CSC 423 Homework - Chapter 6 & 7

Jasmine Dumas

October 27, 2015

page 341 #6.8

(a) Build a model for y, LOWBID and apply stepwise regression

```
load("~/Desktop/depaul/CSC423/rdata/R/Exercises&Examples/FLAG2.Rdata")
head(FLAG2)
```

```
##    LOWBID  DOTEST LBERATIO STATUS DISTRICT NUMBIDS DAYSEST RDLNGTH
## 1  362916  385963  0.94029      0        1       3     100   7.200
## 2  152056  175396  0.86693      1        1       3      75   0.000
## 3  239665  194650  1.23126      1        1       3      65   0.206
## 4 1559368 1925307  0.80993      0        1      10     250   3.600
## 5  144062  252925  0.56958      0        1       8      90  23.700
## 6 1187104 1573451  0.75446      0        1       5     230   2.600
##    PCTASPH  PCTBASE PCTEXCAV PCTMOBIL PCTSTRUC PCTTRAFF SUBCONT
## 1 0.626369 0.000000 0.091475 0.019977 0.116770  0.09170       0
## 2 0.153284 0.012364 0.142221 0.047351 0.018020  0.14764       0
## 3 0.082670 0.000000 0.105314 0.049235 0.222602  0.06029       0
## 4 0.189911 0.298819 0.238240 0.011543 0.166206  0.04054       0
## 5 0.316169 0.245298 0.159904 0.003471 0.082562  0.00882       0
## 6 0.281976 0.214393 0.241728 0.008424 0.147573  0.02250       0
```

```
library(MASS)
FLAG2 <- na.omit(FLAG2) # step() requires removal of missing data before
full.road.model <- lm(LOWBID ~ DOTEST + LBERATIO + STATUS + DISTRICT + NUMBIDS + DAYSEST + RDLNGTH + PCTASPH +
road.model <- step(full.road.model, direction="both")
```

```
## Start:  AIC=5403.07
## LOWBID ~ DOTEST + LBERATIO + STATUS + DISTRICT + NUMBIDS + DAYSEST +
##     RDLNGTH + PCTASPH + PCTBASE + PCTEXCAV + PCTMOBIL + PCTSTRUC +
##     PCTTRAFF + SUBCONT
##
##             Df  Sum of Sq        RSS    AIC
## - STATUS     1 4.6438e+08 1.2300e+13 5401.1
## - SUBCONT    1 3.3494e+09 1.2303e+13 5401.1
## - PCTMOBIL   1 9.9185e+09 1.2310e+13 5401.2
## - DAYSEST    1 1.2081e+10 1.2312e+13 5401.3
## - PCTEXCAV   1 2.7176e+10 1.2327e+13 5401.5
## - PCTSTRUC   1 3.9054e+10 1.2339e+13 5401.8
## - PCTBASE    1 4.1987e+10 1.2342e+13 5401.8
## - PCTASPH    1 5.2671e+10 1.2352e+13 5402.0
## - DISTRICT   1 5.7062e+10 1.2357e+13 5402.1
## <none>                    1.2300e+13 5403.1
## - NUMBIDS    1 1.3402e+11 1.2434e+13 5403.4
## - PCTTRAFF   1 1.7964e+11 1.2479e+13 5404.2
## - RDLNGTH    1 1.9001e+11 1.2490e+13 5404.4
## - LBERATIO   1 2.5788e+12 1.4878e+13 5442.4
## - DOTEST     1 2.0108e+14 2.1338e+14 6020.3
##
## Step:  AIC=5401.08
## LOWBID ~ DOTEST + LBERATIO + DISTRICT + NUMBIDS + DAYSEST + RDLNGTH +
```

```
##     PCTASPH + PCTBASE + PCTEXCAV + PCTMOBIL + PCTSTRUC + PCTTRAFF +
##     SUBCONT
##
##            Df  Sum of Sq        RSS    AIC
## - SUBCONT   1 3.4416e+09 1.2304e+13 5399.1
## - PCTMOBIL  1 9.6486e+09 1.2310e+13 5399.2
## - DAYSEST   1 1.1934e+10 1.2312e+13 5399.3
## - PCTEXCAV  1 2.6773e+10 1.2327e+13 5399.6
## - PCTSTRUC  1 3.8966e+10 1.2339e+13 5399.8
## - PCTBASE   1 4.1902e+10 1.2342e+13 5399.8
## - PCTASPH   1 5.2296e+10 1.2352e+13 5400.0
## - DISTRICT  1 5.6615e+10 1.2357e+13 5400.1
## <none>                   1.2300e+13 5401.1
## - NUMBIDS   1 1.4105e+11 1.2441e+13 5401.6
## - PCTTRAFF  1 1.9454e+11 1.2495e+13 5402.5
## - RDLNGTH   1 1.9536e+11 1.2495e+13 5402.5
## + STATUS    1 4.6438e+08 1.2300e+13 5403.1
## - LBERATIO  1 3.0655e+12 1.5366e+13 5447.4
## - DOTEST    1 2.0162e+14 2.1392e+14 6018.8
##
## Step:  AIC=5399.14
## LOWBID ~ DOTEST + LBERATIO + DISTRICT + NUMBIDS + DAYSEST + RDLNGTH +
##     PCTASPH + PCTBASE + PCTEXCAV + PCTMOBIL + PCTSTRUC + PCTTRAFF
##
##            Df  Sum of Sq        RSS    AIC
## - PCTMOBIL  1 9.8548e+09 1.2313e+13 5397.3
## - DAYSEST   1 1.1597e+10 1.2315e+13 5397.3
## - PCTEXCAV  1 2.4896e+10 1.2328e+13 5397.6
## - PCTSTRUC  1 3.7407e+10 1.2341e+13 5397.8
## - PCTBASE   1 3.9831e+10 1.2343e+13 5397.8
## - PCTASPH   1 5.4635e+10 1.2358e+13 5398.1
## - DISTRICT  1 6.0916e+10 1.2364e+13 5398.2
## <none>                   1.2304e+13 5399.1
## - NUMBIDS   1 1.3886e+11 1.2442e+13 5399.6
## - PCTTRAFF  1 1.9868e+11 1.2502e+13 5400.6
## - RDLNGTH   1 2.0352e+11 1.2507e+13 5400.7
## + SUBCONT   1 3.4416e+09 1.2300e+13 5401.1
## + STATUS    1 5.5665e+08 1.2303e+13 5401.1
## - LBERATIO  1 3.1328e+12 1.5436e+13 5446.4
## - DOTEST    1 2.0562e+14 2.1792e+14 6020.9
##
## Step:  AIC=5397.31
## LOWBID ~ DOTEST + LBERATIO + DISTRICT + NUMBIDS + DAYSEST + RDLNGTH +
##     PCTASPH + PCTBASE + PCTEXCAV + PCTSTRUC + PCTTRAFF
##
##            Df  Sum of Sq        RSS    AIC
## - DAYSEST   1 1.2480e+10 1.2326e+13 5395.5
## - PCTEXCAV  1 2.1870e+10 1.2335e+13 5395.7
## - PCTSTRUC  1 4.4378e+10 1.2358e+13 5396.1
## - PCTASPH   1 4.6786e+10 1.2360e+13 5396.1
## - PCTBASE   1 5.1779e+10 1.2365e+13 5396.2
## - DISTRICT  1 5.8843e+10 1.2372e+13 5396.3
## <none>                   1.2313e+13 5397.3
## - NUMBIDS   1 1.3107e+11 1.2444e+13 5397.6
## - RDLNGTH   1 2.0402e+11 1.2517e+13 5398.9
## - PCTTRAFF  1 2.1367e+11 1.2527e+13 5399.0
## + PCTMOBIL  1 9.8548e+09 1.2304e+13 5399.1
## + SUBCONT   1 3.6479e+09 1.2310e+13 5399.2
## + STATUS    1 2.5500e+08 1.2313e+13 5399.3
## - LBERATIO  1 3.1853e+12 1.5499e+13 5445.2
```

```
## - DOTEST    1 2.0630e+14 2.1861e+14 6019.5
##
## Step:  AIC=5395.53
## LOWBID ~ DOTEST + LBERATIO + DISTRICT + NUMBIDS + RDLNGTH + PCTASPH +
##     PCTBASE + PCTEXCAV + PCTSTRUC + PCTTRAFF
##
##            Df  Sum of Sq        RSS    AIC
## - PCTEXCAV  1 2.0631e+10 1.2346e+13 5393.9
## - PCTASPH   1 4.7710e+10 1.2374e+13 5394.4
## - DISTRICT  1 5.6217e+10 1.2382e+13 5394.5
## - PCTBASE   1 6.0726e+10 1.2387e+13 5394.6
## - PCTSTRUC  1 7.0941e+10 1.2397e+13 5394.8
## <none>                   1.2326e+13 5395.5
## - NUMBIDS   1 1.4567e+11 1.2472e+13 5396.1
## - RDLNGTH   1 2.1004e+11 1.2536e+13 5397.2
## + DAYSEST   1 1.2480e+10 1.2313e+13 5397.3
## + PCTMOBIL  1 1.0738e+10 1.2315e+13 5397.3
## + SUBCONT   1 3.2974e+09 1.2323e+13 5397.5
## - PCTTRAFF  1 2.2886e+11 1.2555e+13 5397.5
## + STATUS    1 1.3700e+08 1.2326e+13 5397.5
## - LBERATIO  1 3.2476e+12 1.5573e+13 5444.3
## - DOTEST    1 4.9294e+14 5.0526e+14 6199.3
##
## Step:  AIC=5393.9
## LOWBID ~ DOTEST + LBERATIO + DISTRICT + NUMBIDS + RDLNGTH + PCTASPH +
##     PCTBASE + PCTSTRUC + PCTTRAFF
##
##            Df  Sum of Sq        RSS    AIC
## - PCTASPH   1 2.7081e+10 1.2374e+13 5392.4
## - DISTRICT  1 5.1931e+10 1.2398e+13 5392.8
## - PCTBASE   1 8.0408e+10 1.2427e+13 5393.3
## <none>                   1.2346e+13 5393.9
## - PCTSTRUC  1 1.3984e+11 1.2486e+13 5394.3
## - NUMBIDS   1 1.6263e+11 1.2509e+13 5394.7
## - RDLNGTH   1 2.0844e+11 1.2555e+13 5395.5
## + PCTEXCAV  1 2.0631e+10 1.2326e+13 5395.5
## - PCTTRAFF  1 2.1010e+11 1.2557e+13 5395.6
## + DAYSEST   1 1.1241e+10 1.2335e+13 5395.7
## + PCTMOBIL  1 7.6014e+09 1.2339e+13 5395.8
## + SUBCONT   1 1.5808e+09 1.2345e+13 5395.9
## + STATUS    1 1.7789e+05 1.2346e+13 5395.9
## - LBERATIO  1 3.2547e+12 1.5601e+13 5442.7
## - DOTEST    1 5.1940e+14 5.3175e+14 6208.4
##
## Step:  AIC=5392.37
## LOWBID ~ DOTEST + LBERATIO + DISTRICT + NUMBIDS + RDLNGTH + PCTBASE +
##     PCTSTRUC + PCTTRAFF
##
##            Df  Sum of Sq        RSS    AIC
## - DISTRICT  1 5.0410e+10 1.2424e+13 5391.3
## <none>                   1.2374e+13 5392.4
## - NUMBIDS   1 1.6343e+11 1.2537e+13 5393.2
## - RDLNGTH   1 1.8148e+11 1.2555e+13 5393.5
## - PCTTRAFF  1 1.8891e+11 1.2562e+13 5393.7
## + PCTASPH   1 2.7081e+10 1.2346e+13 5393.9
## - PCTBASE   1 2.1514e+11 1.2589e+13 5394.1
## + DAYSEST   1 1.3323e+10 1.2360e+13 5394.1
## + SUBCONT   1 5.0512e+09 1.2369e+13 5394.3
## + PCTMOBIL  1 2.3704e+09 1.2371e+13 5394.3
## + STATUS    1 1.8037e+07 1.2374e+13 5394.4
```

```
## + PCTEXCAV  1 1.7260e+06 1.2374e+13 5394.4
## - PCTSTRUC  1 3.6964e+11 1.2743e+13 5396.8
## - LBERATIO  1 3.2645e+12 1.5638e+13 5441.2
## - DOTEST    1 5.3642e+14 5.4879e+14 6213.3
##
## Step:  AIC=5391.25
## LOWBID ~ DOTEST + LBERATIO + NUMBIDS + RDLNGTH + PCTBASE + PCTSTRUC +
##     PCTTRAFF
##
##            Df  Sum of Sq        RSS    AIC
## <none>                   1.2424e+13 5391.3
## - NUMBIDS   1 1.4792e+11 1.2572e+13 5391.8
## - RDLNGTH   1 1.5686e+11 1.2581e+13 5392.0
## - PCTTRAFF  1 1.7080e+11 1.2595e+13 5392.2
## + DISTRICT  1 5.0410e+10 1.2374e+13 5392.4
## + PCTASPH   1 2.5560e+10 1.2398e+13 5392.8
## - PCTBASE   1 2.0907e+11 1.2633e+13 5392.9
## + DAYSEST   1 1.0806e+10 1.2413e+13 5393.1
## + SUBCONT   1 9.6063e+09 1.2414e+13 5393.1
## + PCTMOBIL  1 1.6821e+09 1.2422e+13 5393.2
## + STATUS    1 7.9704e+07 1.2424e+13 5393.3
## + PCTEXCAV  1 5.1890e+07 1.2424e+13 5393.3
## - PCTSTRUC  1 3.7626e+11 1.2800e+13 5395.7
## - LBERATIO  1 3.2966e+12 1.5721e+13 5440.3
## - DOTEST    1 5.5223e+14 5.6466e+14 6217.5
```

```
summary(road.model)
```

```
##
## Call:
## lm(formula = LOWBID ~ DOTEST + LBERATIO + NUMBIDS + RDLNGTH +
##     PCTBASE + PCTSTRUC + PCTTRAFF, data = FLAG2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1769335   -55472     6727    61641  1083827
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.801e+05  1.245e+05  -6.264 2.10e-09 ***
## DOTEST       9.095e-01  9.436e-03  96.384  < 2e-16 ***
## LBERATIO     8.297e+05  1.114e+05   7.447 2.47e-12 ***
## NUMBIDS     -1.315e+04  8.334e+03  -1.577   0.1162
## RDLNGTH      6.601e+03  4.064e+03   1.624   0.1058
## PCTBASE      3.104e+05  1.655e+05   1.875   0.0621 .
## PCTSTRUC     3.351e+05  1.332e+05   2.516   0.0126 *
## PCTTRAFF    -6.652e+05  3.924e+05  -1.695   0.0916 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 243800 on 209 degrees of freedom
## Multiple R-squared:  0.9823, Adjusted R-squared:  0.9818
## F-statistic:  1662 on 7 and 209 DF,  p-value: < 2.2e-16
```

(b) Interpret the β's

For every 1-unit increase in DOTEST would be multiplied by 9.095e-01
For every 1-unit increase in LBERATIO would be multiplied by 8.297e+05
For every 1-unit increase in NUMBIDS would be multiplied by -1.315e+04

For every 1-unit increase in RDLNGTH would be multiplied by 6.601e+03
For every 1-unit increase in PCTBASE would be multiplied by 3.104e+05
For every 1-unit increase in PCTSTRUC would be multiplied by 3.351e+05
For every 1-unit increase in PCTTRAFF would be multiplied by -6.652e+05

(c) The dangers of drawing inferences from a stepwise model: A large number of t-test's have been conducted leading to a high probability of making one or more Type I or Type II errors being that we have included some unimportant independent variables in the the model, and second that we have eliminated some important independent variables. Another danger is that we only tested a first order model/main effects and didn't include and higher order terms or interaction terms. We primarily use stepwise regression just to tell us which of the independent variable out of the masses are important to include into the model.

page 343 #6.9

```
# Best Subset
library(leaps)
yvar = c("LOWBID")
xvars = c("DOTEST","LBERATIO" ,"STATUS","DISTRICT" ,"NUMBIDS", "DAYSEST" , "RDLNGTH" , "PCTASPH" , "PCTBASE",
best.model = leaps(x = FLAG2[,xvars], y=FLAG2[,yvar], names=xvars, nbest=3, method="adjr2")
best.model$which # shows the T or F of variable inclusion in the model
```

```
##    DOTEST LBERATIO STATUS DISTRICT NUMBIDS DAYSEST RDLNGTH PCTASPH PCTBASE
## 1    TRUE    FALSE  FALSE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 1   FALSE    FALSE  FALSE    FALSE   FALSE    TRUE   FALSE   FALSE   FALSE
## 1   FALSE    FALSE  FALSE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 2    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 2    TRUE    FALSE   TRUE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 2    TRUE    FALSE  FALSE    FALSE    TRUE   FALSE   FALSE   FALSE   FALSE
## 3    TRUE     TRUE  FALSE    FALSE   FALSE    TRUE   FALSE   FALSE   FALSE
## 3    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 3    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 4    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE   FALSE   FALSE   FALSE
## 4    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE   FALSE    TRUE   FALSE
## 4    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE    TRUE    TRUE   FALSE
## 5    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE    TRUE    TRUE   FALSE
## 5    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE    TRUE   FALSE    TRUE
## 5    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE    TRUE   FALSE   FALSE
## 6    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE    TRUE    TRUE   FALSE
## 6    TRUE     TRUE  FALSE    FALSE    TRUE   FALSE    TRUE    TRUE   FALSE
## 6    TRUE     TRUE  FALSE    FALSE   FALSE   FALSE    TRUE   FALSE    TRUE
## 7    TRUE     TRUE  FALSE    FALSE    TRUE   FALSE    TRUE   FALSE    TRUE
## 7    TRUE     TRUE  FALSE    FALSE    TRUE   FALSE    TRUE    TRUE   FALSE
## 7    TRUE     TRUE  FALSE    FALSE    TRUE   FALSE    TRUE    TRUE   FALSE
## 8    TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE   FALSE    TRUE
## 8    TRUE     TRUE  FALSE    FALSE    TRUE   FALSE    TRUE    TRUE    TRUE
## 8    TRUE     TRUE  FALSE    FALSE    TRUE    TRUE    TRUE   FALSE    TRUE
## 9    TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE    TRUE    TRUE
## 9    TRUE     TRUE  FALSE     TRUE    TRUE    TRUE    TRUE   FALSE    TRUE
## 9    TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE   FALSE    TRUE
## 10   TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE    TRUE    TRUE
## 10   TRUE     TRUE  FALSE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 10   TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE    TRUE    TRUE
## 11   TRUE     TRUE  FALSE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 11   TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE    TRUE    TRUE
## 11   TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE    TRUE    TRUE
## 12   TRUE     TRUE  FALSE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 12   TRUE     TRUE  FALSE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 12   TRUE     TRUE  FALSE     TRUE    TRUE   FALSE    TRUE    TRUE    TRUE
## 13   TRUE     TRUE  FALSE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 13   TRUE     TRUE   TRUE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 13   TRUE     TRUE   TRUE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
## 14   TRUE     TRUE   TRUE     TRUE    TRUE    TRUE    TRUE    TRUE    TRUE
##    PCTEXCAV PCTMOBIL PCTSTRUC PCTTRAFF SUBCONT
## 1     FALSE    FALSE    FALSE    FALSE   FALSE
## 1     FALSE    FALSE    FALSE    FALSE   FALSE
## 1      TRUE    FALSE    FALSE    FALSE   FALSE
## 2     FALSE    FALSE    FALSE    FALSE   FALSE
## 2     FALSE    FALSE    FALSE    FALSE   FALSE
## 2     FALSE    FALSE    FALSE    FALSE   FALSE
## 3     FALSE    FALSE    FALSE    FALSE   FALSE
## 3     FALSE    FALSE    FALSE     TRUE   FALSE
## 3     FALSE    FALSE     TRUE    FALSE   FALSE
```

```
## 4     FALSE    FALSE     TRUE     TRUE   FALSE
## 4     FALSE    FALSE    FALSE     TRUE   FALSE
## 4     FALSE    FALSE    FALSE    FALSE   FALSE
## 5     FALSE    FALSE    FALSE     TRUE   FALSE
## 5     FALSE    FALSE     TRUE    FALSE   FALSE
## 5     FALSE    FALSE     TRUE     TRUE   FALSE
## 6      TRUE    FALSE    FALSE     TRUE   FALSE
## 6     FALSE    FALSE    FALSE     TRUE   FALSE
## 6     FALSE    FALSE     TRUE     TRUE   FALSE
## 7     FALSE    FALSE     TRUE     TRUE   FALSE
## 7      TRUE    FALSE    FALSE     TRUE   FALSE
## 7     FALSE    FALSE     TRUE     TRUE   FALSE
## 8     FALSE    FALSE     TRUE     TRUE   FALSE
## 8     FALSE    FALSE     TRUE     TRUE   FALSE
## 8     FALSE    FALSE     TRUE     TRUE   FALSE
## 9     FALSE    FALSE     TRUE     TRUE   FALSE
## 9     FALSE    FALSE     TRUE     TRUE   FALSE
## 9     FALSE    FALSE     TRUE     TRUE    TRUE
## 10     TRUE    FALSE     TRUE     TRUE   FALSE
## 10    FALSE    FALSE     TRUE     TRUE   FALSE
## 10    FALSE     TRUE     TRUE     TRUE   FALSE
## 11     TRUE    FALSE     TRUE     TRUE   FALSE
## 11     TRUE     TRUE     TRUE     TRUE   FALSE
## 11     TRUE    FALSE     TRUE     TRUE    TRUE
## 12     TRUE     TRUE     TRUE     TRUE   FALSE
## 12     TRUE    FALSE     TRUE     TRUE    TRUE
## 12     TRUE     TRUE     TRUE     TRUE    TRUE
## 13     TRUE     TRUE     TRUE     TRUE    TRUE
## 13     TRUE     TRUE     TRUE     TRUE   FALSE
## 13     TRUE    FALSE     TRUE     TRUE    TRUE
## 14     TRUE     TRUE     TRUE     TRUE    TRUE
```

```
best.model$adjr2
```

```
##  [1] 0.9761895 0.6373629 0.1435015 0.9811456 0.9771655 0.9767419 0.9813131
##  [8] 0.9812961 0.9812842 0.9814788 0.9814661 0.9814535 0.9816169 0.9815503
## [15] 0.9815495 0.9816543 0.9816467 0.9816281 0.9817574 0.9816795 0.9816793
## [22] 0.9817441 0.9817074 0.9816856 0.9816960 0.9816756 0.9816634 0.9816379
## [29] 0.9816239 0.9816185 0.9815670 0.9815644 0.9815533 0.9814915 0.9814821
## [36] 0.9814787 0.9814055 0.9814012 0.9813912 0.9813142
```

```
df = best.model$which
```

Yes, as evident from row `which(df[19,])` the "best subset" model did select the same 7 variables as chosen in the stepwise regression method. `apply(df, 1, which)`

```
##   DOTEST LBERATIO  NUMBIDS  RDLNGTH  PCTBASE PCTSTRUC PCTTRAFF
##        1        2        5        7        9       12       13
```

page 343 #6.10

(a) Stepwise regression (with stepwise selection) to find the "best predictors" of heat rate (y)

```
load("~/Desktop/depaul/CSC423/rdata/R/Exercises&Examples/GASTURBINE.Rdata")
head(GASTURBINE, n=5)
```

```
##        ENGINE SHAFTS   RPM CPRATIO INLETTEMP EXHTEMP AIRFLOW POWER
## 1 Traditional      1 27245     9.2      1134     602       7  1630
## 2 Traditional      1 14000    12.2       950     446      15  2726
## 3 Traditional      1 17384    14.8      1149     537      20  5247
## 4 Traditional      1 11085    11.8      1024     478      27  6726
## 5 Traditional      1 14045    13.2      1149     553      29  7726
##   HEATRATE  LHV ISOWORK
## 1    14622 24.6  232.86
## 2    13196 27.3  181.73
## 3    11948 30.1  262.35
## 4    11289 31.9  249.11
## 5    11964 30.1  266.41
```

```
tail(GASTURBINE, n=5)
```

```
##         ENGINE SHAFTS   RPM CPRATIO INLETTEMP EXHTEMP AIRFLOW POWER
## 63 Aeroderiv        2 18910    14.0      1066     532       8  1845
## 64 Aeroderiv        3  3600    35.0      1288     448     152 57930
## 65 Aeroderiv        3  3600    20.0      1160     456      84 25600
## 66 Aeroderiv        2 16000    10.6      1232     560      14  3815
## 67 Aeroderiv        1 14600    13.4      1077     536      20  4942
##    HEATRATE  LHV ISOWORK
## 63    12766 28.2  230.63
## 64     8714 41.3  341.64
## 65     9469 38.0  304.76
## 66    11948 30.1  272.50
## 67    12414 29.0  247.10
```

```
# stepwise/stepwise regression
GASTURBINE <- na.omit(GASTURBINE) # step() requires removal of missing data before if any
full.gas.model <- lm(HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + EXHTEMP + AIRFLOW + POWER + LHV +
gas.model <- step(full.gas.model, direction="both")
```

```
## Start:  AIC=694.84
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + EXHTEMP +
##     AIRFLOW + POWER + LHV + ISOWORK
##
##             Df Sum of Sq     RSS    AIC
## - EXHTEMP    1        24 1494459 692.84
## - POWER      1      1173 1495609 692.89
## - ISOWORK    1     21683 1516119 693.81
## - AIRFLOW    1     44256 1538692 694.80
## <none>                   1494435 694.84
## - SHAFTS     1     66628 1561063 695.76
## - INLETTEMP  1    104046 1598481 697.35
## - ENGINE     2    321528 1815963 703.90
## - CPRATIO    1    319879 1814314 705.84
## - RPM        1    358280 1852716 707.24
## - LHV        1   5653285 7147721 797.70
##
```

```
## Step:  AIC=692.84
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + AIRFLOW +
##     POWER + LHV + ISOWORK
##
##             Df Sum of Sq      RSS    AIC
## - POWER      1      1150  1495609 690.89
## - ISOWORK    1     29435  1523894 692.15
## <none>                    1494459 692.84
## - AIRFLOW    1     47493  1541952 692.94
## - SHAFTS     1     66669  1561128 693.77
## + EXHTEMP    1        24  1494435 694.84
## - INLETTEMP  1    195895  1690354 699.10
## - ENGINE     2    322091  1816550 701.92
## - RPM        1    359684  1854143 705.29
## - CPRATIO    1    391431  1885890 706.43
## - LHV        1   9370579 10865038 823.76
##
## Step:  AIC=690.89
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + AIRFLOW +
##     LHV + ISOWORK
##
##             Df Sum of Sq      RSS    AIC
## - ISOWORK    1     41215  1536824 690.72
## <none>                    1495609 690.89
## - SHAFTS     1     71532  1567141 692.02
## + POWER      1      1150  1494459 692.84
## + EXHTEMP    1         1  1495609 692.89
## - INLETTEMP  1    204758  1700367 697.49
## - ENGINE     2    376822  1872431 701.95
## - RPM        1    455572  1951181 706.71
## - CPRATIO    1    480939  1976549 707.58
## - AIRFLOW    1    956542  2452151 722.02
## - LHV        1  10230013 11725622 826.86
##
## Step:  AIC=690.72
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + AIRFLOW +
##     LHV
##
##             Df Sum of Sq      RSS    AIC
## <none>                    1536824 690.72
## + ISOWORK    1     41215  1495609 690.89
## - SHAFTS     1     56239  1593063 691.12
## + POWER      1     12930  1523894 692.15
## + EXHTEMP    1      8952  1527873 692.32
## - INLETTEMP  1    168769  1705593 695.70
## - ENGINE     2    427892  1964716 703.17
## - CPRATIO    1    453650  1990474 706.05
## - RPM        1    585287  2122112 710.34
## - AIRFLOW    1   1434857  2971682 732.90
## - LHV        1  13752704 15289529 842.65
```

```
summary(gas.model)
```

```
##
## Call:
## lm(formula = HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP +
##     AIRFLOW + LHV, data = GASTURBINE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -428.83  -96.99  -19.01   87.82  362.67
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.447e+04  5.176e+02  47.267  < 2e-16 ***
## ENGINEAeroderiv   -2.545e+02  1.277e+02  -1.993 0.050921 .
## ENGINETraditional -3.196e+02  7.969e+01  -4.011 0.000175 ***
## SHAFTS             1.173e+02  8.051e+01   1.457 0.150547
## RPM                2.861e-02  6.088e-03   4.700 1.65e-05 ***
## CPRATIO            4.121e+01  9.959e+00   4.138 0.000115 ***
## INLETTEMP         -1.085e+00  4.301e-01  -2.524 0.014372 *
## AIRFLOW            1.207e+00  1.641e-01   7.359 7.31e-10 ***
## LHV               -3.980e+02  1.747e+01 -22.782  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 162.8 on 58 degrees of freedom
## Multiple R-squared:  0.9908, Adjusted R-squared:  0.9896
## F-statistic: 784.8 on 8 and 58 DF,  p-value: < 2.2e-16
```

(b) Stepwise with backward elimination

```
gas.model <- step(full.gas.model, direction="backward")
```

```
## Start:  AIC=694.84
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + EXHTEMP +
##     AIRFLOW + POWER + LHV + ISOWORK
##
##             Df Sum of Sq     RSS    AIC
## - EXHTEMP    1        24 1494459 692.84
## - POWER      1      1173 1495609 692.89
## - ISOWORK    1     21683 1516119 693.81
## - AIRFLOW    1     44256 1538692 694.80
## <none>                   1494435 694.84
## - SHAFTS     1     66628 1561063 695.76
## - INLETTEMP  1    104046 1598481 697.35
## - ENGINE     2    321528 1815963 703.90
## - CPRATIO    1    319879 1814314 705.84
## - RPM        1    358280 1852716 707.24
## - LHV        1   5653285 7147721 797.70
##
## Step:  AIC=692.84
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + AIRFLOW +
##     POWER + LHV + ISOWORK
##
##             Df Sum of Sq      RSS    AIC
## - POWER      1      1150  1495609 690.89
## - ISOWORK    1     29435  1523894 692.15
## <none>                   1494459 692.84
## - AIRFLOW    1     47493  1541952 692.94
## - SHAFTS     1     66669  1561128 693.77
## - INLETTEMP  1    195895  1690354 699.10
## - ENGINE     2    322091  1816550 701.92
## - RPM        1    359684  1854143 705.29
## - CPRATIO    1    391431  1885890 706.43
## - LHV        1   9370579 10865038 823.76
##
## Step:  AIC=690.89
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + AIRFLOW +
```

```
##     LHV + ISOWORK
##
##             Df Sum of Sq      RSS    AIC
## - ISOWORK    1     41215  1536824 690.72
## <none>                    1495609 690.89
## - SHAFTS     1     71532  1567141 692.02
## - INLETTEMP  1    204758  1700367 697.49
## - ENGINE     2    376822  1872431 701.95
## - RPM        1    455572  1951181 706.71
## - CPRATIO    1    480939  1976549 707.58
## - AIRFLOW    1    956542  2452151 722.02
## - LHV        1  10230013 11725622 826.86
##
## Step:  AIC=690.72
## HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP + AIRFLOW +
##     LHV
##
##             Df Sum of Sq      RSS    AIC
## <none>                    1536824 690.72
## - SHAFTS     1     56239  1593063 691.12
## - INLETTEMP  1    168769  1705593 695.70
## - ENGINE     2    427892  1964716 703.17
## - CPRATIO    1    453650  1990474 706.05
## - RPM        1    585287  2122112 710.34
## - AIRFLOW    1   1434857  2971682 732.90
## - LHV        1  13752704 15289529 842.65
```

```
summary(gas.model)
```

```
##
## Call:
## lm(formula = HEATRATE ~ ENGINE + SHAFTS + RPM + CPRATIO + INLETTEMP +
##     AIRFLOW + LHV, data = GASTURBINE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -428.83  -96.99  -19.01   87.82  362.67
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)        2.447e+04  5.176e+02  47.267  < 2e-16 ***
## ENGINEAeroderiv   -2.545e+02  1.277e+02  -1.993 0.050921 .
## ENGINETraditional -3.196e+02  7.969e+01  -4.011 0.000175 ***
## SHAFTS             1.173e+02  8.051e+01   1.457 0.150547
## RPM                2.861e-02  6.088e-03   4.700 1.65e-05 ***
## CPRATIO            4.121e+01  9.959e+00   4.138 0.000115 ***
## INLETTEMP         -1.085e+00  4.301e-01  -2.524 0.014372 *
## AIRFLOW            1.207e+00  1.641e-01   7.359 7.31e-10 ***
## LHV               -3.980e+02  1.747e+01 -22.782  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 162.8 on 58 degrees of freedom
## Multiple R-squared:  0.9908, Adjusted R-squared:  0.9896
## F-statistic: 784.8 on 8 and 58 DF,  p-value: < 2.2e-16
```

(c) All-possible-regressions-selection / "best subset"

```r
library(leaps)
y = c("HEATRATE")
x = c("SHAFTS","RPM","CPRATIO","INLETTEMP","EXHTEMP","AIRFLOW","POWER","HEATRATE", "LHV","ISOWORK")
best.model <- leaps(x = GASTURBINE[,x], y=GASTURBINE[,y], names=x, nbest=3, method="adjr2")
best.model$which # shows the T or F of variable inclusion in the model
```

```
##    SHAFTS   RPM CPRATIO INLETTEMP EXHTEMP AIRFLOW POWER HEATRATE   LHV
## 1   FALSE FALSE   FALSE     FALSE   FALSE   FALSE FALSE     TRUE FALSE
## 1   FALSE FALSE   FALSE     FALSE   FALSE   FALSE FALSE    FALSE  TRUE
## 1   FALSE FALSE   FALSE     FALSE   FALSE   FALSE FALSE    FALSE FALSE
## 2   FALSE  TRUE   FALSE     FALSE   FALSE   FALSE FALSE     TRUE FALSE
## 2   FALSE FALSE    TRUE     FALSE   FALSE   FALSE FALSE     TRUE FALSE
## 2   FALSE FALSE   FALSE     FALSE   FALSE   FALSE FALSE     TRUE  TRUE
## 3   FALSE  TRUE    TRUE     FALSE   FALSE   FALSE FALSE     TRUE FALSE
## 3    TRUE  TRUE   FALSE     FALSE   FALSE   FALSE FALSE     TRUE FALSE
## 3   FALSE  TRUE   FALSE      TRUE   FALSE   FALSE FALSE     TRUE FALSE
## 4   FALSE  TRUE    TRUE      TRUE   FALSE   FALSE FALSE     TRUE FALSE
## 4    TRUE  TRUE   FALSE      TRUE   FALSE   FALSE FALSE     TRUE FALSE
## 4    TRUE  TRUE    TRUE     FALSE   FALSE   FALSE FALSE     TRUE FALSE
## 5    TRUE  TRUE    TRUE      TRUE   FALSE   FALSE FALSE     TRUE FALSE
## 5    TRUE  TRUE   FALSE      TRUE    TRUE   FALSE FALSE     TRUE FALSE
## 5   FALSE  TRUE    TRUE      TRUE    TRUE   FALSE FALSE     TRUE FALSE
## 6    TRUE  TRUE    TRUE      TRUE    TRUE   FALSE FALSE     TRUE FALSE
## 6    TRUE  TRUE    TRUE      TRUE   FALSE   FALSE  TRUE     TRUE FALSE
## 6    TRUE  TRUE   FALSE      TRUE    TRUE   FALSE FALSE     TRUE FALSE
## 7    TRUE  TRUE    TRUE      TRUE    TRUE   FALSE FALSE     TRUE FALSE
## 7    TRUE  TRUE    TRUE      TRUE    TRUE    TRUE FALSE     TRUE FALSE
## 7   FALSE  TRUE    TRUE      TRUE    TRUE    TRUE  TRUE     TRUE FALSE
## 8    TRUE  TRUE    TRUE      TRUE    TRUE    TRUE  TRUE     TRUE FALSE
## 8   FALSE  TRUE    TRUE      TRUE    TRUE    TRUE  TRUE     TRUE FALSE
## 8    TRUE  TRUE   FALSE      TRUE    TRUE    TRUE  TRUE     TRUE FALSE
## 9    TRUE  TRUE    TRUE      TRUE    TRUE    TRUE  TRUE     TRUE  TRUE
## 9    TRUE  TRUE    TRUE      TRUE    TRUE    TRUE  TRUE     TRUE FALSE
## 9    TRUE  TRUE   FALSE      TRUE    TRUE    TRUE  TRUE     TRUE  TRUE
## 10   TRUE  TRUE    TRUE      TRUE    TRUE    TRUE  TRUE     TRUE  TRUE
##    ISOWORK
## 1    FALSE
## 1    FALSE
## 1     TRUE
## 2    FALSE
## 2    FALSE
## 2    FALSE
## 3    FALSE
## 3    FALSE
## 3    FALSE
## 4    FALSE
## 4    FALSE
## 4    FALSE
## 5    FALSE
## 5    FALSE
## 5    FALSE
## 6    FALSE
## 6    FALSE
## 6     TRUE
## 7     TRUE
## 7    FALSE
## 7    FALSE
## 8    FALSE
## 8     TRUE
```

```
## 8     TRUE
## 9    FALSE
## 9     TRUE
## 9     TRUE
## 10    TRUE
```

```
apply(best.model$which, 1, which) # consise print-out of each best model
```

```
## $`1`
## HEATRATE
##        8
##
## $`1`
## LHV
##   9
##
## $`1`
## ISOWORK
##      10
##
## $`2`
##      RPM HEATRATE
##        2        8
##
## $`2`
##  CPRATIO HEATRATE
##        3        8
##
## $`2`
## HEATRATE      LHV
##        8        9
##
## $`3`
##      RPM  CPRATIO HEATRATE
##        2        3        8
##
## $`3`
##   SHAFTS      RPM HEATRATE
##        1        2        8
##
## $`3`
##       RPM INLETTEMP  HEATRATE
##         2         4         8
##
## $`4`
##       RPM   CPRATIO INLETTEMP  HEATRATE
##         2         3         4         8
##
## $`4`
##    SHAFTS       RPM INLETTEMP  HEATRATE
##         1         2         4         8
##
## $`4`
##   SHAFTS      RPM  CPRATIO HEATRATE
##        1        2        3        8
##
## $`5`
##    SHAFTS       RPM   CPRATIO INLETTEMP  HEATRATE
##         1         2         3         4         8
##
```

```
## $`5`
##    SHAFTS       RPM INLETTEMP   EXHTEMP  HEATRATE
##         1         2         4         5         8
##
## $`5`
##       RPM   CPRATIO INLETTEMP   EXHTEMP  HEATRATE
##         2         3         4         5         8
##
## $`6`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP  HEATRATE
##         1         2         3         4         5         8
##
## $`6`
##    SHAFTS       RPM   CPRATIO INLETTEMP     POWER  HEATRATE
##         1         2         3         4         7         8
##
## $`6`
##    SHAFTS       RPM INLETTEMP   EXHTEMP  HEATRATE   ISOWORK
##         1         2         4         5         8        10
##
## $`7`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP  HEATRATE   ISOWORK
##         1         2         3         4         5         8        10
##
## $`7`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW  HEATRATE
##         1         2         3         4         5         6         8
##
## $`7`
##       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW     POWER  HEATRATE
##         2         3         4         5         6         7         8
##
## $`8`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW     POWER
##         1         2         3         4         5         6         7
##  HEATRATE
##         8
##
## $`8`
##       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW     POWER  HEATRATE
##         2         3         4         5         6         7         8
##   ISOWORK
##        10
##
## $`8`
##    SHAFTS       RPM INLETTEMP   EXHTEMP   AIRFLOW     POWER  HEATRATE
##         1         2         4         5         6         7         8
##   ISOWORK
##        10
##
## $`9`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW     POWER
##         1         2         3         4         5         6         7
##  HEATRATE       LHV
##         8         9
##
## $`9`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW     POWER
##         1         2         3         4         5         6         7
##  HEATRATE   ISOWORK
```

```
##         8        10
##
## $`9`
##    SHAFTS       RPM INLETTEMP   EXHTEMP   AIRFLOW     POWER  HEATRATE
##         1         2         4         5         6         7         8
##       LHV   ISOWORK
##         9        10
##
## $`10`
##    SHAFTS       RPM   CPRATIO INLETTEMP   EXHTEMP   AIRFLOW     POWER
##         1         2         3         4         5         6         7
##  HEATRATE       LHV   ISOWORK
##         8         9        10
```

```
best.model$adjr2
```

```
##  [1] 1.0000000 0.9697587 0.7447262 1.0000000 1.0000000 1.0000000 1.0000000
##  [8] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [15] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
## [22] 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
```

(d) Which Independent variable are consistency selected out of the previous results:
Stepwise: EXHTEMP, POWER, ISOWORK
Backwards elimination: ENGINE, SHAFTS, RPM, CPRATIO, INLETTEMP, AIRFLOW, LHV
best subset: POWER, LHV, CPRATIO, RPM

(e) I would use the previous results from each of the model selection techniques to select the most important variables (that also exist most frequently in across each model) based on each final outcome of the elimination techniques and take into account their p-value and AIC score.

page 377 #7.2

(a) The problems that exist when multicollinearity exist are: high correlations among independent variables increase the likelihood of rounding errors in the calculations of the beta estimate from the underlying matrix operation from the computers difficulty in inverting the information matrix. Multicollinearity can also cause misleading and confusing results of the signs of the parameter estimates than what is expected.

(b) You can detect multicollinearity with several methods:

- Calculate the coefficient of correlation between each pair of independent variables in the model; if one or more of the r values is close to 1 or -1, the variables are highly correlated and a severe milticollinearity problem may be present.
- Non-significant t-test's for the individual beta parameters when the F-test for overall model adequacy is significant and estimates with opposite signs from what is expected.
- Calculation of the variance inflation factor (vif) for the individual factors.

(c) Solutions to address multicollinearity are:

- Drop one or more of the correlated independent variables
- if the correlate variables are kept in the model, avoid making inferences about the individual parameters.
- Use a designed experiment

page 377 #7.5

(a) No extreme multicollinearity exists according in the correlation matrix

(b) According to the multiple regression output on page 190, There is evidence for multicollinearity: Significant F-test, with p-value < 0.001 when there are multiple non-significant (high p-values) independent x variables.

page 378 #7.10

(a) fit a first-order model to the data, E(y) = 5.71059 + 0.62597 LIVEWT

```
load("~/Desktop/depaul/CSC423/rdata/R/Exercises&Examples/STEERS.Rdata")
head(STEERS)
```

```
##   LIVEWT DRESSWT
## 1    420     280
## 2    380     250
## 3    480     310
## 4    340     210
## 5    450     290
## 6    460     280
```

```
DRESSWT = STEERS$DRESSWT
LIVEWT = STEERS$LIVEWT
steer.model <- lm(DRESSWT ~ LIVEWT, data=STEERS)
summary(steer.model)
```

```
##
## Call:
## lm(formula = DRESSWT ~ LIVEWT, data = STEERS)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.656  -4.877   2.603   3.824  11.382
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.71059   26.31520   0.217    0.834
## LIVEWT       0.62597    0.06331   9.887 2.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.303 on 7 degrees of freedom
## Multiple R-squared:  0.9332, Adjusted R-squared:  0.9236
## F-statistic: 97.75 on 1 and 7 DF,  p-value: 2.306e-05
```

(b) 95% prediction interval for the dressed weight of a 300 pound steer

```
predict(steer.model, newdata=data.frame(DRESSWT = 300), interval="prediction", level=0.95)
```

```
## Warning: 'newdata' had 1 row but variables found have 9 rows
```

```
##        fit      lwr      upr
## 1 268.6176 247.8971 289.3381
## 2 243.5788 222.2892 264.8684
## 3 306.1757 283.1984 329.1530
## 4 218.5401 195.1119 241.9682
## 5 287.3966 265.9846 308.8087
## 6 293.6563 271.8125 315.5002
## 7 274.8773 254.0309 295.7236
## 8 237.3191 215.6297 259.0085
## 9 249.8385 228.8493 270.8277
```

(c) Yes, The interval is tight as being a prediction interval and with the single x-variable of 300 the interval is a close approximation as to have minimal complaints from customers.

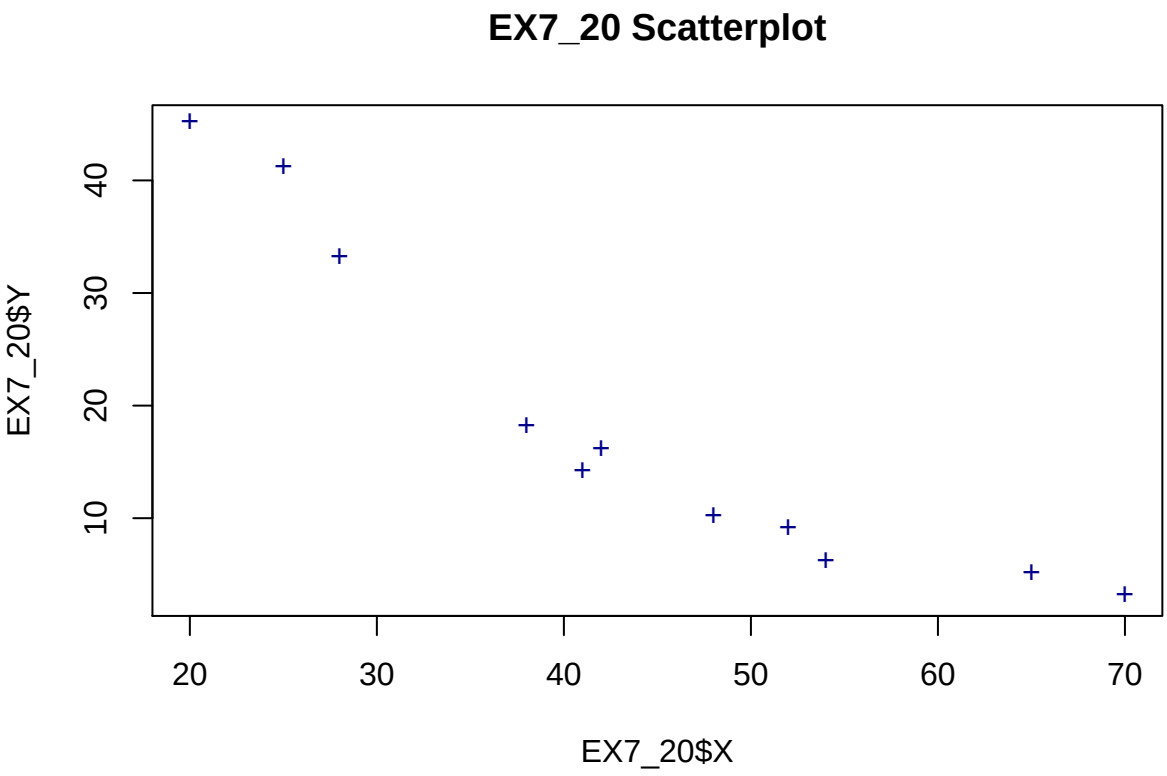
page 381 #7.20

(a) Scatter plot of points. There seems to be a negative linear relationship between X and Y possibly even curvilinear sloping up.

```
load("~/Desktop/depaul/CSC423/rdata/R/Exercises&Examples/EX7_20.Rdata")
EX7_20
```

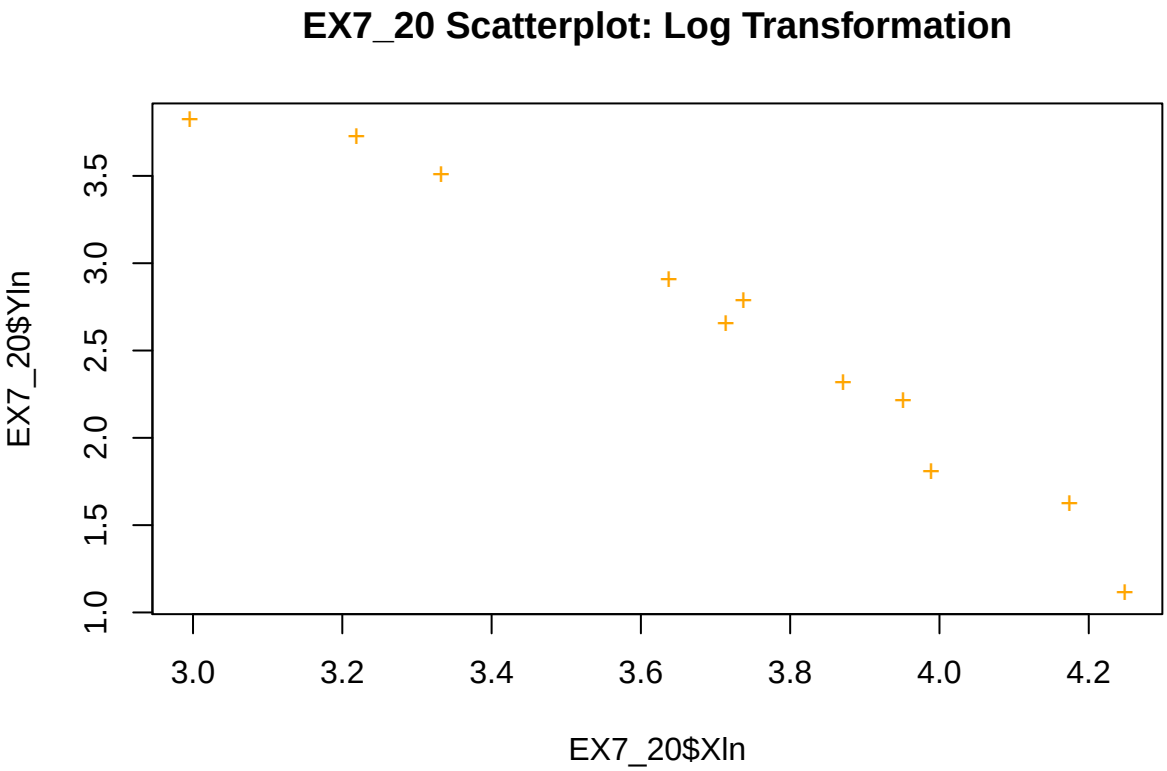
```
##     X  Y
## 1  54  6
## 2  42 16
## 3  28 33
## 4  38 18
## 5  25 41
## 6  70  3
## 7  48 10
## 8  41 14
## 9  20 45
## 10 52  9
## 11 65  5
```

```
plot(EX7_20$X, EX7_20$Y, main="EX7_20 Scatterplot", pch="+", col="darkblue")
```



(b) calculate ln x and ln y, then plot the log transformed data on another scatter plot. This plot shows similarly to the other plot a negative linear relationship and possibly a curvilinear sloping down.

```
EX7_20$Xln <- log(EX7_20$X)
EX7_20$Yln <- log(EX7_20$Y)
plot(EX7_20$Xln, EX7_20$Yln, main="EX7_20 Scatterplot: Log Transformation", pch="+", col="orange")
```



(c) Fit transformed data to model equation. The F-statistic: 180.7 on 1 and 9 DF, p-value: 2.911e-07 is high with a significant p-value above the alpha = 0.05.

```
Yln = EX7_20$Yln
Xln = EX7_20$Xln
logs.on.logs <- lm(Yln ~ Xln, data=EX7_20)
summary(logs.on.logs)
```

```
##
## Call:
## lm(formula = Yln ~ Xln, data = EX7_20)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.32942 -0.07912  0.06168  0.11249  0.24640
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.6364     0.6028   17.64 2.73e-08 ***
## Xln          -2.1699     0.1614  -13.44 2.91e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2021 on 9 degrees of freedom
## Multiple R-squared:  0.9526, Adjusted R-squared:  0.9473
## F-statistic: 180.7 on 1 and 9 DF,  p-value: 2.911e-07
```

(d) prediction of y, when x=30

```
exp(predict(logs.on.logs, newdata = data.frame(Xln = 30), interval="prediction", level=0.95))
```

```
##            fit          lwr          upr
## 1 2.231857e-24 1.497426e-28 3.326499e-20
```

page 381 #7.21

(a) Coefficient of correlation between y and x1. Since the value is so low (0.0025) there seems to be no evidence of a linear relationship between y and x1

```
load("~/Desktop/depaul/CSC423/rdata/R/Exercises&Examples/HAMILTON.Rdata")
head(HAMILTON)
```

```
##     X1   X2     Y
## 1 22.3 96.6 123.7
## 2 25.7 89.4 126.6
## 3 38.7 44.0 120.0
## 4 31.0 66.4 119.3
## 5 33.9 49.1 110.6
## 6 28.3 85.2 130.3
```

```
cor(HAMILTON$X1, HAMILTON$Y)
```

```
## [1] 0.002497966
```

(b) Coefficient of correlation between y and x2. Since the value is so low (0.43) there seems to be no evidence of a linear relationship between y and x2

```
cor(HAMILTON$X2, HAMILTON$Y)
```

```
## [1] 0.4340688
```

(c) Based on the previous results, I do not think that the model will be a useful predictor of sale price

(d) Fit the model: y = -45.154136 + 3.097008 X1 + 1.031859 X2
The R^2 value & adjusted R^2 is very high (0.9998) and the F-statistic has a significant p-value so that would imply that the model disagrees with the findings in the previous answer in part c.

```
ham <- lm(Y ~ X1 + X2, data=HAMILTON)
summary(ham)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2, data = HAMILTON)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.13632 -0.09452 -0.02279  0.08629  0.16325
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -45.154136   0.611418  -73.85   <2e-16 ***
## X1            3.097008   0.012274  252.31   <2e-16 ***
## X2            1.031859   0.003684  280.08   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1072 on 12 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 3.922e+04 on 2 and 12 DF,  p-value: < 2.2e-16
```

(e) Coefficient of correlation between x1 and x2. The result is close to -1 implying there is high correlation between x1 and x2.

```
cor(HAMILTON$X1, HAMILTON$X2)
```

```
## [1] -0.8997765
```

(f) I would not recommend this strategy for this example. The confidence for E(y) and prediction intervals for y generally remain unaffected as long as the values of the independent variables used to predict y follow the same pattern of multicollinearity exhibited in the sample data. The x1 and x2 variables may not be very redundant.

page 381 #7.22

(a) Independent variables that are moderately or highly correlated:

(5) Foreign Status with (3) Race -0.515
(6) Years in graduate program with (7) Year GRE was taken -0.602

(b) If those independent variables are left in the model, you could observe unreliable beta estimates and incorrect signs.