Extra Practice for Exam1 KEY

1. BodyFat

A. VIF for Full Model

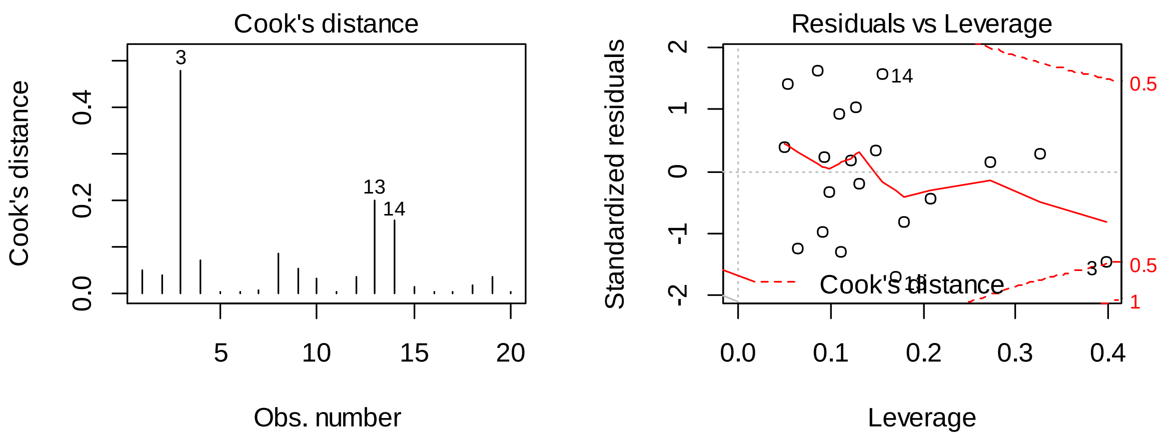
Triceps Thigh Midarm
708.8429 564.3434 104.6060
All VIF values > 10 indicate collinearity!

Hand Calculation:
R2 (Triceps vs other predictors) = 0.9986
VIF = 1/(1-0.9986) = 714.2857 (Does not quite match above due to rounding error.)

B. VIF for model with just Triceps and Midarm

Triceps Midarm
1.265118 1.265118
These look MUCH better than before!

C. Influence Diagnostics



Obs #3 has the largest value of Cook's distance.

```
     dfb.1_   dfb.Trcp dfb.Mdrm   dffit cov.r  cook.d    hat inf
3   1.05631  0.052521 -1.05722 -1.2439 1.327 0.47826 0.3988   *
```

Rules of Thumb:
|DFBETA| > 2/sqrt(n) = 0.447
|DFFITS| = 2*sqrt((k+1)/n) = 0.774
Cooks D > 1

So Obs#3 is influential for DFBETAS and DFFITS, but not Cook's distance.

2. Mortality

A. Correlation

```
           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
```

B. The model selected includes:
`DENSITY, EDUC, HOUSE, JANTEMP, JULYTEMP, NONWHITE, PRECIP`

C. lgHC and lgNOX are added to the model.

D. Checking the vif values indicates that there is collinearity between logHC, logNOX.

```
 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. A one unit increase in logHC is associated with a <u>28.48 unit decrease</u> in predicted Mortality, holding other variables in the model constant.
A one unit increase in logNOX is associated with a <u>37.65 unit increase</u> in predicted Mortality, holding other variables in the model constant.

Another clue that there is an issue is how the estimated coefficient for logHC changes sign depending on whether logNOX is in the model.

```
Model selection table
  (Intrc)    DENSI   EDUC  HOUSE  JANTE   JULYT  logHC lgNOX  lgSO2 NONWH PRECI df
4    1353 0.003128 -13.76 -55.03 -1.980 -2.1730 -28.48 37.65        5.032 1.766 11
8    1460 0.003873 -16.16 -69.76 -2.548 -2.3700 -29.19 45.53 -7.793 5.254 1.903 12
3    1229 0.002266 -16.04 -51.90 -2.265 -0.5672        15.66        4.500 2.067 10
7    1325 0.002939 -18.32 -65.54 -2.801 -0.7131        22.49 -7.252 4.694 2.201 11
2    1322 0.003490 -18.78 -59.26 -2.306 -0.9799  10.73              4.955 1.807 10
1    1525 0.005555 -20.03 -70.03 -2.123 -2.7280                     5.892 1.276  9
```

#1 BodyFat

```
library(car)
HW2Data <- read.csv("C:/hess/STAT512/HW_2018/HW2/BodyFat.csv")
#A
Model1 <- lm(BodyFat ~ ., data = HW2Data)
vif(Model1)
Model2 <- lm(Triceps ~ Thigh + Midarm, data = HW2Data)
summary(Model2)
#B
Model3 <- lm(BodyFat ~ Triceps + Midarm, data = HW2Data)
vif(Model3)
#C
```

```
plot(Model3, which = c(1:2,4:5))
influence.measures(Model3)
```

#2 Mortality

```
library(MuMIn)
library(car)
Mortality <-
read.csv("C:/hess/STAT512/Exams_2018/Exam1_Practice/Mortality.cs
v",row.names = 1)
Mortality$logHC <- log(Mortality$HC)
Mortality$logNOX <- log(Mortality$NOX)
Mortality$logSO2 <- log(Mortality$SO2)
colnames(Mortality)
#A
pairs(Mortality[,c(1, 17:19)])
cor(Mortality[,c(1, 17:19)])
#B
FullModel1 <- lm(MORTALITY ~ PRECIP + HUMIDITY + JANTEMP
                              +JULYTEMP + OVER65 + HOUSE + EDUC
                              + SOUND + DENSITY + NONWHITE
                              + WHITECOL + POOR, data =
                              Mortality)
options(na.action = "na.fail")
AllSubsets1 <- dredge(FullModel1, rank="AIC")
#C
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")
PolluteModel <- lm(MORTALITY ~ DENSITY + EDUC + HOUSE + JANTEMP
+ JULYTEMP + NONWHITE + PRECIP + logHC + logNOX, data =
Mortality)
#D
vif(PolluteModel)
```