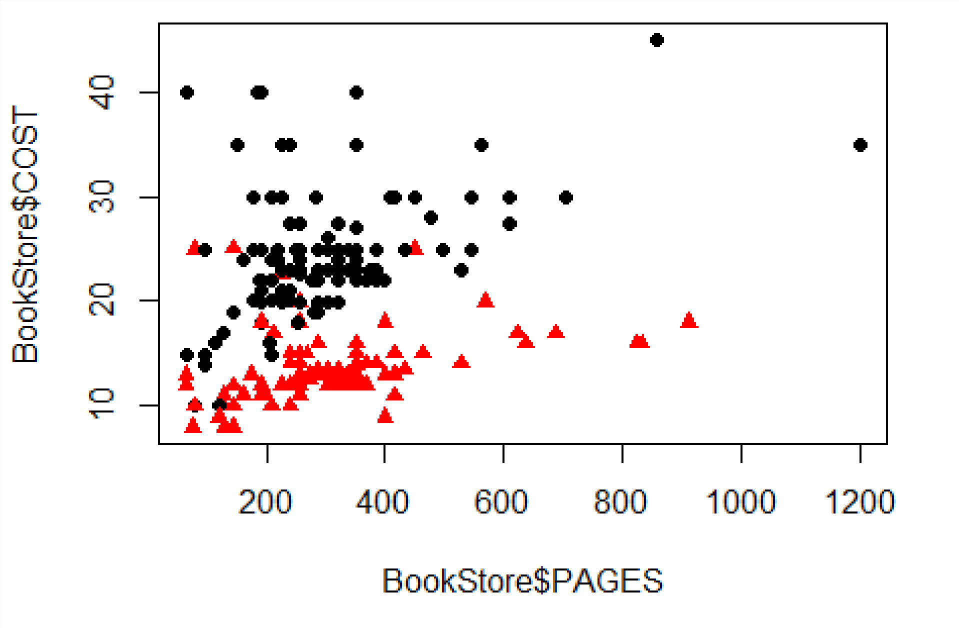
HW 08 Part I

```
setwd ("c:/MSA5020/files/")
BookStore=read.csv("BOOKCOST7.csv")
#Part A (10 pts)
plot(BookStore$PAGES, BookStore$COST, pch=BookStore$SOFTCOVER+16, col=BookSto
re$SOFTCOVER+1)
```



The plot confirms that hardcover books cost more than soft cover ones. There also seem to some extremely large values in the upper right corner of the graph. There exist a positive linear relation between the book cost and the number of pages based on the cover type.

```r
#Part B (10 pts)
Model1=lm(COST~PAGES+SOFTCOVER, data=BookStore)
summary(Model1)

##
## Call:
## lm(formula = COST ~ PAGES + SOFTCOVER, data = BookStore)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.2465  -2.2260  -0.7968   1.1107  18.4347
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.786733   0.738462  28.149  < 2e-16 ***
## PAGES         0.012164   0.002048   5.939 1.22e-08 ***
## SOFTCOVER   -10.798613   0.652305 -16.555  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.599 on 204 degrees of freedom
## Multiple R-squared:  0.6003, Adjusted R-squared:  0.5964
## F-statistic: 153.2 on 2 and 204 DF,  p-value: < 2.2e-16
```

Since the p-values for the coefficients of the number of pages (p-value =0.00) and the type of book cover (p-value=0.00) are both less than 0.05, we can conclude at 95% confidence that both variables are significant predictors of the cost of a book.

```r
#Part C(10 pts)
```

COST = 20.8 + 0.0122 PAGES - 10.8 SOFTCOVER

$\beta_1 = 0.0122$; The average cost of the book increases by 1.22cents for every additional page in the book.

$\beta_2 = 10.8$; The hardcover books cost about $10.8 more than softcover books on average.

```
#Part D (15 pts)
library("alr3")

#Lack of fit test
```

a) Lack of fit test

For the lack of fit, we test the hypothesis that

$H_0: y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
$H_a: y \neq \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

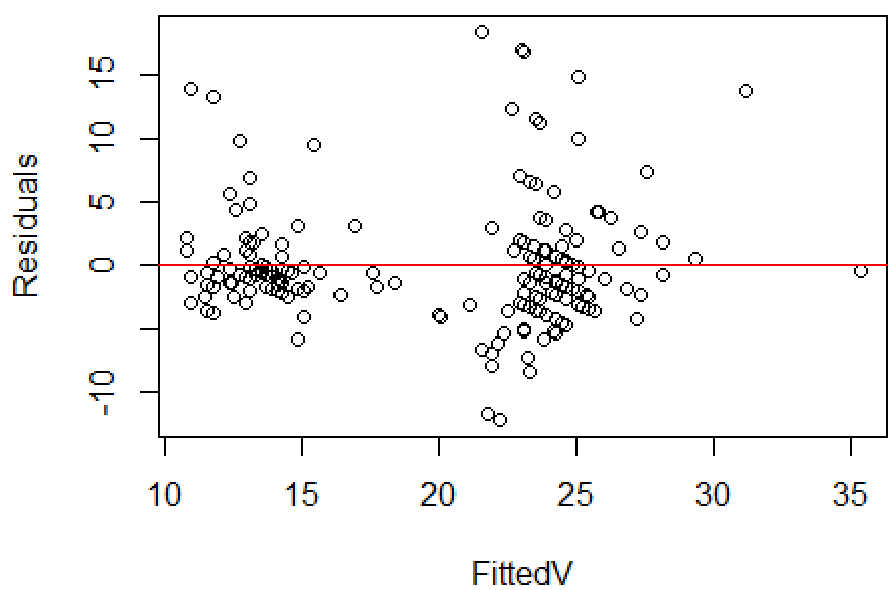
```
pureErrorAnova(Model1)

## Analysis of Variance Table
##
## Response: COST
##                 Df Sum Sq Mean Sq  F value    Pr(>F)
## PAGES            1  684.7   684.7  35.4494 2.761e-08 ***
## SOFTCOVER        1 5796.5  5796.5 300.0948 < 2.2e-16 ***
## Residuals      204 4314.8    21.2
##  Lack of fit    86 2035.6    23.7   1.2254    0.1524
##  Pure Error    118 2279.2    19.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value for the lack of fit test is 0.152 (or 0.10)> 0.05 we fail to reject the null hypothesis. There is **no** sufficient evidence at 95% confidence level to conclude that the linear model assumption is not adequate.

```
#Constant variance test
```

H_0: Variance is constant
H_a: Variance is not constant




```
ncvTest(Model1)
##Non-constant Variance Score Test

##Variance formula: ~ fitted.values
##Chisquare = 5.28059, Df = 1, p = 0.021564
```

Since the p-value is less than 0.05, we reject the null hypothesis. There is sufficient evidence at 95% confidence level to suggest that the variance is not constant.

<u>Normality Test</u>

H_0: The error term is normally distributed
H_a: The error term is not normally distributed

```
library("nortest")

## Warning: package 'nortest' was built under R version 3.1.3

ad.test(Model1Res$Residuals)
```

```
##
##  Anderson-Darling normality test
##
## data:  Model1Res$Residuals
## A = 9.738, p-value < 2.2e-16
```

Part E **(10 pts)**

From the output above, the Anderson-Darling value is 9.738 and the p-value for the normality test is less than 0.005.

Since p-value <0.005 is less than 0.05, we reject H_0. There is sufficient evidence to conclude that the normality assumption of the error term is violated.

```
### Part F (10 pts)
BookStore$PAGES.SOFT=BookStore$PAGES*BookStore$SOFTCOVER
Model2=lm(COST~PAGES+SOFTCOVER+PAGES.SOFT, data=BookStore)
summary(Model2)

##
## Call:
## lm(formula = COST ~ PAGES + SOFTCOVER + PAGES.SOFT, data = BookStore)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4766  -2.2143  -0.8453   1.0037  19.4456
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.500428   0.913658  21.343  < 2e-16 ***
## PAGES        0.016468   0.002734   6.023 7.93e-09 ***
## SOFTCOVER   -7.921170   1.386921  -5.711 3.95e-08 ***
## PAGES.SOFT  -0.009543   0.004072  -2.344   0.0201 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.549 on 203 degrees of freedom
## Multiple R-squared:  0.6109, Adjusted R-squared:  0.6051
## F-statistic: 106.2 on 3 and 203 DF,  p-value: < 2.2e-16

library("HH")

vif(Model2)

##      PAGES  SOFTCOVER PAGES.SOFT
##   1.821837   4.621290   5.494748
```

Although multicollinearity may be evident, especially with the interaction term, there does not seem to be a serious issue of multicollinearity (the coefficients are significant).

```
### Part G (10 pts)
```

The regression equation is

COST = 19.5 + 0.0165 PAGES - 7.92 SOFTCOVER - 0.00954 PAGES-SOFT

$\beta_1 = 0.0165$; The average cost of the hardcover books increases by 1.65 cents for every additional page.

$\beta_1 = 0.0165 - .00954 = .00696$; The average cost of the softcover books increases by 0.696 cents for every additional page.

```
### Part H (15 pts)
#lack of fit test
pureErrorAnova(Model2)

## Analysis of Variance Table
##
## Response: COST
##                 Df Sum Sq Mean Sq  F value    Pr(>F)
## PAGES            1  684.7   684.7  35.4494 2.761e-08 ***
## SOFTCOVER        1 5796.5  5796.5 300.0948 < 2.2e-16 ***
## PAGES.SOFT       1  113.7   113.7   5.8857   0.01678 *
## Residuals      203 4201.1    20.7
##  Lack of fit    85 1921.9    22.6   1.1706   0.21338
##  Pure Error    118 2279.2    19.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##Alteratively
anova(Model2, Factor.Model)

## Analysis of Variance Table
##
## Model 1: COST ~ PAGES + SOFTCOVER + PAGES.SOFT
## Model 2: COST ~ factor(PAGES) + factor(SOFTCOVER)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    203 4201.1
## 2    139 2692.0 64    1509.2 1.2176 0.1695
```

For the lack of fit, we test the hypothesis that

H_0: $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 + \varepsilon$
H_a: $y \neq \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 + \varepsilon$

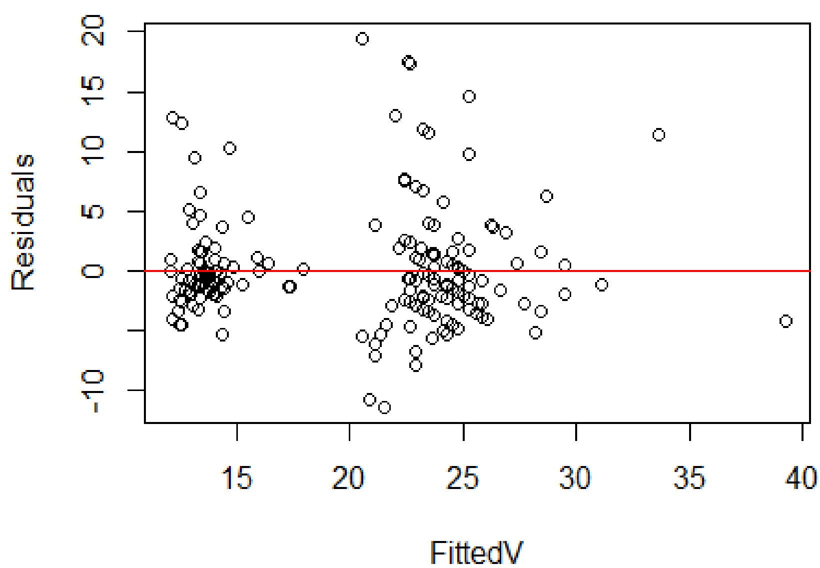
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	6594.9	2198.3	106.22	0.000
Residual Error	203	4201.1	20.7		
Lack of Fit	85	1921.9	22.6	1.17	0.213
Pure Error	118	2279.2	19.3		
Total	206	10796.0			

Since the p-value for the lack of fit test is 0.213 (or 0.1695)> 0.05 we fail to reject the null hypothesis. There is no sufficient evidence at 95% confidence level to conclude that the linear model is not adequate.

```
#Constant variance test
```

H_0: Variance is constant
H_a: Variance is not constant




```
#It seems that when fitted values are increasing the variation is increasing.
Hence we need to order data on fitted values in a assending fashion.
ncvTest(Model2)

## Non-constant Variance Score Test

## Variance formula: ~ fitted.values

## Chisquare =4.151172, Df =1, p =0.041606
```

Since the p-value is less than 0.05 we reject the null hypothesis. We have sufficient evidence at 95% confidence level to conclude that the error variance is not constant.

```
##Normality test
```

Ho: The residuals are normally distributed
Ha: The residuals are not normally distributed

```
ad.test(Model2Res$Residuals)

##
##  Anderson-Darling normality test
##
## data:  Model2Res$Residuals
## A = 9.8377, p-value < 2.2e-16
```

From the output above, the Anderson-Darling value is 9.838 and the p-value for the normality test is less than 0.005.

Since p-value <0.005 is less than 0.05 we reject H_0. We have sufficient evidence to conclude at 95% confidence level that the error terms are not normally distributed.

```
### PART I (10 pts)
anova(Model1, Model2)

## Analysis of Variance Table
##
## Model 1: COST ~ PAGES + SOFTCOVER
## Model 2: COST ~ PAGES + SOFTCOVER + PAGES.SOFT
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    204 4314.8
## 2    203 4201.1  1    113.69 5.4934 0.02006 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The inclusion of the interaction term improves the model; P value of the F statistic is less then significant level of 5%. Hence, adding an interaction term improves the model.

HW 08 Part II

```
setwd ("c:/MSA5020/files/")
Data=read.csv("LaQuinta.csv")


attach(Data)
library("leaps")

############## PART A  (20 PTS)
####Using CP
Model.allfit.Cp=leaps(x=cbind(ROOMS, NEAREST, OFFICE, COLLEGE, INCOME, DISTTW
N), y=MARGIN, method="Cp", nbest=2)

Model.allfit.Cp.Table = cbind(Model.allfit.Cp$which, Model.allfit.Cp$size, Mo
del.allfit.Cp$Cp)

n= length(Model.allfit.Cp$size)
dimnames(Model.allfit.Cp.Table) <- list(1:n,c("Rooms","Nearest","Office","Col
lege","Income","DisttWn","Size","Cp"))

round(Model.allfit.Cp.Table ,digits=3)

##    Rooms Nearest Office College Income DisttWn Size     Cp
## 1      0       0      1       0      0       0    2 54.201
## 2      1       0      0       0      0       0    2 59.126
## 3      1       0      1       0      0       0    3 19.933
## 4      1       0      0       0      1       0    3 46.793
## 5      1       0      1       0      1       0    4 14.016
## 6      1       1      1       0      0       0    4 14.414
## 7      1       1      1       0      1       0    5  7.773
## 8      1       0      1       1      1       0    5 13.132
## 9      1       1      1       0      1       1    6  7.572
## 10     1       1      1       1      1       0    6  7.623
## 11     1       1      1       1      1       1    7  7.000

######## R2
Model.allfit.r2=leaps(x=cbind(ROOMS, NEAREST, OFFICE, COLLEGE, INCOME, DISTTW
N), y=MARGIN, method="r2", nbest=2)

Model.allfit.r2.Table = cbind(Model.allfit.r2$which, Model.allfit.r2$size, Mo
del.allfit.r2$r2*100)


n= length(Model.allfit.r2$size)
dimnames(Model.allfit.r2.Table) <- list(1:n,c("Rooms","Nearest","Office","Col
lege","Income","DisttWn","Size","R2"))

round(Model.allfit.r2.Table ,digits=2)
```

```
##    Rooms Nearest Office College Income DisttWn Size    R2
## 1      0       0      1       0      0       0    2 24.66
## 2      1       0      0       0      0       0    2 22.19
## 3      1       0      1       0      0       0    3 42.85
## 4      1       0      0       0      1       0    3 29.38
## 5      1       0      1       0      1       0    4 46.82
## 6      1       1      1       0      0       0    4 46.62
## 7      1       1      1       0      1       0    5 50.96
## 8      1       0      1       1      1       0    5 48.27
## 9      1       1      1       0      1       1    6 52.06
## 10     1       1      1       1      1       0    6 52.04
## 11     1       1      1       1      1       1    7 53.35
```

```
######## Adj R2
Model.allfit.adjr2=leaps(x=cbind(ROOMS, NEAREST, OFFICE, COLLEGE, INCOME, DIS
TTWN), y=MARGIN, method="adjr2", nbest=2)

Model.allfit.adjr2.Table = cbind(Model.allfit.adjr2$which, Model.allfit.adjr2
$size, Model.allfit.adjr2$adjr2*100)
n= length(Model.allfit.adjr2$size)
dimnames(Model.allfit.adjr2.Table) <- list(1:n,c("Rooms","Nearest","Office","
College","Income","DisttWn","Size","AdjR2"))

round(Model.allfit.adjr2.Table ,digits=2)
```

```
##    Rooms Nearest Office College Income DisttWn Size AdjR2
## 1      0       0      1       0      0       0    2 23.89
## 2      1       0      0       0      0       0    2 21.40
## 3      1       0      1       0      0       0    3 41.67
## 4      1       0      0       0      1       0    3 27.92
## 5      1       0      1       0      1       0    4 45.16
## 6      1       1      1       0      0       0    4 44.96
## 7      1       1      1       0      1       0    5 48.89
## 8      1       0      1       1      1       0    5 46.09
## 9      1       1      1       0      1       1    6 49.51
## 10     1       1      1       1      1       0    6 49.49
## 11     1       1      1       1      1       1    7 50.34
```

It seems that model with Rooms, Nearest, Office, and Income performs fairly well with respect to all metrics.

################ PART B **(20 PTS)**

```
library("MASS")
Model.Full=lm(MARGIN~ ROOMS + NEAREST + OFFICE + COLLEGE + INCOME + DISTTWN)

dropterm(Model.Full, test="F")
```

```
## Single term deletions
##
```

```
## Model:
## MARGIN ~ ROOMS + NEAREST + OFFICE + COLLEGE + INCOME + DISTTWN
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>                2779.4 346.48
## ROOMS    1   1145.99 3925.4 379.00  38.346 1.591e-08 ***
## NEAREST  1    257.80 3037.2 353.35   8.626  0.004177 **
## OFFICE   1    960.37 3739.8 374.16  32.135 1.606e-07 ***
## COLLEGE  1     76.87 2856.3 347.21   2.572  0.112150
## INCOME   1    260.81 3040.2 353.45   8.727  0.003971 **
## DISTTWN  1     78.38 2857.8 347.26   2.623  0.108733
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
###Drop COLLEGE variable

Model.backward=update(Model.Full, .~.-COLLEGE)
dropterm(Model.backward, test="F")
```

```
## Single term deletions
##
## Model:
## MARGIN ~ ROOMS + NEAREST + OFFICE + INCOME + DISTTWN
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>                2856.3 347.21
## ROOMS    1   1193.99 4050.3 380.14  39.295 1.097e-08 ***
## NEAREST  1    278.98 3135.2 354.53   9.181  0.003158 **
## OFFICE   1    960.25 3816.5 374.19  31.602 1.930e-07 ***
## INCOME   1    238.65 3094.9 353.23   7.854  0.006158 **
## DISTTWN  1     65.77 2922.0 347.49   2.165  0.144557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Drop DISTTWN

Model.backward=update(Model.backward, .~.-DISTTWN)
dropterm(Model.backward, test="F")
```

```
## Single term deletions
##
## Model:
## MARGIN ~ ROOMS + NEAREST + OFFICE + INCOME
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>                2922.0 347.49
## ROOMS    1   1243.35 4165.4 380.94  40.423 7.095e-09 ***
## NEAREST  1    246.34 3168.4 353.58   8.009  0.005680 **
## OFFICE   1    940.58 3862.6 373.39  30.580 2.811e-07 ***
## INCOME   1    258.25 3180.3 353.96   8.396  0.004668 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## STOP Nothing to exit
```

The resulting model with backward elimination is the same suggested model in all regressions. The 4 variables selected are ROOMS, NEAREST, OFFICE and INCOME.

```
############### Part C
```
(20 PTS)

```
Model.null=lm(MARGIN~1)
addterm(Model.null,Model.Full, test= "F")

## Single term additions
##
## Model:
## MARGIN ~ 1
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>               5958.2 410.74
## ROOMS    1   1322.17 4636.1 387.65  27.949 7.561e-07 ***
## NEAREST  1    196.68 5761.6 409.38   3.345   0.07043 .
## OFFICE   1   1469.34 4488.9 384.42  32.078 1.482e-07 ***
## COLLEGE  1     95.09 5863.1 411.13   1.589   0.21040
## INCOME   1    373.71 5584.5 406.26   6.558   0.01197 *
## DISTTWN  1     70.33 5887.9 411.55   1.171   0.28195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##First OFFICE Variable enters the model

Model.forward=update(Model.null,.~.+OFFICE)
addterm(Model.forward,Model.Full, test= "F")

## Single term additions
##
## Model:
## MARGIN ~ OFFICE
##         Df Sum of Sq    RSS    AIC F Value    Pr(F)
## <none>               4488.9 384.42
## ROOMS    1   1083.90 3405.0 358.78 30.8778 2.41e-07 ***
## NEAREST  1    125.18 4363.7 383.59  2.7825  0.09852 .
## COLLEGE  1     93.99 4394.9 384.30  2.0744  0.15301
## INCOME   1    186.09 4302.8 382.19  4.1951  0.04324 *
## DISTTWN  1     95.43 4393.5 384.27  2.1068  0.14987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Rooms enters the model

Model.forward=update(Model.forward,.~.+ROOMS)
addterm(Model.forward,Model.Full, test= "F")
```

```
## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS
##         Df Sum of Sq    RSS    AIC F Value    Pr(F)
## <none>                3405.0 358.78
## NEAREST  1   224.711 3180.3 353.96  6.7831 0.010665 *
## COLLEGE  1    60.224 3344.8 359.00  1.7285 0.191734
## INCOME   1   236.623 3168.4 353.58  7.1695 0.008723 **
## DISTTWN  1    48.017 3357.0 359.36  1.3731 0.244174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##INCOME enters

Model.forward=update(Model.forward,.~.+INCOME)
addterm(Model.forward,Model.Full, test= "F")
```

```
## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME
##         Df Sum of Sq    RSS    AIC F Value   Pr(F)
## <none>                3168.4 353.58
## NEAREST  1   246.342 2922.0 347.49  8.0090 0.00568 **
## COLLEGE  1    86.172 3082.2 352.82  2.6560 0.10647
## DISTTWN  1    33.140 3135.2 354.53  1.0042 0.31885
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
###NEAREST

Model.forward=update(Model.forward,.~.+NEAREST)
addterm(Model.forward,Model.Full, test= "F")
```

```
## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST
##         Df Sum of Sq    RSS    AIC F Value  Pr(F)
## <none>                2922.0 347.49
## COLLEGE  1    64.265 2857.8 347.26  2.1138 0.1493
## DISTTWN  1    65.774 2856.3 347.21  2.1646 0.1446
```

```
### DISTTWN

Model.forward=update(Model.forward,.~.+DISTTWN)
addterm(Model.forward,Model.Full, test= "F")
```

```
## Single term additions
##
## Model:
```

```
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST + DISTTWN
##          Df Sum of Sq    RSS    AIC F Value  Pr(F)
## <none>                2856.3 347.21
## COLLEGE  1    76.871 2779.4 346.48  2.5722 0.1122
```

```
# add COLLEGE as welll
Model.forward=update(Model.forward,.~.+COLLEGE)
```

The forward selection approach does not really help much in this particular example since it just selects all variables. This may be due to setting alpha-to-enter at .25.

```
############## PART D (20 PTS)
addterm(Model.null,Model.Full, test= "F")
```

```
## Single term additions
##
## Model:
## MARGIN ~ 1
##          Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>                5958.2 410.74
## ROOMS    1   1322.17 4636.1 387.65  27.949 7.561e-07 ***
## NEAREST  1    196.68 5761.6 409.38   3.345   0.07043 .
## OFFICE   1   1469.34 4488.9 384.42  32.078 1.482e-07 ***
## COLLEGE  1     95.09 5863.1 411.13   1.589   0.21040
## INCOME   1    373.71 5584.5 406.26   6.558   0.01197 *
## DISTTWN  1     70.33 5887.9 411.55   1.171   0.28195
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##First OFFICE Variable enters the model

Model.stepwise=update(Model.null,.~.+OFFICE)
dropterm(Model.stepwise, test="F")
```

```
## Single term deletions
##
## Model:
## MARGIN ~ OFFICE
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>               4488.9 384.42
## OFFICE  1    1469.3 5958.2 410.74  32.078 1.482e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Nothing exit

addterm(Model.stepwise,Model.Full, test= "F")
```

```
## Single term additions
##
```

```
## Model:
## MARGIN ~ OFFICE
##          Df Sum of Sq    RSS    AIC F Value    Pr(F)
## <none>                4488.9 384.42
## ROOMS    1   1083.90 3405.0 358.78 30.8778 2.41e-07 ***
## NEAREST  1    125.18 4363.7 383.59  2.7825  0.09852 .
## COLLEGE  1     93.99 4394.9 384.30  2.0744  0.15301
## INCOME   1    186.09 4302.8 382.19  4.1951  0.04324 *
## DISTTWN  1     95.43 4393.5 384.27  2.1068  0.14987
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Rooms enters the model
Model.stepwise=update(Model.stepwise,.~.+ROOMS)

dropterm(Model.stepwise, test="F")

## Single term deletions
##
## Model:
## MARGIN ~ OFFICE + ROOMS
##        Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>               3405.0 358.78
## OFFICE  1    1231.1 4636.1 387.65  35.070 4.823e-08 ***
## ROOMS   1    1083.9 4488.9 384.42  30.878 2.410e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Nothing exit
addterm(Model.stepwise,Model.Full, test= "F")

## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS
##          Df Sum of Sq    RSS    AIC F Value    Pr(F)
## <none>                3405.0 358.78
## NEAREST  1   224.711 3180.3 353.96  6.7831 0.010665 *
## COLLEGE  1    60.224 3344.8 359.00  1.7285 0.191734
## INCOME   1   236.623 3168.4 353.58  7.1695 0.008723 **
## DISTTWN  1    48.017 3357.0 359.36  1.3731 0.244174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##INCOME enters

Model.stepwise=update(Model.stepwise,.~.+INCOME)

dropterm(Model.stepwise, test="F")

## Single term deletions
##
```

```
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME
##        Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>              3168.4 353.58
## OFFICE  1   1039.35 4207.7 379.95  31.492 1.935e-07 ***
## ROOMS   1   1134.44 4302.8 382.19  34.373 6.421e-08 ***
## INCOME  1    236.62 3405.0 358.78   7.170  0.008723 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Nothing exit
addterm(Model.stepwise,Model.Full, test= "F")

## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME
##         Df Sum of Sq    RSS    AIC F Value   Pr(F)
## <none>               3168.4 353.58
## NEAREST  1   246.342 2922.0 347.49  8.0090 0.00568 **
## COLLEGE  1    86.172 3082.2 352.82  2.6560 0.10647
## DISTTWN  1    33.140 3135.2 354.53  1.0042 0.31885
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

###NEAREST

Model.stepwise=update(Model.stepwise,.~.+NEAREST)

dropterm(Model.stepwise, test="F")

## Single term deletions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>               2922.0 347.49
## OFFICE   1    940.58 3862.6 373.39  30.580 2.811e-07 ***
## ROOMS    1   1243.35 4165.4 380.94  40.423 7.095e-09 ***
## INCOME   1    258.25 3180.3 353.96   8.396  0.004668 **
## NEAREST  1    246.34 3168.4 353.58   8.009  0.005680 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Nothing exit
addterm(Model.stepwise,Model.Full, test= "F")

## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST
##         Df Sum of Sq    RSS    AIC F Value  Pr(F)
```

```
## <none>              2922.0 347.49
## COLLEGE  1    64.265 2857.8 347.26  2.1138 0.1493
## DISTTWN  1    65.774 2856.3 347.21  2.1646 0.1446

### DISTTWN

Model.stepwise=update(Model.stepwise,.~.+DISTTWN)

dropterm(Model.stepwise, test="F")

## Single term deletions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST + DISTTWN
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>              2856.3 347.21
## OFFICE   1    960.25 3816.5 374.19  31.602 1.930e-07 ***
## ROOMS    1   1193.99 4050.3 380.14  39.295 1.097e-08 ***
## INCOME   1    238.65 3094.9 353.23   7.854  0.006158 **
## NEAREST  1    278.98 3135.2 354.53   9.181  0.003158 **
## DISTTWN  1     65.77 2922.0 347.49   2.165  0.144557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Nothing exit
addterm(Model.stepwise,Model.Full, test= "F")

## Single term additions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST + DISTTWN
##         Df Sum of Sq    RSS    AIC F Value  Pr(F)
## <none>              2856.3 347.21
## COLLEGE  1    76.871 2779.4 346.48  2.5722 0.1122

### Add COllege
Model.stepwise=update(Model.stepwise,.~.+COLLEGE)

dropterm(Model.stepwise, test="F")

## Single term deletions
##
## Model:
## MARGIN ~ OFFICE + ROOMS + INCOME + NEAREST + DISTTWN + COLLEGE
##         Df Sum of Sq    RSS    AIC F Value     Pr(F)
## <none>              2779.4 346.48
## OFFICE   1    960.37 3739.8 374.16  32.135 1.606e-07 ***
## ROOMS    1   1145.99 3925.4 379.00  38.346 1.591e-08 ***
## INCOME   1    260.81 3040.2 353.45   8.727  0.003971 **
## NEAREST  1    257.80 3037.2 353.35   8.626  0.004177 **
## DISTTWN  1     78.38 2857.8 347.26   2.623  0.108733
## COLLEGE  1     76.87 2856.3 347.21   2.572  0.112150
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Nothing exit
```

The stepwise procedure, while setting alpha-to-enter = alpha-to-remove = .25, falls in the same predicament as the forward selection (this is perhaps not surprising since stepwise starts with forward selection). Again, it is most likely due to not setting the threshold stringent enough.

PART E (20 PTS)

Based on the conclusions in (a), (b), (c) and (d) above, the consensus seems to suggest that the 5 variables ROOMS, NEAREST, OFFICE, COLLEGE and INCOME should be selected as the predictor variables as one begins to build the model for predicting the operating margins.