Group Members: Bradley Dodds

Class: CPSC 375

Professor: Dr. Panangadan

Due Date: 12/3/2019

Project 1: Body Fat Percentage

Project Description:

This project will identify 13 predictors from the data set *Bodyfat.csv* to determine the most accurate estimate of body fat percentage. Each variable will be visualized and cleaned by removing any outliers. Different combinations of the variables will be used for linear modeling. A body fat prediction function will be written in R to predict the most accurate body fat percentage. The project will demonstrate exploratory data analysis and modeling for the data set that represents estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men.

Variables:

Age, Weight, Height, Neck, Chest, Abdomen, Hip, Thigh, Knee, Ankle, Biceps, Forearm, Wrist

Circumference Measurements:

cm

Data Examined:

http://staff.pubhealth.ku.dk/~tag/Teaching/share/data/Bodyfat.html

Data Set File:

Bodyfat.csv

Software Used:

RStudio

Modeling: bodyfat_model

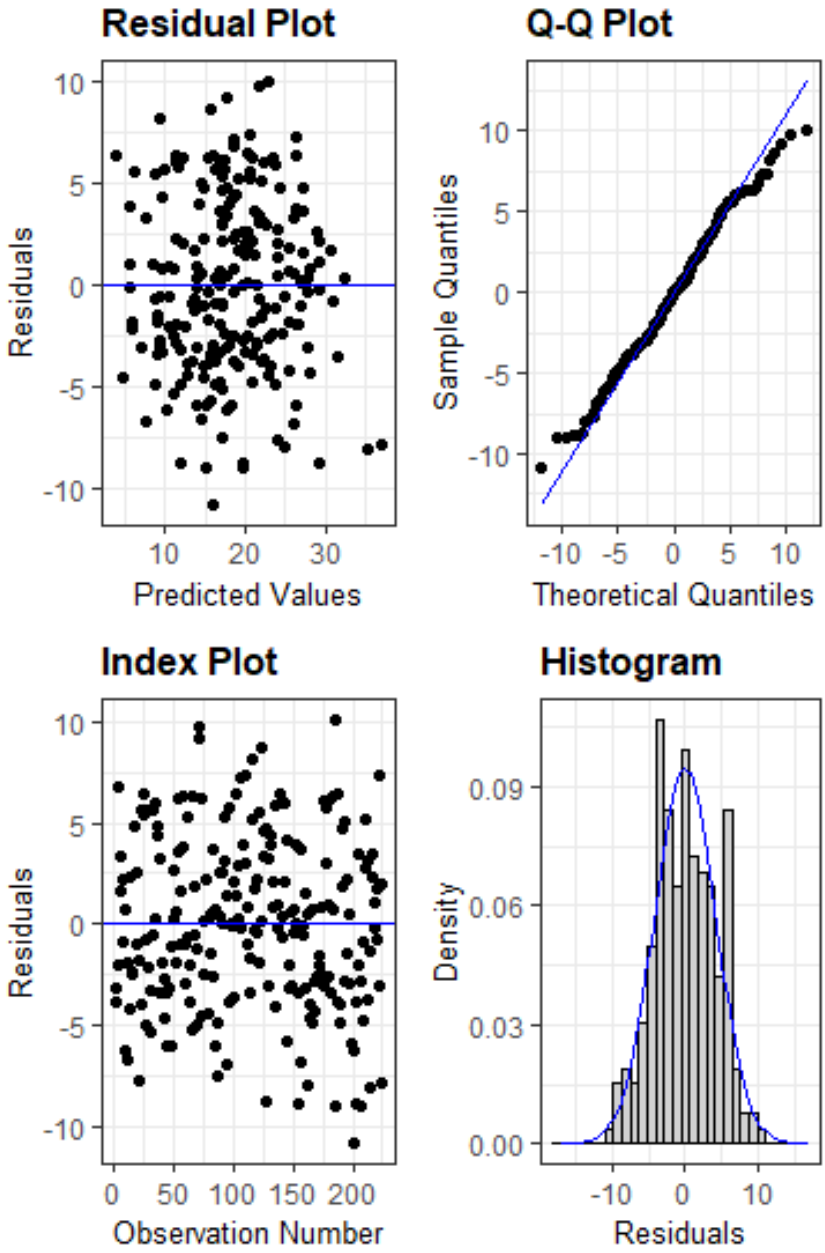
```
    Density         bodyfat          Age            Weight          Height          Neck            Chest      
 Min.   :0.995   Min.   : 0.00   Min.   :22.00   Min.   :118.5   Min.   :29.50   Min.   :31.10   Min.   : 79.30  
 1st Qu.:1.041   1st Qu.:12.47   1st Qu.:35.75   1st Qu.:159.0   1st Qu.:68.25   1st Qu.:36.40   1st Qu.: 94.35  
 Median :1.055   Median :19.20   Median :43.00   Median :176.5   Median :70.00   Median :38.00   Median : 99.65  
 Mean   :1.056   Mean   :19.15   Mean   :44.88   Mean   :178.9   Mean   :70.15   Mean   :37.99   Mean   :100.82  
 3rd Qu.:1.070   3rd Qu.:25.30   3rd Qu.:54.00   3rd Qu.:197.0   3rd Qu.:72.25   3rd Qu.:39.42   3rd Qu.:105.38  
 Max.   :1.109   Max.   :47.50   Max.   :81.00   Max.   :363.1   Max.   :77.75   Max.   :51.20   Max.   :136.20  
    Abdomen          Hip             Thigh           Knee           Ankle           Biceps          Forearm     
 Min.   : 69.40   Min.   : 85.0   Min.   :47.20   Min.   :33.00   Min.   :19.1   Min.   :24.80   Min.   :21.00  
 1st Qu.: 84.58   1st Qu.: 95.5   1st Qu.:56.00   1st Qu.:36.98   1st Qu.:22.0   1st Qu.:30.20   1st Qu.:27.30  
 Median : 90.95   Median : 99.3   Median :59.00   Median :38.50   Median :22.8   Median :32.05   Median :28.70  
 Mean   : 92.56   Mean   : 99.9   Mean   :59.41   Mean   :38.59   Mean   :23.1   Mean   :32.27   Mean   :28.66  
 3rd Qu.: 99.33   3rd Qu.:103.5   3rd Qu.:62.35   3rd Qu.:39.92   3rd Qu.:24.0   3rd Qu.:34.33   3rd Qu.:30.00  
 Max.   :148.10   Max.   :147.7   Max.   :87.30   Max.   :49.10   Max.   :33.9   Max.   :45.00   Max.   :34.90  
     Wrist      
 Min.   :15.80  
 1st Qu.:17.60  
 Median :18.30  
 Mean   :18.23  
 3rd Qu.:18.80  
 Max.   :21.40  
```

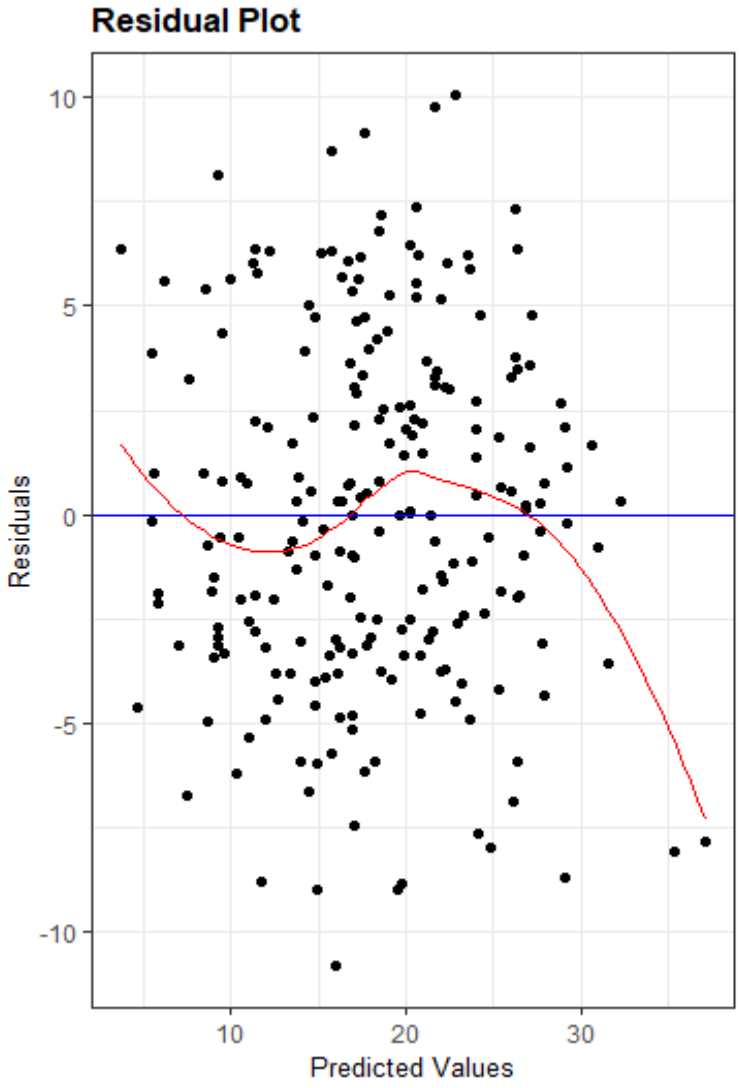
```
lm(formula = bodyfat ~ Age + Weight + Height + Neck + Chest + 
    Abdomen + Hip + Thigh + Knee + Ankle + Biceps + Forearm + 
    Wrist, data = bodyfat)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.7832  -3.1087  -0.0108   3.1079  10.0344 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  7.709952  27.130436   0.284  0.77655    
Age          0.072063   0.034855   2.068  0.03990 *  
Weight       0.007312   0.078293   0.093  0.92568    
Height      -0.262554   0.218470  -1.202  0.23079    
Neck        -0.444355   0.279539  -1.590  0.11342    
Chest       -0.109051   0.118840  -0.918  0.35986    
Abdomen      0.871696   0.100245   8.696 9.77e-16 ***
Hip         -0.200770   0.180813  -1.110  0.26810    
Thigh        0.251309   0.167284   1.502  0.13452    
Knee        -0.072022   0.297268  -0.242  0.80880    
Ankle        0.061884   0.400756   0.154  0.87743    
Biceps       0.172411   0.196073   0.879  0.38023    
Forearm      0.230815   0.371744   0.621  0.53534    
Wrist       -1.813656   0.649058  -2.794  0.00568 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.357 on 211 degrees of freedom
Multiple R-squared:  0.694,    Adjusted R-squared:  0.6752 
F-statistic: 36.82 on 13 and 211 DF,  p-value: < 2.2e-16
```


Residual Plot
Residuals
Predicted Values
10
5
0
-5
-10
10
20
30
Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-10
-5
0
5
10
Index Plot
Residuals
Observation Number
0
50
100
150
200
Histogram
Density
0.09
0.06
0.03
0.00
Residuals


Residual Plot
10
5
0
-5
-10
Residuals
10
20
30
Predicted Values

Model 2:

```
lm(formula = bodyfat ~ Abdomen + Wrist, data = bodyfat)

Residuals:
     Min       1Q   Median       3Q      Max
-10.0861  -3.0967  -0.1398   3.2598  11.2217

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -14.45729    6.53831  -2.211    0.028 *
Abdomen       0.81598    0.04175  19.543  < 2e-16 ***
Wrist        -2.29132    0.43579  -5.258 3.42e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.466 on 222 degrees of freedom
Multiple R-squared:  0.6618,   Adjusted R-squared:  0.6588
F-statistic: 217.3 on 2 and 222 DF,  p-value: < 2.2e-16
```

Model 3:

```
lm(formula = bodyfat ~ Height + Abdomen + Wrist, data = bodyfat)

Residuals:
    Min      1Q  Median      3Q     Max
-9.6534 -3.1590 -0.2565  3.1576  9.7649

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.55689    8.80745   0.631   0.5287
Height      -0.40881    0.12366  -3.306   0.0011 **
Abdomen      0.81561    0.04085  19.966  < 2e-16 ***
Wrist       -1.80844    0.45068  -4.013 8.22e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.369 on 221 degrees of freedom
Multiple R-squared:  0.6778,   Adjusted R-squared:  0.6734
F-statistic:   155 on 3 and 221 DF,  p-value: < 2.2e-16
```

Model 4:

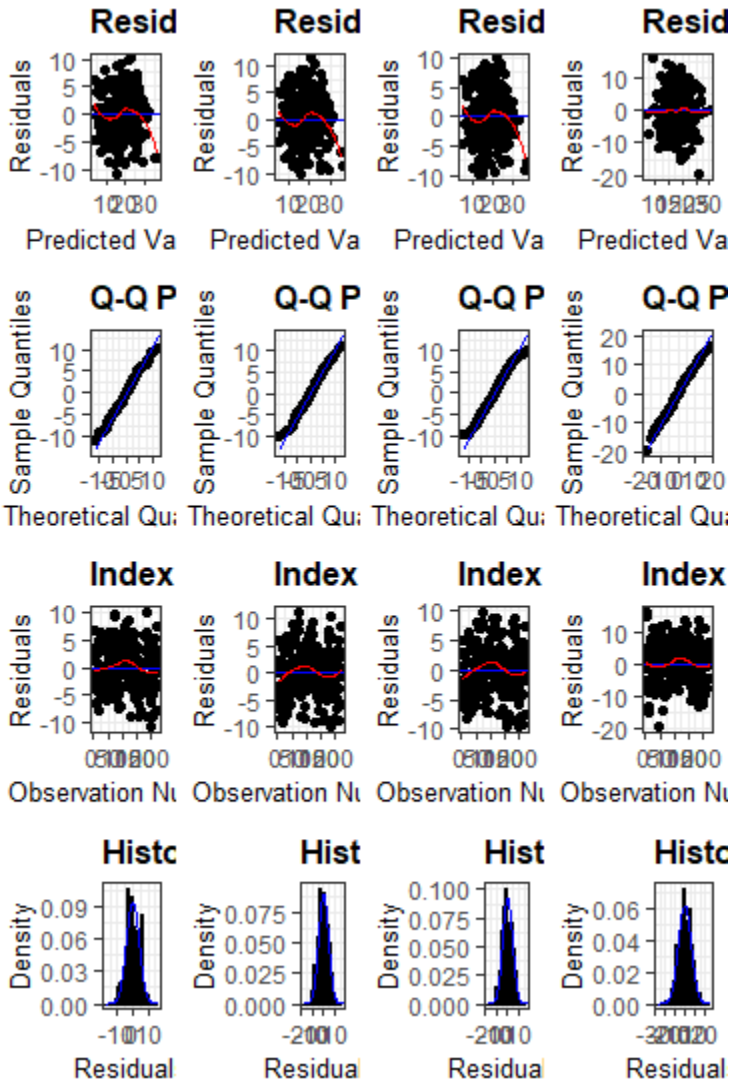
```
lm(formula = bodyfat ~ Age + Forearm + Biceps, data = bodyfat)

Residuals:
     Min       1Q   Median       3Q      Max
-19.5018  -4.5259  -0.0493   4.8279  16.0469

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -33.6754     7.3512  -4.581 7.73e-06 ***
Age           0.1968     0.0343   5.737 3.15e-08 ***
Forearm       0.2807     0.4073   0.689    0.491
Biceps        1.0946     0.2619   4.180 4.21e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.474 on 221 degrees of freedom
Multiple R-squared:  0.2925,   Adjusted R-squared:  0.2829
F-statistic: 30.46 on 3 and 221 DF,  p-value: < 2.2e-16
```

Compare Models:



Out of the four models examined, the linear model with Height + Abdomen + Wrist had the highest adjusted r2, equally distributed residuals around zero, and low error.