Math 644, Fall 2012

Solution for Homework 6

Problem 1.

(a)

Hard hat: $E(Y) = \beta_0 + \beta_2 + \beta_1 X_1$

Bump hat: $E(Y) = \beta_0 + \beta_3 + \beta_1 X_1$

None: $E(Y) = \beta_0 + \beta_1 X_1$

(b)

(1) $H_0 : \beta_3 = 0; \quad H_a : \beta_3 \neq 0;$

(1) $H_0 : \beta_3 = \beta_2; \quad H_a : \beta_3 \neq \beta_2;$

Problem 2.

(1) β_3 means the difference in the intercepts between M2 and M4

(2) $\beta_4 - \beta_3$ is the difference in the intercepts between M2 and M3

(3) when all the other factors are fixed, the expected increment of Y as X_1 increase by one unit.

(4) $\beta_7 = 0$ means there is no difference on the effect of X_1 on the response for M3 and M4.

(5) $\beta_5 - \beta_6$ is difference on the effect of X_1 on the response for M1 and M2.

Problem 3.

See the following R code:

```
xy = read.table('CH08PR06.txt')
y = xy[,1]
x = xy[,2]
x2 = x*x

# part (a)
ra = lm(y~x+x2)
summary(ra)
plot(x, residuals(ra))

# the fit appears to be good, R^2 = 0.8143
```

```
# part (b)
# H_0: beta_1 = beta_2 = 0
# F =  52.63 > F(0.99, 2, 24) therefore we reject H_0 and there is
# regression relation

# p-value =  1.678e-09 < 0.01
# we also conclude that there is regression relation

# part (c)
xnew = data.frame(x = 15, x2 = 225)
predict(ra, xnew, interval="confidence", level=0.99)

# the CI is [17.64092, 22.63492]

# part (d)

# H0: beta_2 = 0;
# since the t-vale
# |t| = 5.045 > t(0.995, 24), we reject  H0, i.e.  the quadratic term cannot
be dropped.
```

Problem 4.

See the following R code:

```
xy = read.table('CH06PR18.txt')
y = xy[,1]
x1 = xy[,2]
x2 = xy[,3]
x2 = xy[,4]
x4 = xy[,5]

x11 = x1*x1

# part (a)
ra = lm(y~x1+x2+x4+x11)
e = residuals(ra)
par(mfrow = c(2, 2))
plot(x1, e)
plot(x2, e)
plot(x4, e)

# the plots show no clear trend pattern. The model fit well

# part (b)
summary(ra)
# Ra^2 = 0.4534

# part (c)
#approach 1
# t-value = 0.845  < t(0.95, 76) = 1.984
# we accept that x11 can be dropped
```

```
#approach 2
rb = lm(y~x1+x2+x4)
# F-value = (123.990-122.834)/(122.834/76) = 0.7152417 < F(0.95,1, 76) =
3.936256
# we accept that x11 can be dropped

# part (d)
xnew = data.frame(x1=8, x2 = 16, x4 = 250000, x11 = 64)
predict(rb, xnew, interval='confidence', level=0.95)

# the interval is [-40.91043, 29.07418]
```

Problem 5.

a.

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \\ 1 & 0 \end{pmatrix} \quad \text{thus} \quad X'X = \begin{pmatrix} n & n_1 \\ n_1 & n_1 \end{pmatrix}$$

We have

$$(X'X)^{-1} = \begin{pmatrix} \frac{1}{n-n_1} & \frac{-1}{n-n_1} \\ \frac{-1}{n-n_1} & \frac{n}{(n-n_1)n_1} \end{pmatrix}$$

b.

$$X'Y = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n_1} Y_i \end{pmatrix}$$

Thus

$$b = \begin{pmatrix} \bar{Y}_d \\ \frac{n}{n-n_1}(-\bar{Y} + \bar{Y}_c) \end{pmatrix}$$

where $\bar{Y}_c, \bar{Y}_d$ is means for incorporated firm and non-incorporated firm respectively. We have

$$\hat{Y} = \bar{Y}_d \qquad \text{if non-incorporated firm}$$

and

$$\hat{Y} = \bar{Y}_c \qquad \text{if incorporated firm}$$

c.

$$SSE = \sum_{i=1}^{n_1}(Y_i - \bar{Y}_c)^2 + \sum_{i=n_1+1}^{n}(Y_i - \bar{Y}_d)^2$$

and

$$SSR = n_1(\bar{Y}_c - \bar{Y})^2 + (n - n_1)(\bar{Y}_d - \bar{Y})^2$$

Problem 6.

a. The fitted model is

$$
\begin{array}{ccccccccc}
Y & = & \text{-207.5} & + & 0.0005515X_1 & + & 0.1070X_2 & + & 149.0D_1 \\
\text{(SE)} & & (70.28\ ) & & (0.0002835) & & (0.01325) & & (86.83) \\
 & & & & & + & 145.5D_2 & + & 191.2D_3 \\
 & & & & & & (85.15) & & (80.03)
\end{array}
$$

SSE = 139093455,

b. The CIs for β_3 and β_4 are respectively

$$149.0 \pm 1.64 * 86.83 = [6.5988, 291.4012]$$

and

$$145.5 \pm 1.64 * 85.15 = [5.8540, 285.1460]$$

They have overlap (and include the other in the CI), thus, there is no difference in the regional effect

Another approach is by considering to models

$$Full: \qquad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 D_1 + \beta_4 D_4 + +\beta_5 D_5 + \varepsilon$$

and

$$Reduced: \qquad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (D_1 + D_4) + \beta_5 D_5 + \varepsilon$$

c. H_0 : there is no geographic effects The reduced model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

For teh reduced model $SSE = 140967081$. we have

$F = (140967081 - 139093455)/3/(139093455/434) = 1.948699 < F(0.99, 3, 434) = 2.62$

There is no geographic effects

Also, see the following R code:

```
xy = read.table('APPENC02.txt')
y = xy[,8]

x1 = xy[,5]
x2 = xy[,16]
R = xy[,17]

D1 = (R==1)+0
D2 = (R==2)+0
D3 = (R==3)+0

# part (a)

ra = lm(y~x1+x2+D1+D2+D3)
summary(ra)

# part (c)
rc = lm(y~x1+x2)
anova(rc)
```