \beta_1+17\beta_2+17^2\beta_3+8\beta_4+8^2\beta_5+8\times17\beta_6 > \beta_1+16\beta_2+16^2\beta_3+18\beta_4+18^2\beta_5+18\times16\beta_6$

$\Rightarrow H_0: \beta_2+33\beta_3-10\beta_4-260\beta_5-152\beta_6 \le 0$ →右尾

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

$t=\dfrac{-0.03588-0}{0.021489}=-1.6697<1.646$

$\Rightarrow$ 不拒絕 H_0, H_1為真

(g) $H_0: \beta_1+17\beta_2+17^2\beta_3+16\beta_4+16^2\beta_5+16\times17\beta_6 \le \beta_1+16\beta_2+16^2\beta_3+26\beta_5+26\times16\beta_6$

$H_1: \beta_1+17\beta_2+17^2\beta_3+16\beta_4+16^2\beta_5+16\times17\beta_6 > \beta_1+16\beta_2+16^2\beta_3+26\beta_5+26\times16\beta_6$

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

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

$t=\dfrac{0.03092-0}{0.01499}=2.0267>1.646$

$\Rightarrow$ 拒絕 H_0, H_1為真

(h) marginal effect $=\beta_4+2\beta_5 EXPER+\beta_6 EDUC$

$H_0: \beta_4+2\times17\beta_5+12\beta_6=\beta_4+2\times11\beta_5+16\beta_6$

$H_1: \beta_4+2\times17\beta_5+12\beta_6\ne\beta_4+2\times11\beta_5+16\beta_6$

$\Rightarrow H_0: 12\beta_5-4\beta_6=0$

$H_1: 12\beta_5-4\beta_6\ne 0$

$t=\dfrac{-0.001575}{0.0015334}=-1.02713>-1.962$

$\Rightarrow$ 不拒絕 H_0

(i) $\beta_4+2x\beta_5\times(11+x)+16\beta_6=0$

$x=\dfrac{-\beta_4-16\beta_6}{2\beta_5}-11=19.67706$

$se=\sqrt{\left(\frac{-1}{2\beta_5}\right)^2 var(\beta_4)+\left(\frac{\beta_4+16\beta_6}{2\beta_5^2}\right)^2 var(\beta_5)+\left(\frac{-8}{\beta_5}\right)^2 var(\beta_6)+2\left(\frac{-1}{2\beta_5}\right)\left(\frac{\beta_4+16\beta_6}{2\beta_5^2}\right)cov(\beta_4,\beta_5)+2\times\left(\frac{\beta_4+16\beta_6}{2\beta_5^2}\right)\left(\frac{-8}{\beta_5}\right)cov(\beta_2,\beta_3)+2\times\left(\frac{-1}{2\beta_5}\right)\left(\frac{-8}{\beta_5}\right)cov(\beta_1,\beta_3)}=1.8957$

confidence interval $\Rightarrow 19.6771 \pm t\times1.8957 \Rightarrow [15.958, 23.396]$