Exam1PraacticeSolutions

Question #1

```r
library(car)
```

```
## Loading required package: carData
```

```r
bodyfat <- read.csv("~/Dropbox/STAT512/Assigns/Assign4/BodyFat.csv")
#A
Model1 <- lm(BodyFat ~ ., data = bodyfat)
vif(Model1)
```

```
##  Triceps    Thigh   Midarm
## 708.8429 564.3434 104.6060
```

```r
Model2 <- lm(Triceps ~ Thigh + Midarm, data = bodyfat)
1/(1-summary(Model2)$r.squared)
```

```
## [1] 708.8429
```

```r
#B
Model3 <- lm(BodyFat ~ Triceps + Midarm, data = bodyfat)
vif(Model3)
```

```
##  Triceps   Midarm
## 1.265118 1.265118
```

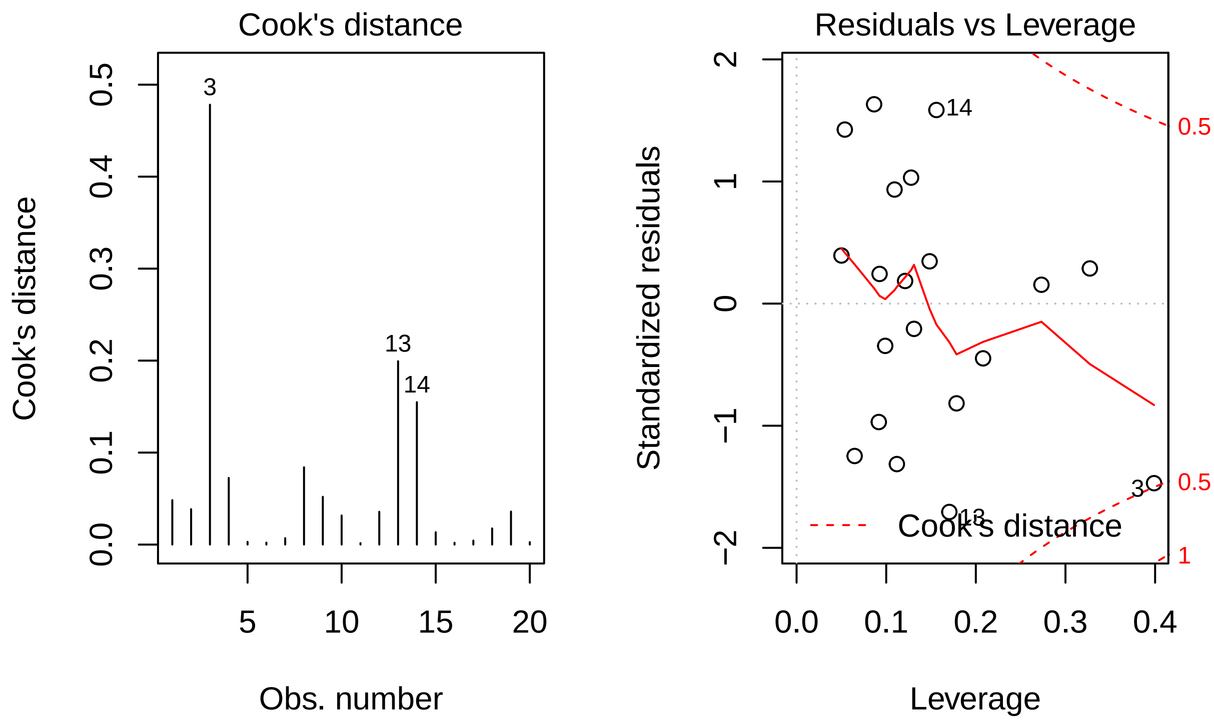
```r
#The above VIF's are the same because,
1/(1-cor(bodyfat$Triceps,bodyfat$Midarm)^2)
```

```
## [1] 1.265118
```

```r
#is the same as due to commutative property of only have 2 predictors
1/(1- summary(lm(Midarm ~ Triceps, data = bodyfat))$r.squared)
```

```
## [1] 1.265118
```

```r
par(mfrow = c(1,2))
plot(Model3, which = c(4,5))
```




```r
influence.measures(Model3)
```

```
## Influence measures of
##   lm(formula = BodyFat ~ Triceps + Midarm, data = bodyfat) :
##
##     dfb.1_  dfb.Trcp dfb.Mdrm   dffit cov.r  cook.d    hat inf
## 1  -0.01417  0.308728 -0.21515 -0.3768 1.295 0.04831 0.1785
## 2   0.00575 -0.075532  0.08371  0.3514 0.865 0.03851 0.0538
## 3   1.05631  0.052521 -1.05722 -1.2439 1.327 0.47826 0.3988   *
## 4   0.30021 -0.168675 -0.20053 -0.4774 0.979 0.07251 0.1119
## 5  -0.01333 -0.074759  0.06649  0.0919 1.643 0.00299 0.2732   *
## 6  -0.05045 -0.030637  0.06128 -0.0780 1.370 0.00215 0.1310
## 7  -0.01524  0.114639 -0.05289  0.1408 1.379 0.00697 0.1484
## 8  -0.29157  0.066364  0.27028  0.5306 0.787 0.08408 0.0865
## 9   0.33347 -0.023774 -0.26269  0.3956 1.133 0.05195 0.1278
## 10 -0.17910 -0.102926  0.20727 -0.3076 1.113 0.03165 0.0917
## 11 -0.03196  0.042322  0.00616  0.0668 1.356 0.00158 0.1211
## 12 -0.06659  0.236274 -0.07068  0.3259 1.150 0.03568 0.1093
## 13 -0.70148  0.378629  0.34248 -0.8242 0.823 0.19936 0.1704
## 14  0.07310 -0.579789  0.36482  0.7163 0.879 0.15486 0.1560
## 15  0.16720 -0.117851 -0.06671  0.1953 1.757 0.01344 0.3272   *
## 16 -0.03801  0.033493  0.01903  0.0754 1.308 0.00201 0.0926
## 17 -0.03646 -0.065715  0.06818 -0.1116 1.303 0.00438 0.0989
## 18 -0.06520 -0.171702  0.16163 -0.2242 1.462 0.01760 0.2081
## 19 -0.11269  0.153656 -0.03214 -0.3343 0.961 0.03595 0.0648
## 20  0.01396 -0.000594 -0.00237  0.0881 1.228 0.00273 0.0501
```

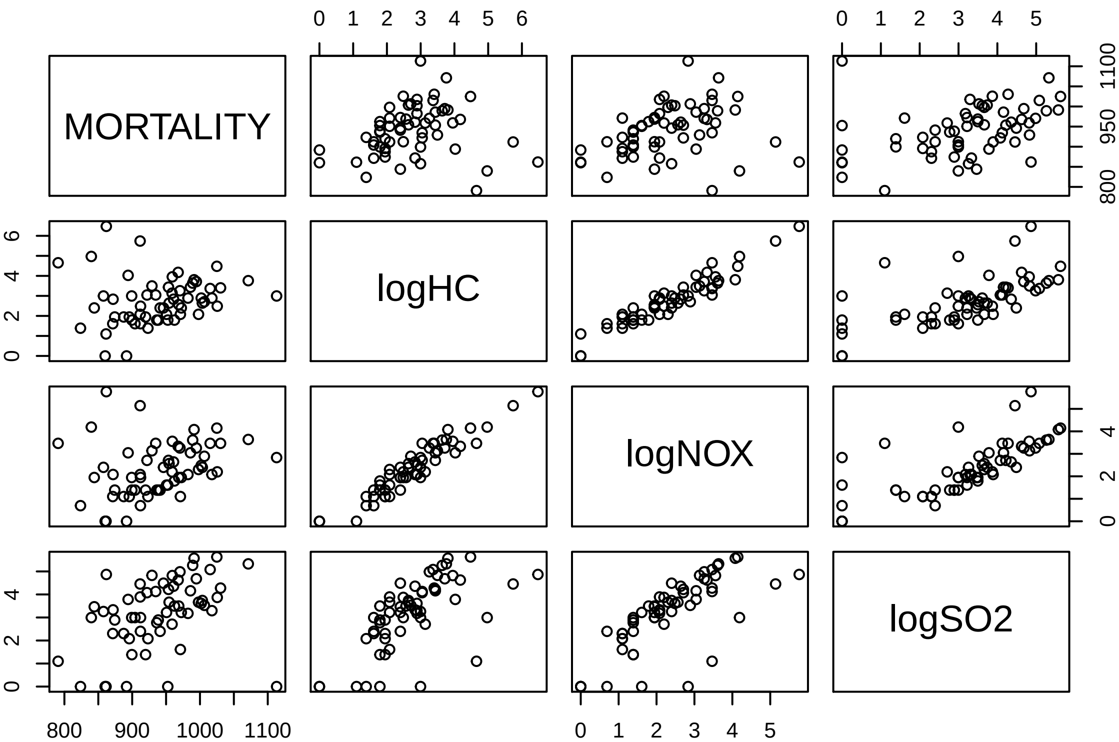
Rules of Thumb: |DFBETA| > 2/sqrt(n) = 0.447 |DFFITS| = 2*sqrt((k+1)/n) = 0.774 Cooks D > 1 Obs#3 is influential for DFBETAS and DFFITS, but not Cook's distance

Question #2

```r
library(MuMIn)
library(car)
Mortality <- read.csv("~/Dropbox/STAT512/ExamMaterial/Exam1/Mortality.csv", row.names=1)
Mortality$logHC <- log(Mortality$HC)
Mortality$logNOX <- log(Mortality$NOX)
Mortality$logSO2 <- log(Mortality$SO2)
colnames(Mortality)
```

```
##  [1] "MORTALITY" "PRECIP"    "HUMIDITY"  "JANTEMP"   "JULYTEMP"
##  [6] "OVER65"    "HOUSE"     "EDUC"      "SOUND"     "DENSITY"
## [11] "NONWHITE"  "WHITECOL"  "POOR"      "HC"        "NOX"
## [16] "SO2"       "logHC"     "logNOX"    "logSO2"
```

```r
#A
pairs(Mortality[,c(1, 17:19)])
```




```r
cor(Mortality[,c(1, 17:19)])
```

```
##           MORTALITY     logHC    logNOX    logSO2
## MORTALITY 1.0000000 0.1507561 0.2919995 0.4031280
## logHC     0.1507561 1.0000000 0.9474952 0.6408135
## logNOX    0.2919995 0.9474952 1.0000000 0.7328074
## logSO2    0.4031280 0.6408135 0.7328074 1.0000000
```

```r
#B
FullModel1 <- lm(MORTALITY ~ PRECIP + HUMIDITY + JANTEMP +
                   JULYTEMP + OVER65 + HOUSE + EDUC + SOUND +
                   DENSITY + NONWHITE + WHITECOL + POOR, data = Mortality)
options(na.action = "na.fail")
```

```r
AllSubsets1 <- dredge(FullModel1, rank="AIC")
```

```
## Fixed term is "(Intercept)"
```

The Final model has these seven predictors: DENSITY EDUC HOUSE JANTEMP JULYTEMP NONWHITE PRECIP

C

```r
FullModel2 <- lm(MORTALITY ~ DENSITY + EDUC + HOUSE + JANTEMP
                 + JULYTEMP + NONWHITE + PRECIP + logHC + logNOX + logSO2, data = Mortality)
AllSubsets2 <- dredge(FullModel2,
                      fixed = c("DENSITY", "EDUC", "HOUSE", "JANTEMP",
                             "JULYTEMP","NONWHITE","PRECIP"),rank="AIC")
```

```
## Fixed terms are "DENSITY", "EDUC", "HOUSE", "JANTEMP", "JULYTEMP", "NONWHITE", "PRECIP" and "(Interc
```

```r
AllSubsets2
```

```
## Global model call: lm(formula = MORTALITY ~ DENSITY + EDUC + HOUSE + JANTEMP + JULYTEMP +
##     NONWHITE + PRECIP + logHC + logNOX + logSO2, data = Mortality)
## ---
## Model selection table
##   (Intrc)    DENSI   EDUC  HOUSE  JANTE   JULYT   logHC lgNOX  lgSO2 NONWH
## 4    1353 0.003128 -13.76 -55.03 -1.980 -2.1730 -28.480 37.65        5.032
## 8    1460 0.003873 -16.16 -69.76 -2.548 -2.3700 -29.190 45.53 -7.793 5.254
## 3    1229 0.002266 -16.04 -51.90 -2.265 -0.5672         15.66        4.500
## 7    1325 0.002939 -18.32 -65.54 -2.801 -0.7131         22.49 -7.252 4.694
## 2    1322 0.003490 -18.78 -59.26 -2.306 -0.9799  10.730              4.955
## 1    1525 0.005555 -20.03 -70.03 -2.123 -2.7280                      5.892
## 5    1339 0.003809 -16.70 -52.19 -1.731 -1.8270               6.007 5.228
## 6    1295 0.003266 -17.84 -55.00 -2.132 -0.9944   8.683       2.126 4.899
##   PRECI df   logLik   AIC delta weight
## 4 1.766 11 -287.984 598.0  0.00  0.411
## 8 1.903 12 -287.140 598.3  0.31  0.352
## 3 2.067 10 -290.177 600.4  2.39  0.125
## 7 2.201 11 -289.499 601.0  3.03  0.090
## 2 1.807 10 -292.917 605.8  7.87  0.008
## 1 1.276  9 -294.284 606.6  8.60  0.006
## 5 1.450 10 -293.368 606.7  8.77  0.005
## 6 1.768 11 -292.850 607.7  9.73  0.003
## Models ranked by AIC(x)
```

D

2 of the 3 new natural logged pollution variables are added

```r
PolluteModel <- lm(MORTALITY ~ DENSITY + EDUC + HOUSE + JANTEMP +
                     JULYTEMP + NONWHITE + PRECIP + logHC + logNOX, data = Mortality)
vif(PolluteModel)
```

```
##   DENSITY      EDUC     HOUSE   JANTEMP  JULYTEMP  NONWHITE    PRECIP
##  1.480251  1.871509  2.112059  1.980619  3.420447  3.929744  2.350081
```

```
##     logHC    logNOX
## 16.835589 12.660836
```

E

One unit increase in logHC is associated with a 28.48 unit DECREASE in predicted Mortality, holding all other variables in the model constant. One unit increase in logNOX is associated with a 37.65 unit INCREASE in predicted Mortality, holding all otehr variable in the model constant.

**Also, note that the estimated partial regression coefficients associated with logHC in different 'runner up' models has the sign changes when logNOX is in the mode which indicate sensitivity to including both variables in a model.