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.

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

$\alpha = 0.05$

test statistic: $\dfrac{b_2 - 0}{SE(b_2)} \sim t(60)$

$$t^* = \frac{3-0}{\sqrt{4}} = 1.5 < 2 = t_{0.975}(60)$$

don't reject H_0

b.

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

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

$\alpha = 0.05$

test statistic: $\dfrac{(b_1 + 2b_2) - 5}{SE(b_1 + 2b_2)} \sim t(60)$

$$t^* = \frac{2 + 2\times 3 - 5}{\sqrt{Var(b_1 + 2b_2)}} = \frac{3}{\sqrt{3 + 2^2 \times 4 + 2\times 2(-2)}} = \frac{3}{\sqrt{11}} = 0.905 < 2 = t_{0.95}(60)$$

don't reject H_0

c.

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

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

$\alpha = 0.05$

test statistic: $\dfrac{b_1 - b_2 + b_3 - 4}{SE(b_1 - b_2 + b_3)} \sim t(60)$

$$t^* = \frac{2 - 3 + (-1) - 4}{\sqrt{3^2 + 4^2 + 3^2 - 2\times(-2) + 2\times 1 - 2\times 0}} = -0.95 > -2 = t_{0.025}(60)$$

don't reject H_0

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.

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

β_1: Bill's expected commute time when he leavse Cornegie at 6:30 AM and encounters no red lights and no trains is estimated to be 20.87 minutes.

β_2: Assuming the number of red lights and train are constant. If Bill leaves later than 6:30 AM, the increase in his expected traveling time is estimated to be 3.681 minutes for every 10 minutes that his departure time is later than 6:30 AM.

β_3: The expected increase in traveling time form each red lights, with departure time and number of trains held constant, is estimated to be 1.5219 minutes.

β_4: The expected increase in traveling time from each train, with departure time and the number of red lights held constant, is estimated to be 3.0237 minutes.

```
> summary(model)

Call:
lm(formula = time ~ ., 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
```

b.

β_1 : $[17.5694,\ 24.1709]$

β_2 : $[0.2990,\ 0.4373]$

β_3 : $[1.1575,\ 1.8864]$

β_4 : $[1.7749,\ 4.2725]$

In the context of driving time, these intervals are relatively narrow ones. We have obtained precise estimates of each of the coefficient.

c.

$H_0: \beta_3 \geq 2$

$H_1: \beta_3 < 2$

$\alpha = 0.05$

test statistic: $\dfrac{b_3 - 2}{SE(b_3)} \sim t(245)$

$t^* = \dfrac{1.5219 - 2}{0.185} = -2.58 < -1.65 = t_{0.05}(245)$

$\therefore$ reject H_0

We conclude that the expected delay from each red lights is less than 2 minutes.

d.

$H_0: \beta_4 = 3$

$H_1: \beta_4 \neq 3$

$\alpha = 0.1$

test statistic: $\dfrac{b_4 - 3}{SE(b_4)} \sim t(245)$

$t^* = \dfrac{3.0237 - 3}{0.634} = 0.037 < 1.65 = t_{1-0.1/2}(245)$

$\therefore$ don't reject

We can't conclude that the expected delay from train is different with 3 minutes.

e.

$H_0: \beta_2 \geq \frac{1}{3}$

$H_1: \beta_2 < \frac{1}{3}$

$\alpha = 0.05$

test statistic: $\dfrac{b_2 - \frac{1}{3}}{SE(b_2)} \sim t(245)$

$t^* = \dfrac{0.3681 - \frac{1}{3}}{0.0351} = 0.99 > -1.65 = t_{0.05}(245)$

$\therefore$ don't reject H_0

The data are consistent with the hypothesis that delaying departure time by 30 minutes increases expected traveling time by at least 10 minutes.

f.

$H_0: \beta_4 \geq 3\beta_3$

$H_1: \beta_4 < 3\beta_3$

$\alpha = 0.05$

$$\text{test statistic}: \frac{b_4 - 3b_3 - 0}{SE(b_4 - 3b_3)} \sim t(245)$$

$$SE(b_4 - 3b_3) = \sqrt{0.634^2 + 3^2 \times 0.185^2 - 2 \times 3 \times (-0.0006)} = 0.84$$

$$t^* = \frac{3.0237 - 3 \times 1.5219 - 0}{0.84} = -1.84 < -1.65 = t_{0.05}(245)$$

$\therefore$ reject H_0

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

g.

$H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \leq 45$

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

$\alpha = 0.05$

$$\text{test statistic}: \frac{b_1 + 30b_2 + 6b_3 + b_4 - 45}{SE(b_1 + 30b_2 + 6b_3 + b_4)}$$

$$SE(b_1 + 30b_2 + 6b_3 + b_4) = 0.539$$

$$t^* = \frac{20.8701 + 30 \times 0.3681 + 6 \times 1.5219 + 3.0237 - 45}{0.338}$$

$$= -1.73 < 1.65 \quad t_{0.95}(245)$$

$\therefore$ don't reject H_0

There is no insufficient evidence to conclude that Bill will get to Univerity after 7:45 AM

h.

If it is imperative that Bill is not late for his meeting, he will wish to establish, with a high degree of probability, that his commute time will be less than 45 minutes. To do so, having a commute time less than 45 minutes should be alternative hypothesis. Having it as the null hypothesis, and failing to reject the null hypothesis, doesn't imply the commute time will necessarily be less than 45 minutes. If hypotheses are reserved as

$H_0: \beta_1 + 30\beta_2 + 6\beta_3 + \beta_4 \geq 45$

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

$$t^* = -1.73 < -1.65 = t_{0.05}(245)$$

In this case, we reject H_0

Bill's expected commute time is such 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"?

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 1% level of significance with the expection of that for $EDUC^2$ which is significant at a 12% significance level.

b.

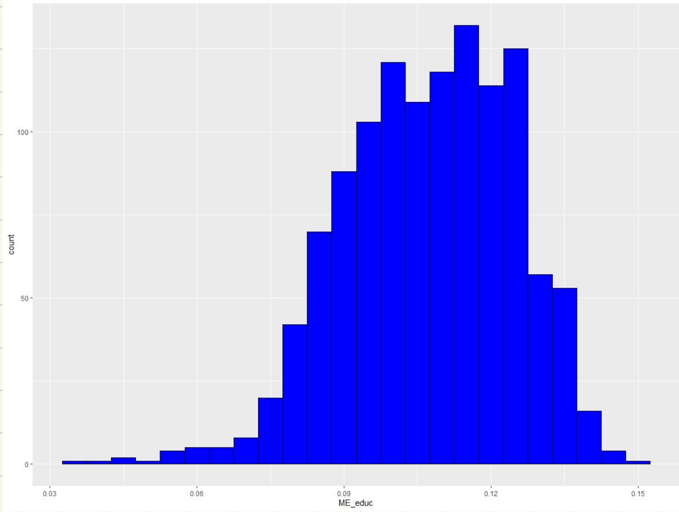
$$\hat{ME}_{EDUC} = \frac{\partial E(\ln(WAGE) \mid EDUC, EXPER)}{\partial EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$$

Its estimate is $\hat{ME}_{EDUC} = 0.0895 + 2 \times 0.00145\, EDUC - 0.00101\, EXPER$

$= 0.0895 + 0.0029\, EDUC - 0.00101\, EXPER$

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

c.




```
    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.03565 0.09513 0.10843 0.10735 0.12050 0.14787
> quantile(cps5_small$ME_educ, probs = c(0.05, 0.95))
        5%        95%
0.08008187 0.13361880
```

We observe that the marginal effects range from 0.036 to 0.148 and the histogram seems to be skewed to the left.

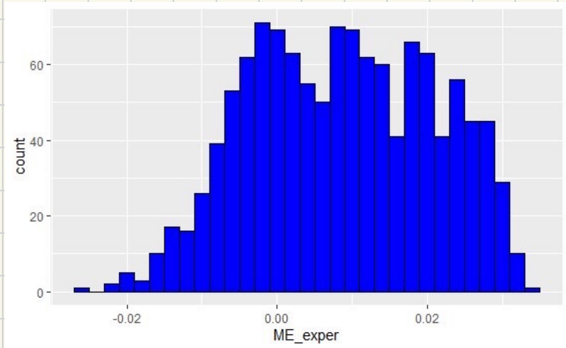
d.

$\hat{ME}_{EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$

$= 0.0449 - 0.00094\, EXPER - 0.00101$

The margin effect of EXPER increase as the level of education decreases, and as the years of experience decreases.

e.




```
> summary(cps5_small$ME_exper)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.025279 -0.001034  0.008419  0.008652  0.018586  0.033989
> quantile(cps5_small$ME_exper, probs = c(0.05, 0.95))
         5%         95%
-0.01037621  0.02793115
```

The values range from -0.025 to 0.0340

$\hat{ME}_{EXPER, 0.05} = -0.0104$

$\hat{ME}_{EXPER, 0.5} = 0.0084$

$\hat{ME}_{EXPER, 0.95} = 0.0279$

f.

$$\begin{array}{r} \beta_1 + 16\beta_2 + 16^2\beta_3 + 18\beta_4 + 18^2\beta_5 + 16\times 18\beta_6 \\ -\ \beta_1 + 17\beta_2 + 17^2\beta_3 + 8\beta_4 + 8^2\beta_5 + 17\times 8\beta_6 \\ \hline -\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 \end{array}$$

$H_0: -\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6 \geq 0$

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

$\alpha = 0.05$

test statistic : $\dfrac{-b_2 - 33b_3 + 10b_4 + 260b_5 + 152b_6 - 0}{SE(-b_2 - 33b_3 + 10b_4 + 260b_5 + 152b_6)} \sim t(1200-6)$

$SE(-b_2 - 33b_3 + 10b_4 + 260b_5 + 152b_6) = 0.0215$

$t^* = \dfrac{0.03588}{0.0215} = 1.669 > -1.646 = t_{0.05}(1194)$

$\therefore$ don't reject H_0

There is insufficient evidence to conclude that David's log-wage is greater.

g.

$$\begin{array}{r} \beta_1 + 16\beta_2 + 16^2\beta_3 + 26\beta_4 + 26^2\beta_5 + 16\times 26\beta_6 \\ -\ \beta_1 + 17\beta_2 + 17^2\beta_3 + 16\beta_4 + 16^2\beta_5 + 17\times 16\beta_6 \\ \hline -\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 \end{array}$$

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

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

$\alpha = 0.05$

test statistic : $\dfrac{-\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6 - 0}{SE(-\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6)} \sim t(1194)$

$t^* = \dfrac{-0.0309}{0.015} = -2.06 < -1.646 = t_{0.05}(1194)$

$\therefore$ reject H_0

We conclude that David's log-wage is greater. This test result is not the same as in part f. The difference in outcomes is attributable to diminishing returns to experience. Because Svetlana initially had 18 years of experience, her extra years of experience had relatively small impact on her log-wage.

h.

$$\widehat{ME}_{EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$$

$$\begin{array}{r} \beta_4 + 34\beta_5 + 12\beta_6 \\ - \quad \beta_4 + 22\beta_5 + 16\beta_6 \\ \hline 12\beta_5 - 4\beta_6 \end{array}$$

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

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

$\alpha = 0.05$

test statistic : $\dfrac{12b_5 - 4b_6 - 0}{SE(12b_5 - 4b_6)} \sim t(1194)$

$t^* = \dfrac{-0.00157}{0.00153} = -1.03 > -1.96 = t_{0.05/2}(1194)$

don't reject H_0

There is no evidence to suggest that the marginal effects from extra experience are different for Jill and Wendy.

i

$$\widehat{ME}_{EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$$

We assume that Jill gains more experience, but no more education.

$b_4 + 2 \times b_5 \times (11 + X) + b_6 \times 16 < 0$

$11 + X > \dfrac{-b_4 - b_6 \times 16}{2 \times b_5}$

$X > 30.677 - 11$

$X > 19.677$

$$var(X) = Var\left(\frac{-b_4 - b_6 \times 16}{2 \times b_5} - 11\right)$$

$$= \left(\frac{-1}{2 \times b_5}\right)^2 Var(b_4) + \left(\frac{b_4 + 16b_6}{2b_5^2}\right)^2 Var(b_5) + \left(\frac{-16}{2b_5}\right)^2 Var(b_6)$$

$$+ 2\left(\frac{-1}{2 \times b_5}\right)\left(\frac{b_4 + 16b_6}{2b_5^2}\right) cov(b_4, b_5) + 2\left(\frac{b_4 + 16b_6}{2b_5}\right)\left(\frac{-16}{2b_5}\right) cov(b_5, b_6)$$

$$+ 2\left(\frac{-1}{2b_5}\right)\left(\frac{-16}{2b_5}\right) cov(b_4, b_6) = 3.59$$

$SE(X) = \sqrt{Var(X)} = 1.9$

95 % interval estimate : $19.677 \pm 1.9\, t_{0.025}(1194)$

$\Rightarrow [15.953, 23.401]$