https://cran.r-project.org/web/packages/outliers/index.html

Remove the outlier function

```
"outlier" <-
function (x, opposite = FALSE, logical = FALSE)
{
   if (is.matrix(x))
      apply(x, 2, outlier, opposite = opposite, logical = logical)
   else if (is.data.frame(x))
      sapply(x, outlier, opposite = opposite, logical = logical)
   else {
        if (xor(((max(x,na.rm=TRUE) - mean(x,na.rm=TRUE)) < (mean(x,na.rm=TRUE) -
min(x,na.rm=TRUE))),opposite))
                {
                        if (!logical) min(x,na.rm=TRUE)
                        else x == min(x,na.rm=TRUE)
                }
                else
                {
                        if (!logical) max(x,na.rm=TRUE)
                        else x == max(x,na.rm=TRUE)
                }
        }
}

BodyFat = read.csv("BodyFat.csv")
BodyFat = BodyFat[-c(39,41),]
outlier(BodyFat$BODYFAT)
outlier(BodyFat$AGE)
outlier(BodyFat$DENSITY)
outlier(BodyFat$WEIGHT)
outlier(BodyFat$HEIGHT)
outlier(BodyFat$NECK)
outlier(BodyFat$ADIPOSITY)
outlier(BodyFat$CHEST)
outlier(BodyFat$ABDOMEN)
outlier(BodyFat$HIP)
outlier(BodyFat$BICEPS)
outlier(BodyFat$THIGH)


BodyFat = read.csv("BodyFat.csv")
outlier(BodyFat$BODYFAT)
```

Suggestion
Hi guys! I talked to the TA this morning and he give some suggestions about our project. First, according to him it is probably not very scientific for us to take out the 9 outliers (we could take out 3 for sure, but the other ones we do not have strong evidence to take). The best way to eliminate those outliers is basing on their cook's distance value (close to 1). Therefore, we might need to compute the cook's distance's value of each outlier to make sure it is an outlier. Second -- if I understand correctly -- we use SLR methods to detect outliers for each variable against the body fat and choose 4. Then we use MLR R function to get coefficient of the four variables and form an equation. But this is also not very scientific as there are influences between variables and moreover we could not just use the SLR results and put them in MLR. The better way of doing it is to calculate the R-square (correlation) of each variable against the body fat. Then we choose one of the most correlative variable and put it in MLR model. Then we get its residual plots and etc. Then we choose another variable if we decided to do MLR – we choose another variable, but we need to compare the results of having and not having the second variable using R squares. That way we make sure in MLR one variable does not influence another and they are all strongly correlated to the bodyfat.
I have put all ta's suggestions in blue in the middle section where we pose our questions. I think we should ask the professor whether Ta's suggestion is practical or not.

Rank by correlation next one --- next one--

Reading Data into dataframe/ analyzing outliers

```
BodyFat = read.csv("BodyFat.csv")
BodyFat

pairs(BodyFat)
?pairs
```

POTENTIAL OUTLIERS:

```
BodyFat[BodyFat$WEIGHT >= 350,] #this shows that we shouldn't use this row to check, row 39
BodyFat[BodyFat$HEIGHT <= 30,] # major height outlier, row 42
*BodyFat[BodyFat$AGE > 75,] # row 79
```

```
BodyFat[BodyFat$ADIPOSITY >= 45, ]#row 39
BodyFat[BodyFat$NECK >= 50, ] #row 39
BodyFat[BodyFat$CHEST >= 35, ] #row 39
BodyFat[BodyFat$ABDOMEN >= 140, ]#row 39
BodyFat[BodyFat$HIP >= 135, ]#row 39
BodyFat[BodyFat$THIGH >= 85 , ] #row39
BodyFat[BodyFat$KNEE >= 47 , ] #row39
BodyFat[BodyFat$ANKLE >= 30 , ]#row 31,86
BodyFat[BodyFat$BICEPS >= 30 , ]# row 39
BodyFat[BodyFat$WRIST >= 20.9, ] #row 39 and 41
BodyFat[BodyFat$FOREARM <= 22, ] #row 226
```

CURRENT MODEL

```
BodyFat = read.csv("BodyFat.csv") #Read data into r

BodyFat = BodyFat[-c(39,41,216),] #many outliers

m= lm(BODYFAT ~ ABDOMEN, data=BodyFat)
m
summary(m)

t = predict(m, interval="confidence")
```

HISTOGRAM DATA

```
par(mfrow=c(4,4)) #Makes a two-by-two, i.e. (2,2), plotting window
par(mgp=c(1.8,.5,0), mar=c(3,3,1,1)) #"Beautifies" plots when creating multiple figures. Google this for more info.
attach(BodyFat)
hist(BODYFAT,breaks=30,
    main="Histogram of Body Fat %",xlab="Body Fat %")
hist(AGE,breaks=30,
    main="Histogram of Age",xlab="Age (yrs)")
hist(WEIGHT,breaks=30,
    main="Histogram of Weight",xlab="Weight (lbs)")
hist(HEIGHT,breaks=30,
    main="Histogram of HEIGHT",xlab="Knee Circumference (cm)")
hist(ADIPOSITY,breaks=30,
    main="Histogram of ADIPOSITY %",xlab="Adiposity")
```

```
hist(NECK,breaks=30,
    main="Histogram of NECK %",xlab="Neck (inches)")
hist(DENSITY,breaks=30,
    main="Histogram of DENSITY %",xlab="Density")
hist(HIP,breaks=30,
    main="Histogram of HIP %",xlab="Hip (inches)")
hist(THIGH,breaks=30,
    main="Histogram of THIGH %",xlab="Thigh (inches)")
hist(ANKLE,breaks=30,
    main="Histogram of ANKLE %",xlab="Ankle (inches)")
hist(BICEPS,breaks=30,
    main="Histogram of BICEPS%",xlab="Biceps (inches)")
hist(FOREARM,breaks=30,
    main="Histogram of FOREARM %",xlab="Forearm (inches)")
hist(WRIST,breaks=30,
    main="Histogram of WRIST %",xlab="Wrist (inches)")
hist(CHEST,breaks=30,
    main="Histogram of CHEST %",xlab="Chest (inches)")
```

Residual Plots and QQ Plots R code

```
lmmodel = lm(BODYFAT ~ AGE,data = BodyFat)
AGE = BodyFat$AGE
plot(AGE,resid(lmmodel),pch=23,bg="red",cex=1.2,
   xlab="Age", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)


lmmodel = lm(BODYFAT ~ WEIGHT,data = BodyFat)
WEIGHT = BodyFat$WEIGHT
plot(WEIGHT,resid(lmmodel),pch=23,bg="red",cex=1.2,
   xlab="WEIGHT", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)


lmmodel = lm(BODYFAT ~ HEIGHT,data = BodyFat)
```

```
HEIGHT = BodyFat$HEIGHT
plot(HEIGHT,resid(lmmodel),pch=23,bg="red",cex=1.2,
   xlab="HEIGHT", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)


lmmodel = lm(BODYFAT ~ NECK,data = BodyFat)
NECK = BodyFat$NECK
plot(NECK,resid(lmmodel),pch=23,bg="red",cex=1.2,
   xlab="NECK", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)


lmmodel = lm(BODYFAT ~ ADIPOSITY,data = BodyFat)
ADIPOSITY = BodyFat$ADIPOSITY
plot(ADIPOSITY,resid(lmmodel),pch=23,bg="red",cex=1.2,
   xlab="NECK", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)


lmmodel=lm(BODYFAT~CHEST, data=BodyFat)
chest = BodyFat$CHEST
plot(chest,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Chest", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)

lmmodel=lm(BODYFAT~ABDOMEN, data=BodyFat)
abdomen = BodyFat$ABDOMEN
plot(abdomen,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Abdomen", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)
```

```
lmmodel=lm(BODYFAT~HIP, data=BodyFat)
hip = BodyFat$HIP
plot(hip,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Hip", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)

lmmodel=lm(BODYFAT~THIGH, data=BodyFat)
thigh = BodyFat$THIGH
plot(thigh,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Thigh", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)

llmmodel = lm(BODYFAT~WRIST, data=Bodyfat)

wrist_variable=Bodyfat$WRIST
plot(wrist_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="wrist", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)

######################

lmmodel = lm(BODYFAT~FOREARM, data=Bodyfat)

forearm_variable=Bodyfat$FOREARM
plot(forearm_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="forearm", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)

######################
```

```
lmmodel = lm(BODYFAT~BICEPS, data=Bodyfat)

biceps_variable=Bodyfat$BICEPS
plot(biceps_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="biceps", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)
###############
lmmodel = lm(BODYFAT~ANKLE, data=Bodyfat)

ankle_variable=Bodyfat$ANKLE
plot(ankle_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="ankle", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)
###############
lmmodel = lm(BODYFAT~KNEE, data=Bodyfat)

knee_variable=Bodyfat$KNEE
plot(knee_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="knee", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2)
abline(a=0,b=1,col="black",lwd=3)
```

Outlier : 182 (0 bodyfat)

ASSUMPTIONS TABLE

	HOMOSKADESCITY	LINEARITY	NORMALITY	OUTLIER	Independence test (P=pass/F=fail)
AGE	YES		YES	Yes (79)	F
WEIGHT	YES		YES	NO	P
HEIGHT	NO		YES	YES(42)	F

NECK	YES		NO	YES(226)	F
ADPOSITY	YES/NO		YES/NO (depending on an outlier, check with and without row 39)	YES(41,216)	F
CHEST	YES		YES	NO	P
ABDOMEN	YES		YES/NO	NO	P
HIP	YES/NO		YES	NO	P
THIGH	YES		NO	NO	F
WRIST	YES		NO	NO	F
FOREARM	YES		NO	YES(175)	F
BICEPS	YES		NO	NO	F
ANKLE	YES		NO	YES(31,86)	F
KNEE	YES		NO	YES(244)	F

Questions to be answered:

- HOMOSCEDASTICITY (variance in Y-axis, correct? Is height homoskedastic?
- Which test is more appropriate to put in report(F-test or Chi-Squared test)

Chi-squared test is not really appliable here. For, F-test, better use the p-value

- How many outliers should we remove(we have 9 points so far?)

Maybe around 3 that influence every variable. For the others we need cook's distance to prove it

- Should we remove row 172 because it has 1.9% body fat? (our model predicted 9.62% and it has been pretty consistent)

Probably not.

- How can we prevent the plots from "squishing"?
- Make sure we present on Monday in case I qualify for Big Ten's and am out of town next Wednesday.

- F-test p-value: (except chest) values were in the range 2.2*e-16 . Chi-squared test results show that retain the null? Values range from 0.05422 and 0.2175? That says reject null which means its a dependent variable but works well in model? Doesn't make this.

use the lm() statistics. Then get the summary(). We can use the $ sign to access the pvalue. It is better to use the p value to assess whether it is a good fit or not. (Maybe not F-statistic or Chi-squared) . Also, we might use Beta 1 in the summary for prediction.

- Should we mention that for all three confidence intervals (95%, 90%, 99%- p value is all same?) weird or normal?

Both are okay. 95 percent is better.

Conflin:

P value

- What counts as a high F value? Is p value? F-test statistic

No standard. So better use the p-value.

- In the F-test would it be more helpful to just look at the F statistic, or the p-value? P
- Are we right about testing through rows or should we find a random sample or samples?
- Are 3-5 random samples of 10 to 12 rows each sufficient? Or is there another way to do this?

That is kind of unnecessary. We might prioritize caculate R square, cook's distance and etc

- What graphs are suitable with the MLR model-- graphs of qqplot and residual of just the predictors?

Both are very good

- The scatterplot of 4 predictors vs body fat % looks really bad. Should we scale hips, abdomen, and chest to work with weight, or should be plot 2 different plots, 4?
- Should we do an F-test in the end after a random sampling?
- Use Wikipedia to show that our model predicts for the typical BMI not for absurd values? Can we do this?

Maybe not

- Does the team contribution section/works cited have to be included in the 2 page limit?

Yes

TO DO YET:

1. Get the correlation value
2. Make a graph with the 4 predictors on x axis and body fat on the y axis and fit the MLR line on it.
3. Prediction interval

residual plot matrix

```
matrix_want=matrix(c(1,2,3,4,5,6,7,8,9,10,11,12,13,14),nrow=7, ncol=2,byrow=TRUE)
layout(matrix_want)
lmmodel = lm(BODYFAT ~ AGE,data = BodyFat)
AGE = BodyFat$AGE
plot(AGE,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Age", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT ~ WEIGHT,data = BodyFat)
WEIGHT = BodyFat$WEIGHT
plot(WEIGHT,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="WEIGHT", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT ~ HEIGHT,data = BodyFat)
HEIGHT = BodyFat$HEIGHT
plot(HEIGHT,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="HEIGHT", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT ~ NECK,data = BodyFat)
NECK = BodyFat$NECK
plot(NECK,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="NECK", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT ~ ADIPOSITY,data = BodyFat)
ADIPOSITY = BodyFat$ADIPOSITY
plot(ADIPOSITY,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="ADIPOSITY", ylab="Residuals",main="Residual Plot")
```

```
abline(a=0,b=0,col="black",lwd=3)

lmmodel=lm(BODYFAT~CHEST, data=BodyFat)
chest = BodyFat$CHEST
plot(chest,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Chest", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel=lm(BODYFAT~ABDOMEN, data=BodyFat)
abdomen = BodyFat$ABDOMEN
plot(abdomen,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Abdomen", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel=lm(BODYFAT~HIP, data=BodyFat)
hip = BodyFat$HIP
plot(hip,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Hip", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel=lm(BODYFAT~THIGH, data=BodyFat)
thigh = BodyFat$THIGH
plot(thigh,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="Thigh", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT~WRIST, data=Bodyfat)

wrist_variable=BodyFat$WRIST
plot(wrist_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="wrist", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT~FOREARM, data=BodyFat)

forearm_variable=BodyFat$FOREARM
plot(forearm_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="forearm", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT~BICEPS, data=BodyFat)

biceps_variable=BodyFat$BICEPS
```

```
plot(biceps_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="biceps", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT~ANKLE, data=BodyFat)

ankle_variable=BodyFat$ANKLE
plot(ankle_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="ankle", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)

lmmodel = lm(BODYFAT~KNEE, data=BodyFat)

knee_variable=BodyFat$KNEE
plot(knee_variable,resid(lmmodel),pch=23,bg="red",cex=1.2,
    xlab="knee", ylab="Residuals",main="Residual Plot")
abline(a=0,b=0,col="black",lwd=3)
```

---

matrix for histograms

```
matrix_hist=matrix(c(1,2,3,4,5,6,7,8,9,10,11,12,13,14), nrow=7, ncol=2, byrow = TRUE)
layout(matrix_hist)
hist(BODYFAT,breaks=30,
    main="Histogram of Body Fat %",xlab="Body Fat %")
hist(AGE,breaks=30,
    main="Histogram of Age",xlab="Age (yrs)")
hist(WEIGHT,breaks=30,
    main="Histogram of Weight",xlab="Weight (lbs)")
hist(HEIGHT,breaks=30,
    main="Histogram of HEIGHT",xlab="Knee Circumference (cm)")
hist(ADIPOSITY,breaks=30,
    main="Histogram of ADIPOSITY %",xlab="Body Fat %")
hist(NECK,breaks=30,
    main="Histogram of NECK %",xlab="Body Fat %")
hist(DENSITY,breaks=30,
    main="Histogram of DENSITY %",xlab="Body Fat %")
hist(HIP,breaks=30,
    main="Histogram of HIP %",xlab="Body Fat %")
hist(THIGH,breaks=30,
    main="Histogram of THIGH %",xlab="Body Fat %")
hist(ANKLE,breaks=30,
    main="Histogram of ANKLE %",xlab="Body Fat %")
hist(BICEPS,breaks=30,
```

```
    main="Histogram of BICEPS%",xlab="Body Fat %")
hist(FOREARM,breaks=30,
    main="Histogram of FOREARM %",xlab="Body Fat %")
hist(WRIST,breaks=30,
    main="Histogram of WRIST %",xlab="Body Fat %")
hist(CHEST,breaks=30,
    main="Histogram of CHEST %",xlab="Body Fat %")
```

---

matrix for qqplots

```
matrix_qqplot=matrix(c(1,2,3,4,5,6,7,8,9,10,11,12,13,14), nrow=7, ncol=2, byrow = TRUE)
layout(matrix_qqplot)
lmmodel = lm(BODYFAT~AGE, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="age")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~WEIGHT, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="weight")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~HEIGHT, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="height")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~ADIPOSITY, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="adiposity")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~NECK, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="neck")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~CHEST, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="chest")
abline(a=0,b=1,col="black",lwd=3)


lmmodel = lm(BODYFAT~ABDOMEN, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="abdomen")
abline(a=0,b=1,col="black",lwd=3)
```

```
lmmodel = lm(BODYFAT~HIP, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="hip")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~THIGH, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="thigh")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~KNEE, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="knee")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~ANKLE, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="ankle")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~BICEPS, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="biceps")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~FOREARM, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="forearm")
abline(a=0,b=1,col="black",lwd=3)

lmmodel = lm(BODYFAT~WRIST, data=BodyFat)
qqnorm(rstandard(lmmodel),pch=23,bg="red",cex=1.2, main="wrist")
abline(a=0,b=1,col="black",lwd=3)
```

Chi square tests:

```
# chisquares test
age_val=BodyFat$AGE
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi=matrix(data=c(age_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi)

height_val=BodyFat$HEIGHT
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi2=matrix(data=c(height_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi2)
```

```
adiposity_val=BodyFat$ADIPOSITY
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi3=matrix(data=c(adiposity_val,Bodyfat_compare), nrow=245, ncol=2,
byrow=TRUE)
chisq.test(matrix_for_chi3)

hip_val=BodyFat$HIP
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi4=matrix(data=c(hip_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi4)

thigh_val=BodyFat$THIGH
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi7=matrix(data=c(thigh_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi7)

knee_val=BodyFat$KNEE
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi9=matrix(data=c(knee_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi9)

forearm_val=BodyFat$FOREARM
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi13=matrix(data=c(forearm_val,Bodyfat_compare), nrow=245, ncol=2,
byrow=TRUE)
chisq.test(matrix_for_chi13)

wrist_val=BodyFat$WRIST
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi14=matrix(data=c(wrist_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi14)

weight_val=BodyFat$WEIGHT
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi10=matrix(data=c(weight_val,Bodyfat_compare), nrow=245, ncol=2,
byrow=TRUE)
chisq.test(matrix_for_chi10)

neck_val=BodyFat$NECK
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi11=matrix(data=c(neck_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi11)
```

```
chest_val=BodyFat$CHEST
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi5=matrix(data=c(chest_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi5)


abdomen_val=BodyFat$ABDOMEN
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi6=matrix(data=c(abdomen_val,Bodyfat_compare), nrow=245, ncol=2,
byrow=TRUE)
chisq.test(matrix_for_chi6)

ankle_val=BodyFat$ANKLE
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi8=matrix(data=c(ankle_val,Bodyfat_compare), nrow=245, ncol=2, byrow=TRUE)
chisq.test(matrix_for_chi8)

biceps_val=BodyFat$BICEPS
Bodyfat_compare=BodyFat$BODYFAT
matrix_for_chi12=matrix(data=c(biceps_val,Bodyfat_compare), nrow=245, ncol=2,
byrow=TRUE)
chisq.test(matrix_for_chi12)
```

```
Rank by correlation using R^2 value
Check using residual plots and QQ plots
cor(BodyFat$AGE, BodyFat$BODYFAT)
cor(BodyFat$WEIGHT, BodyFat$BODYFAT)
cor(BodyFat$HEIGHT, BodyFat$BODYFAT)
cor(BodyFat$ADIPOSITY, BodyFat$BODYFAT)
cor(BodyFat$NECK, BodyFat$BODYFAT)
cor(BodyFat$CHEST, BodyFat$BODYFAT)
cor(BodyFat$ABDOMEN, BodyFat$BODYFAT)
cor(BodyFat$HIP, BodyFat$BODYFAT)
cor(BodyFat$THIGH, BodyFat$BODYFAT)
```

```
cor(BodyFat$KNEE, BodyFat$BODYFAT)
cor(BodyFat$ANKLE, BodyFat$BODYFAT)
cor(BodyFat$BICEPS, BodyFat$BODYFAT)
cor(BodyFat$FOREARM, BodyFat$BODYFAT)
cor(BodyFat$WRIST, BodyFat$BODYFAT)
```

CORRELATION RANKING (after l.p. and outliers removed)

Predictor	Correlation Coefficient (Raw Data)	R^2 Value (Raw Data)	Correlation Coefficient (Manipulated Data)	R^2 Value (Manipulated Data)
Age	0.289	0.084	0.293	0.085
Weight	0.613	0.376	0.610	0.372
Height	-0.089	0.008	-0.071	-0.005
Adiposity	0.728	0.530	0.735	0.540
Neck	0.491	0.241	0.467	0.218
Chest	0.703	0.494	0.688	0.473
Abdomen	0.814	0.663	0.817	0.667
Hip	0.626	0.392	0.625	0.391
Thigh	0.561	0.315	0.553	0.306
Knee	0.507	0.257	0.520	0.270
Ankle	0.267	0.071	0.236	0.056
Biceps	0.493	0.243	0.476	0.227
Forearm	0.363	0.132	0.361	0.125
Wrist	0.348	0.121	0.320	0.108

Matrix for Cook's distance for each predictor

(finding leverage points- proving row 39 is full of them)

```
y=matrix(c(1,2,3,4,5,6,7,8,9,10,11,12,13,14),nrow=7, ncol=2,byrow=TRUE)
layout(y)
m = lm(BODYFAT ~ AGE, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$BODYFAT)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for AGE)",ylab="Cook's Distance")

m = lm(BODYFAT ~ WEIGHT, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$WEIGHT)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for WEIGHT)",ylab="Cook's Distance")

m = lm(BODYFAT ~ HEIGHT, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$HEIGHT)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for HEIGHT)",ylab="Cook's Distance")

m = lm(BODYFAT ~ ADIPOSITY, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$ADIPOSITY)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for ADIPOSITY)",ylab="Cook's Distance")

m = lm(BODYFAT ~ NECK, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$NECK)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for NECK)",ylab="Cook's Distance")

m = lm(BODYFAT ~ CHEST, data= BodyFat)
```

```
cooki = cooks.distance(m)
n = length(BodyFat$CHEST)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for CHEST)",ylab="Cook's Distance")

m = lm(BODYFAT ~ ABDOMEN, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$ABDOMEN)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for ABDOMEN)",ylab="Cook's Distance")

m = lm(BODYFAT ~ HIP, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$HIP)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for HIP)",ylab="Cook's Distance")

m = lm(BODYFAT ~ THIGH, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$THIGH)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for THIGH)",ylab="Cook's Distance")

m = lm(BODYFAT ~ KNEE, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$KNEE)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for KNEE)",ylab="Cook's Distance")

m = lm(BODYFAT ~ ANKLE, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$ANKLE)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for ANKLE)",ylab="Cook's Distance")

m = lm(BODYFAT ~ BICEPS, data= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$BICEPS)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for BICEPS)",ylab="Cook's Distance")

m = lm(BODYFAT ~ FOREARM, data= BodyFat)
cooki = cooks.distance(m)
```

```
n = length(BodyFat$FOREARM)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for FOREARM)",ylab="Cook's Distance")

m = lm(BODYFAT ~ WRIST, d ata= BodyFat)
cooki = cooks.distance(m)
n = length(BodyFat$WRIST)
plot(1:n,cooki,type="p",pch=23,bg="red",cex=1.2,
    xlab="Index (Each Observation for WRIST)",ylab="Cook's Distance")
```

Abdomen scatter plot with 95% confidence bands

```
 m = lm(BODYFAT ~ ABDOMEN, data=BodyFat)

 v = confint(m, level=0.95)
 plot(BodyFat$ABDOMEN, BodyFat$BODYFAT)
abline(m)
abline(v[1], v[2], col="blue", lty="dotted")
abline(v[3], v[4], col="blue", lty="dotted")
```