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. $\beta_2 = 0$

$$\begin{cases} H_0: \beta_2 = 0 \\ H_a: \beta_2 \neq 0 \end{cases} \quad t = \frac{3-0}{\sqrt{4}} = 1.5 < 2 = t_{0.975,60}$$

not to reject $H_0: \beta_2 = 0$

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

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

$$\begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = 8$$

$$Var(\beta_1 + 2\beta_2) = \begin{bmatrix} 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & 6 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = 11$$

$$t = \frac{8-5}{\sqrt{11}} = 0.905 < 2$$ not to reject $H_0: \beta_1 + 2\beta_2 = 5$

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

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

$$\begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = -2$$

$$Var(\beta_1 - \beta_2 + \beta_3) = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 & 1 \\ -2 & 4 & 0 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 6 & -6 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = 16$$

$$t = \frac{-2-4}{\sqrt{16}} = -1.5 > -2$$ not to reject $H_0: \beta_1 - \beta_2 + \beta_3 = 4$

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.

$\widehat{TIME} = 20.8701 + 0.3681\ DEPART + 1.5219\ REDS + 3.0237\ TRAINS$

```
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 ***
```

If one more minute he depart later than 6:30, TIME increases by 0.3681 minutes, holding other factors constant.
If he encounters one more red light, TIME increases by 1.5219 minutes, holding other factors constant.
If he has to wait one more train, TIME increases by 3.0237 minutes, holding other factors constant.
If he depart at 6:30, does not encounter any red lights and have to wait any trains, he will take 20.8701 minutes to drive to work.

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

```
                 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 are relatively narrow ones, we have 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.

$\begin{cases} H_0: \beta_3 \geq 2 \\ H_a: \beta_3 < 2 \end{cases}$ $t = \dfrac{1.5219 - 2}{0.1850} = -2.5843 < t_{(0.05, 245)} = -1.6511$

Reject the time expected delay from each red light is 2 minutes or more.

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.

$$\begin{cases} H_0: \beta_4 = 3 \\ H_a: \beta_4 \neq 3 \end{cases} \qquad t = \frac{3.0237 - 3}{0.6340} = 0.0374 < t_{(0.95,245)} = 1.6511$$

There is no enough evidence that we can reject H_0
⇒ The time expected delay from each train is 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.)

$$\begin{cases} H_0: 60\beta_2 - 30\beta_2 \geq 10 \\ H_a: 60\beta_2 - 30\beta_2 < 10 \end{cases} \Rightarrow \begin{cases} H_0: \beta_2 \geq \frac{1}{3} \\ H_a: \beta_2 < \frac{1}{3} \end{cases} \quad t = \frac{0.3681 - \frac{1}{3}}{0.0351} = 0.99057 \quad t_{(0.05,245)} = -1.6511$$

There is no enough evidence that we can reject H_0
⇒ The time expected delay at least 10 minutes longer if he leaves more later 30 minutes.

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.

$$\begin{cases} H_0: \beta_4 \geq 3\beta_3 \\ H_0: \beta_4 < 3\beta_3 \end{cases} \qquad t = \frac{\beta_4 - 3\beta_3}{se(\beta_4 - 3\beta_3)} = \frac{-1.542}{0.844992} = -1.8249 < t_{(0.05,245)} = -1.6511$$

Reject $H_0: \beta_4 \geq 3\beta_2$. We can conclude that the expected delay from a train is less than three times greater than that from a red light.

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.]

$$\begin{cases} H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \leq 45 \\ H_a: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 > 45 \end{cases} \qquad E(TIME|X) = 44.06924$$

$$t = \frac{44.06924 - 45}{0.5392687} = -1.726 < t_{(0.95,245)} = 1.6511$$

There is no enough evidence that we can reject H_0
⇒ He encounters six red lights and one train, leaving Carnegie at 7 AM is early enough to get him to the university on or before 7:45 AM

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?

$$\begin{cases} H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \geq 45 \\ H_a: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 < 45 \end{cases}$$ (We always want to reject H_0)

In this case, $-1.726 < t_{0.05,245} = -1.6511$, we can reject H_0 that he can expect 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"?

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)   β1  1.038e+00  2.757e-01   3.764 0.000175 ***
educ          β2  8.954e-02  3.108e-02   2.881 0.004038 ** 
I(educ^2)     β3  1.458e-03  9.242e-04   1.578 0.114855    
exper         β4  4.488e-02  7.297e-03   6.150 1.06e-09 ***
I(exper^2)    β5 -4.680e-04  7.601e-05  -6.157 1.01e-09 ***
I(educ * exper) β6 -1.010e-03  3.791e-04  -2.665 0.007803 ** 
```

All coeffcient estimates are different from zero at 1% level, except for $EDUC^2$, it's different from zero at 12% level.

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.

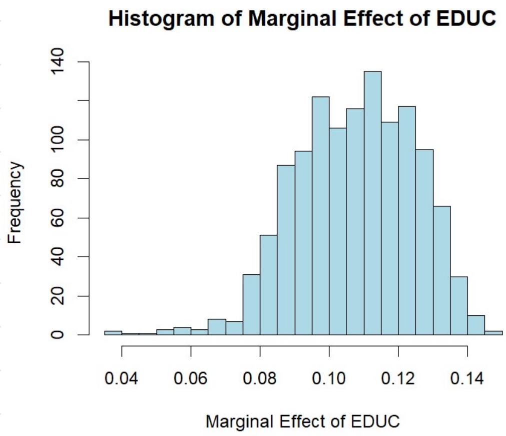
$$\frac{\partial E[\ln(WAGE)|EDUC, EXPER]}{\partial EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$$

$$= 0.08954 + 2 \times 0.001458\, EDUC + (-0.001010) EXPER$$

The marginal effect of education increases as the level of education increases, but decreases with the level of experience.

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.

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



I observed most of the marginal effects falls on [0.08, 0.14]

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.

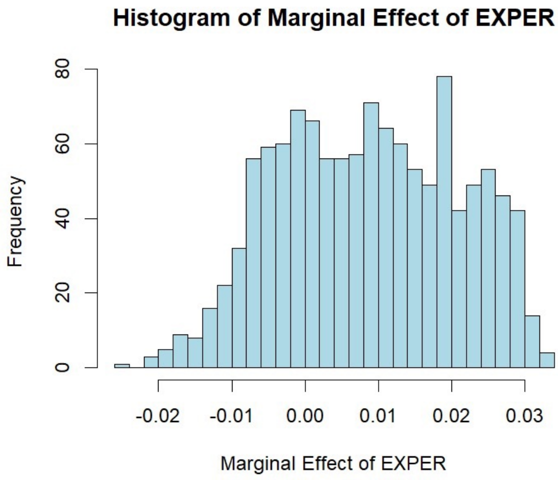
$$\frac{\partial E[\ln(WAGE)|EDUC, EXPER]}{\partial EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$$

$$= 0.04488 + 2 \times (-0.000468) EXPER + (-0.001010) EDUC$$

The marginal effect of experience decreases as the level of education and experience 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.

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



There is a proportion is negative.

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.

$$\begin{cases} H_0: \beta_1 + 17\beta_2 + 17^2\beta_3 + 8\beta_4 + 8^2\beta_5 + 8\times17\beta_6 \\ \quad \leq \beta_1 + 16\beta_2 + 16^2\beta_3 + 18\beta_4 + 18^2\beta_5 + 18\times16\beta_6 \\ H_a: \beta_1 + 17\beta_2 + 17^2\beta_3 + 8\beta_4 + 8^2\beta_5 + 8\times17\beta_6 \\ \quad > \beta_1 + 16\beta_2 + 16^2\beta_3 + 18\beta_4 + 18^2\beta_5 + 18\times16\beta_6 \end{cases}$$

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

$$t = \frac{-0.03588456 - 0}{0.02148902} = -1.6699 < t_{(0.95,1194)} = 1.646$$

we don't have enough evidence to reject H_0.

We can conclude that Svetlana's expected log-wage is equal to or greater than David's expected log-wage.

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?

$$\begin{cases} H_0: \beta_1 + 17\beta_2 + 17^2\beta_3 + 16\beta_4 + 16^2\beta_5 + 16\times17\beta_6 \\ \quad \leq \beta_1 + 16\beta_2 + 16^2\beta_3 + 26\beta_4 + 26^2\beta_5 + 26\times16\beta_6 \\ H_a: \beta_1 + 17\beta_2 + 17^2\beta_3 + 16\beta_4 + 16^2\beta_5 + 16\times17\beta_6 \\ \quad > \beta_1 + 16\beta_2 + 16^2\beta_3 + 26\beta_4 + 26^2\beta_5 + 26\times16\beta_6 \end{cases}$$

$$\Rightarrow \begin{cases} H_0: \beta_2 + 33\beta_3 - 10\beta_4 - 420\beta_5 - 144\beta_6 \leq 0 \\ H_a: \beta_2 + 33\beta_3 - 10\beta_4 - 420\beta_5 - 144\beta_6 > 0 \end{cases}$$

$$t = \frac{0.03091716 - 0}{0.01499112} = 2.06247 > t_{(0.95, 1194)} = 1.646$$

We can reject H_0, conclude that Svetlana's expected log-wage is not equal to or greater than David's expected log-wage after eight years.

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.

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

$$\begin{cases} H_0: \beta_4 + 2\times 17\beta_5 + 12\beta_6 = \beta_4 + 2\times 11\beta_5 + 16\beta_6 \\ H_a: \beta_4 + 2\times 17\beta_5 + 12\beta_6 \neq \beta_4 + 2\times 11\beta_5 + 16\beta_6 \end{cases}$$

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

$$t = \frac{-0.001575327 - 0}{0.001533457} = -1.0273 > -1.962 = t_{(0.025, 1194)}$$

We do not have enough evidence to reject H_0, at 5% significant level.
We can conclude that their marginal effects of extra experience are equal.

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.

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

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

$$se\left(\frac{-\beta_4 - 16\beta_6}{2\beta_5} - 11\right) = se\left(\frac{-\beta_4 - 16\beta_6}{2\beta_5}\right)$$

$$= \sqrt{\left(-\frac{1}{2\beta_5}\right)^2 Var(\beta_4) + \left(\frac{\beta_4 + 16\beta_6}{2\beta_5^2}\right)^2 var(\beta_5) + \left(\frac{-16}{2\beta_5}\right)^2 var(\beta_6) + 2\times\left(-\frac{1}{2\beta_5}\right)\left(\frac{\beta_4+16\beta_6}{2\beta_5^2}\right) cov(\beta_4, \beta_5) + 2\left(-\frac{1}{2\beta_5}\right)\left(\frac{-16}{2\beta_5}\right) cov(\beta_4, \beta_6) + 2\times\left(\frac{-16}{2\beta_5}\right)\left(\frac{\beta_4+16\beta_6}{2\beta_5^2}\right) cov(\beta_5, \beta_6)} = 1.895713$$

$$19.67706 \pm t_{(0.025, 1194)} \cdot 1.895713 = [15.9578, 23.3964]$$