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. $H_0: \beta_2 = 0$
$H_1: \beta_2 \neq 0$ and $\alpha = 0.05$, the t-statistic $t = \dfrac{3-0}{se\, b_2} = \dfrac{3}{2}$

Since $t_{(0.975, 60)} = 2$ and $\left|\frac{3}{2}\right| < 2 \Rightarrow$ fail to reject H_0.

b. $H_0: \beta_1 + 2\beta_2 = 5$
$H_1: \beta_1 + 2\beta_2 \neq 5$ and $\alpha = 0.05$, the t-statistic $t = \dfrac{b_1 + 2b_2 - 5}{se(b_1 + 2b_2)} = \dfrac{3}{\sqrt{11}}$

Since $\left|\frac{3}{\sqrt{11}}\right| < 2 \Rightarrow$ fail to reject H_0.

C. $H_0: \beta_1 - \beta_2 + \beta_3 = 4$
$H_1: \beta_1 - \beta_2 + \beta_3 \neq 4$ and $\alpha = 0.05$, the t-statistic $t = \dfrac{b_1 - b_2 + b_3 - 4}{se(b_1 - b_2 + b_3)} = \dfrac{-6}{4}$

$= \dfrac{-3}{2}$

Since $\left|\frac{-3}{2}\right| < 2 \Rightarrow$ fail to 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

20.88 0.37 1.53 3.03

$$TIME = \beta_1 + \beta_2 DEPART + \beta_3 REDS + \beta_4 TRAINS + e$$

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? narrow?

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. 1.66

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. Since $\beta_1 = 20.88$, If he departs at 6:30 and encounters neither red light nor train, it will take him 20.88 minute to university.

Since $\beta_2 = 0.37$, If he drives on the same way (the number of red light and train are fixed) everyday, as he delay one minute to depart, the total time for commuting will increase 0.37 minutes.

If he departs at the same time everyday:

Since $\beta_3 = 1.53$, Once he chooses a different way with one more red light than usual, the total time for commuting will increase 1.53 minutes.

If the way he choose has same number of red light but one more trains to wait, Since $\beta_4 = 3.03$, the total time for commuting will increase 3.03 minutes.

The Estimation:

```
> M
            (Intercept)        depart          reds        trains
(Intercept)  2.808171830 -0.0260985055 -0.2690250770  0.0010777876
depart      -0.026098505  0.0012321419  0.0004557753 -0.0104185104
reds        -0.269025077  0.0004557753  0.0342390502 -0.0006481936
trains       0.001077788 -0.0104185104 -0.0006481936  0.4019709090
> summary(theMod)

Call:
lm(formula = time ~ depart + reds + trains, data = theData)

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.

the interval:

$\beta_1: [20.88 - 1.97 \times 1.68,\ 20.88 + 1.97 \times 1.68]$

$\beta_2: [0.37 - 1.97 \times 0.036,\ 0.37 + 1.97 \times 0.036]$

$\beta_3: [1.53 - 1.97 \times 0.19,\ 1.53 + 1.97 \times 0.19]$

$\beta_4: [3.03 - 1.97 \times 0.64,\ 3.03 + 1.97 \times 0.64]$

Since all of their standard error is small and p-value < 0.05, We obtained precise estimation!

c. $H_0: \beta_3 \geq 2$

$H_1: \beta_3 < 2$

and $\alpha = 0.05$, t-statistic $t = \frac{1.53-2}{0.19} = -2.48$

and $t_{(0.05,\,245)} = -1.65$

Since $-2.48 < -1.65$, we reject null hypothesis.

d. $H_0: \beta_4 = 3$

$H_1: \beta_4 \neq 3$

, and $\alpha = 0.1$, the t-statistic $t = \frac{3.03-3}{0.64} = 0.047$.

and $t_{(0.95,245)} = 1.65$

Since $0.047 < 1.65$, we fail to reject null hypothesis.

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

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

and $\alpha = 0.05$, the t-statistic $t = \frac{0.37-\frac{1}{3}}{0.035} = 1.048$

and $t_{(0.05,245)} = -1.65$

Since $1.048 > -1.65$, we fail to reject null hypothesis

f. $H_0: 3\beta_3 - \beta_4 \leq 0$

$H_1: 3\beta_3 - \beta_4 > 0$

and $\alpha = 0.05$, the t-statistic $t = \frac{3 \times 1.53 - 3.03}{se(3\beta_3 - \beta_4)}$

$= \frac{1.56}{0.86} = 1.81$

and $t_{(0.95,245)} = 1.65$

Since $1.81 > 1.65$, we reject null hypothesis

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

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

and $\alpha = 0.05$, the t-statistic $t = \frac{20.88 + 30 \times 0.37 + 6 \times 1.53 + 3.03 - 45}{se(\beta_1 + 30\beta_2 + 6\beta_3 + \beta_4)}$

$= \frac{-0.81}{0.53} = -1.53$

and $t_{(0.95,245)} = 1.65$

Since $-1.53 < 1.65$, we fail to reject null hypothesis.

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. mf1

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. mf2

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. $\left|\dfrac{\beta_1}{se(\beta_1)}\right| = \left|\dfrac{1.038}{0.28}\right| = 3.71$

$\left|\dfrac{\beta_2}{se(\beta_2)}\right| = \left|\dfrac{0.09}{0.032}\right| = 2.82$

$\left|\dfrac{\beta_3}{se(\beta_3)}\right| = \left|\dfrac{0.0015}{0.00093}\right| = 1.62$

$\left|\dfrac{\beta_4}{se(\beta_4)}\right| = \left|\dfrac{0.045}{0.0073}\right| = 6.17$

$\left|\dfrac{\beta_5}{se(\beta_5)}\right| = \left|\dfrac{-0.00047}{0.000077}\right| = 6.11$

$\left|\dfrac{\beta_6}{se(\beta_6)}\right| = \left|\dfrac{-0.001}{0.00038}\right| = 2.64$

```
> summary(theMod)

Call:
lm(formula = log(wage) ~ educ + I(educ^2) + exper + I(exper^2) +
    exper * educ, data = theData)

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 ***
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
```

If all of the coefficients all significant different from zero with α level, then $t_{(1-\frac{\alpha}{2},1196)}$ should be lower than all the data above.

when $\alpha = 0.13$, $t_{(0.945,1196)} = 1.599$. Thus, α should be lower than 0.13.

b.

the margin effect $= \dfrac{\partial \ln WAGE}{\partial EDUC}$

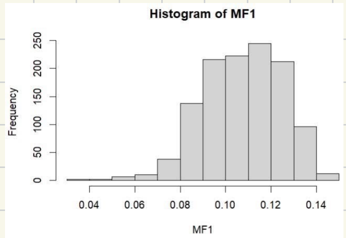
$= \beta_2 + 2\beta_3 EDUC + \beta_6 Exper$

Since $\beta_3 > 0$, as EDUC increase, the margin effect would also increase.

Since $\beta_4 < 0$, as Exper increase, the margin effect would decrease.

c.

According to the picture, the histogram is negative skewness.



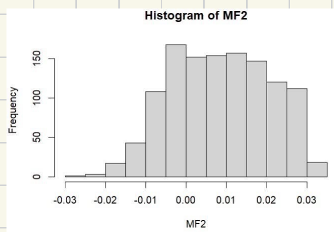

```
> quantile(MF1, c(0.05, 0.95))
        5%        95%
0.08008187 0.13361880
> summary(MF1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.03565 0.09513 0.10843 0.10735 0.12050 0.14787
```

d.

the margin effect $= \dfrac{\partial \ln WAGE}{\partial EXPER} = b_4 + 2b_5 EXPER + b_6 EDUC$

Since both b_5 and b_6 are negative, when either Exper or EDUC increase, the margin effect will decrease.

e. Histogram: the distribution seems more uniform than previous one.




```
> quantile(MF2, c(0.05, 0.95))
         5%         95%
-0.01037621  0.02793115
> summary(MF2)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.025279 -0.001034  0.008419  0.008652  0.018586  0.033989
```

f. H_0: logWAGE of Svetlana − log WAGE of David ≥ 0

H_1: logWAGE of Svetlana − log WAGE of David < 0

, $\alpha = 0.05$, the t-statistic $t = \dfrac{0.036}{se(\text{logWAGE of Svetlana} - \text{log WAGE of David})}$

$= \dfrac{0.036}{se(-\beta_2 - 33\beta_3 + 10\beta_4 + 260\beta_5 + 152\beta_6)}$

$= \dfrac{0.036}{0.022} = 1.64$

and $t_{(0.05,1196)} = -1.65$

Since $-1.65 < 1.64$, we fail to reject H_0.

g. the t-statistic $t = \dfrac{-0.031}{se(-\beta_2 - 33\beta_3 + 10\beta_4 + 420\beta_5 + 144\beta_6)}$

$= \dfrac{-0.031}{0.015}$

$= -2.1$

Since $-1.65 > 2.1$, we reject H_0. It is different from the previous one.

Explain: Since the EDUC does not change, we could treat it as constant. Thus, $\ln WAGE = C + (\beta_4 + \beta_6 EDUC)EXPER + \beta_5 EXPER^2$

According to the equation, the more experience, the less WAGE. (Since $\beta_5 < 0$).

h. $H_0: b_4 + 2b_5 \times 17 + b_6 \times 12 - (b_4 + 2b_5 \times 11 + b_6 \times 16) = 0$, $\alpha = 0.05$

$H_1: 12b_5 - 4b_6 \neq 0$

the t-statistic $t = \dfrac{12b_5 - 4b_6}{se(12b_5 - 4b_6)} = \dfrac{-0.00157}{\sqrt{144 Var b_5 + 16 Var b_6 - 96 cov\, b_5 b_6}}$

$= \dfrac{-0.00157}{0.00153}$

$= -1$

Since $|-1| < 1.97 = t_{(0.975,1196)}$, we fail to reject H_0

1. the margin effect for EXPER of Jill $MF = b_4 + 2b_5 \, EXPER + 16 \times b_6$

$\Rightarrow 0.029 + (-0.00094) EXPER < 0$

$\Rightarrow EXPER > 30.86$

the interval $= [30.86 - 1.97 * se(EXPER), 30.86 + 1.97 * se(EXPER)]$

Calculate $se(EXPER)$:

$$EXPER = \frac{-b_4 - 16b_6}{2b_5}$$

$$\text{the Taylor Expansion} \Rightarrow EXPER = \frac{-b_4^* - 16b_6^*}{2b_5^*} + \frac{-1}{2b_5^*}(b_4 - b_4^*) + \frac{b_4^* + 16b_6^*}{2b_5^{*2}}(b_5 - b_5^*) + \left(\frac{-8}{b_5^*}\right)(b_6 - b_6^*)$$

$$= 30.68 + (-1068.33)(b_4 - 0.045) + 65546.12(b_5 + 0.00047) + 17093.2(b_6 + 0.0011)$$

$$\text{Thus } Var(EXPER) = (1068.33)^2 Var(b_4) + (65546.12)^2 Var(b_5) + (17093.2)^2 Var(b_6)$$

$$+ 2[-1068.33 \times 65546.12 \, cov(b_4, b_5) + 65546.12 \times 17093.2 \times cov(b_5, b_6) + (1068.33 \times 17093.2) cov(b_4, b_6)]$$

$\Rightarrow se(EXPER) = 1.895$

$\Rightarrow$ the interval $= [26.95, 34.42]$

```
wage<-c(cps5_small$wage)
educ<-c(cps5_small$educ)
exper<-c(cps5_small$exper)

#a
theData=data.frame(wage, educ, exper)
theMod<-lm(log(wage)~educ+I(educ^2)+exper+I(exper^2)+exper*educ, data=theData)
summary(theMod)
b1<-coef(theMod)[[1]]
b2<-coef(theMod)[[2]]
b3<-coef(theMod)[[3]]
b4<-coef(theMod)[[4]]
b5<-coef(theMod)[[5]]
b6<-coef(theMod)[[6]]

#b
MF1<-b2+2*b3*educ+b6*exper

#c
hist(MF1)
quantile(MF1, c(0.05, 0.95))
summary(MF1)
#sd(MF1)

#d
MF2<-b4+2*b5*exper+b6*educ

#e
hist(MF2)
quantile(MF2, c(0.05, 0.95))
summary(MF2)
#sd(MF2)
```

```
#f
David<-c(17, 8)
Svet<-c(16, 18)
logWage_David<-b1+b2*(David[1])+b3*(David[1])^2+b4*(David[2])+b5*(David[2])^2+b6*David[1]*David[2]
logWage_Svet<-b1+b2*(Svet[1])+b3*(Svet[1])^2+b4*(Svet[2])+b5*(Svet[2])^2+b6*Svet[1]*Svet[2]
logWage_Svet-logWage_David
M<-vcov(theMod)
sqrt(M[2, 2]+1089*M[3, 3]+100*M[4, 4]+260*260*M[5, 5]+152*152*M[6, 6]
     +2*(33*M[2, 3]+(-10)*M[2, 4]+(-260)*M[2, 5]+(-152)*M[2, 6]+(-330)*M[3, 4]+(-33*260)*M[3, 5]+(-33*152)*M[3, 6]+2600*M[4, 5]+1520*M[4, 6]+260*152*M[5, 6]))

#g
David8<-c(17, 16)
Svet8<-c(16, 26)
logWage_David8<-b1+b2*(David8[1])+b3*(David8[1])^2+b4*(David8[2])+b5*(David8[2])^2+b6*David8[1]*David8[2]
logWage_Svet8<-b1+b2*(Svet8[1])+b3*(Svet8[1])^2+b4*(Svet8[2])+b5*(Svet8[2])^2+b6*Svet8[1]*Svet8[2]
logWage_Svet8-logWage_David8
sqrt(M[2, 2]+1089*M[3, 3]+100*M[4, 4]+420*420*M[5, 5]+144*144*M[6, 6]
     +2*(33*M[2, 3]+(-10)*M[2, 4]+(-420)*M[2, 5]+(-144)*M[2, 6]+(-330)*M[3, 4]+(-33*420)*M[3, 5]+(-33*144)*M[3, 6]+4200*M[4, 5]+1440*M[4, 6]+420*144*M[5, 6]))

#h
Wendy<-c(12, 17)
Jill<-c(16, 11)
logWage_Wendy<-b1+b2*Wendy[1]+b3*(Wendy[1])^2+b4*Wendy[2]+b5*(Wendy[2])^2+b6*Wendy[1]*Wendy[2]
logWage_Jill<-b1+b2*Jill[1]+b3*(Jill[1])^2+b4*Jill[2]+b5*(Jill[2])^2+b6*Jill[1]*Jill[2]
logWage_Wendy-logWage_Jill
sqrt(16*M[2, 2]+112*112*M[3, 3]+36*M[4, 4]+168*168*M[5, 5]+28*28*M[6, 6]
     +2*(448*M[2, 3]+(-24)*M[2, 4]+(-672)*M[2, 5]+(-102)*M[2, 6]+(-672)*M[3, 4]+(-112*168)*M[3, 5]+(-112*28)*M[3, 6]+6*168*M[4, 5]+6*28*M[4, 6]+168*28*M[5, 6]))

#i
g1=(-1/(2*b5))#-1068.33
g2=(b4+16*b6)/(2*b5*b5)#65546.12
g3=(-8/b5)#17093.2
sqrt(g1^2*M[4, 4]+g2^2*M[5, 5]+g3^2*M[6, 6]+2*(g1*g2*M[4, 5]+g1*g3*M[4, 6]+g2*g3*M[5, 6]))
```