Regression Analysis of a Person's Body Fat Percentage

Ravish Kamath*
213893664

Nandisha Uppiah
218345173

Sujoy Ghosh Roy
216805632

Introduction

This project aims to build a model that can predict a person's body fat percentage. The model has been created by analyzing a random group of people based on multiple factors. This will result in a linear model that will be able to adequately predict the body fat percentage for a larger population size. The goal for this project is to be able to predict a person's body fat percentage using a set of quantitative categories based on a person's body. The objective is to do an in-depth analysis and find a model best suited for this, such that it can serve as a cost-effective alternative to more complicated methods of calculating body fat percentage. This study is of particular importance to us as it may be used to determine a person's general health based on their body fat percentage analysis, at a significantly lower cost while still being very effective.

Methodology: Model Building

There are two types of variables in our data collection. The independent variable (explanatory variable) and the dependent variable (response variable). Body fat percentage is the response variable in our study since that is what we are attempting to predict using thirteen explanatory variables. Since the model consists of more than one explanatory variable, we can clearly infer that this model is a multi linear regression model. Age, weight, height, density, wrist circumference, thigh circumference, neck circumference, knee circumference, abdomen circumference, hip circumference, forearm circumference, chest circumference, and bicep circumference were among the 13 explanatory variables in the original data set, but only 12 were chosen. The explanation for this is simple: as the number of independent variables in our model grows, the coefficient of determination improves, thus influencing the significance of the study. Even though an extra explanatory variable may not impact the overall model drastically, we are trying to build up a model that is significant to our analysis. We have seen that density is quite costly to use as compared to the other independent variables. Furthermore, we chose to implement only the variables that according to us, does impact the body fat percentage of an individual. Below is a breakdown of the types of variables in our model:

Response Variable:

Body Fat: Percentage body fat from Siri's (1956) equation (BF)

Explanatory Variables:

Age: Years (A)
Weight: pounds (W)
Height: inches (H)
Neck circumference: cm (N)
Chest circumference: cm (C)
Abdomen circumference: cm (AB)
Hip circumference: cm (H)
Thigh circumference: cm (T)
Knee circumference: cm (K)
Ankle circumference: cm (AN)
Biceps circumference: cm (B)
Forearm circumference: cm (F)
Wrist circumference: cm (W)

Certain manipulations of our dataset were needed to make our data more accurate. Firstly, we permanently removed observation numbers 42, 182 and 172. If we look at the original data set, we see that in observation 42, the individual's height is 29.5 inches however their abdomen is roughly 42 inches (when converted from cm to inches), which can be deemed as impossible. Also, observation 182 and 172 are deemed impossible because these individuals have roughly 0% body fat.

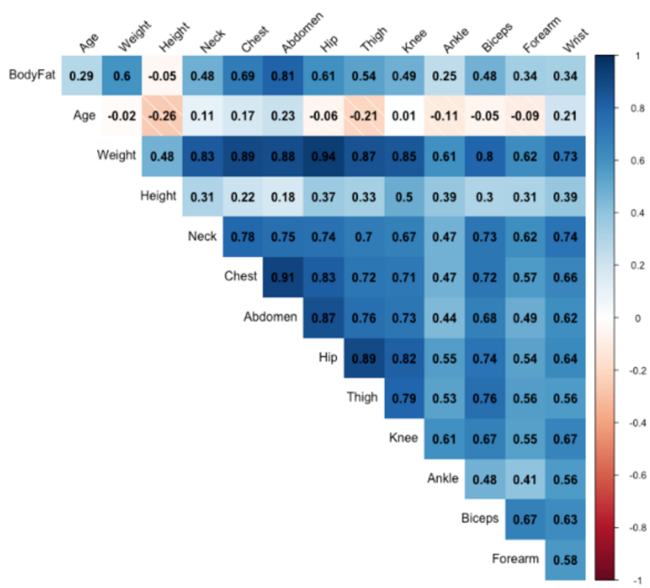


Fig 1: Correlation Matrix (Appendix 1.7)

Body Mass Index (BMI) is an effective method for estimating body fat percentage based on height and weight. As seen from Figure 1, there is a high correlation between height and weight. We can combine height and weight into a single BMI. This contraction can be advantageous because we will have one less explanatory variable in our overall model, without significant change in the R squared.

<u>BMI formula (as recommended by the CDC):</u>

$$BMI = \left(\frac{weight(lbs)}{(height)^2}\right) \times 703$$

<u>Setting up categorical data:</u>

Lastly, we may choose to use age regressor as a category variable to have a detailed understanding of body fat distribution based on everyone's age cohort. Naturally, people who are younger tend to have less body fat than those who are older.

Young Age (YA)	Middle Age (MA)	Old Age (OA)
20-39	40-59	60+

Now that we have formalized our dataset, we will look at creating a linear regression model to see which variables are significant in determining an accurate body fat percentage. By using all the regressors, we may not have the most efficient model, despite it having a good R squared. To test the model's significance, we have certain assumptions that we must consider before coming up with our hypothesis testing. Below is our list of assumptions.

Assumptions:

1. The relationship between the response y and the regressors is linear, at least approximately. (Montgomery, Peck, Vining, 129)
2. The error term ε has zero mean. (Montgomery, Peck, Vining, 129)
3. The error term ε has constant variance σ^2. (Montgomery, Peck, Vining, 129)
4. The errors are uncorrelated. (Montgomery, Peck, Vining, 129)
5. The errors are normally distributed. (Montgomery, Peck, Vining, 129)

To verify these assumptions, we will be looking at the QQ plot to verify our normal distribution, and the residual plots for testing constant variance. In our result section, we will present whether all the assumptions have been met or not. We will also be identifying outliers, influential points and leverage that can have a big impact on our models. To identify such points, we will be looking at our PRESS statistic, Cook's Distance and DFBETAS for influential points. Using the residual plots with Rstudent and leverage, we can identify certain outliers. If any of the assumptions are not met, then a transformation process must be completed on our dataset to accept all assumptions before testing.

After checking all assumptions, we will be looking at subsets of our full model to determine which can accurately describe body fat percentage. This is called model building and we will be using 3 methods to determine the best subset of models. These 3 methods are called forward Selection, backward selection, and stepwise regression. It must be noted that these methods may not determine the best model, however it does get us one step closer to coming up with an accurate model. To use these models, we must pre-determine our significance level. In this report we use our alpha entry to be 0.25 and alpha stay to be 0.1. We determined these based of the textbook, Introduction to Linear Regression Analysis Fifth Edition (pg. 350), where they recommended these levels of significance. Furthermore, certain criterions will help us distinguish the more accurate model. These criterions that we will be testing each model are, Mallow C_p criterion, Akaike Information criterion (AIC), Bayesian Information criterion (BIC) and PRESS criterion (PRESS). For AIC, BIC and PRESS, we are looking for the smallest number to represent the best model, while for C_p we determine the smallest C_p relative to the number of betas in our model will determine a good model. We will also be looking at the R squared, and adjusted R squared as factors in choosing the more accurate models. For these, we will look for the highest number to determine a good model. Note some models may have pass certain criteria while failing in others. In our discussion of results section, we will discuss all the criterions and then recommend the most accurate model based on all the testing.

Before moving onto our results section, multicollinearity can be a factor in the models. To test for multicollinearity, we will use Variance Inflation Factor to determine whether multicollinearity exists. Any variable with a factor above 10, will be deemed an issue. If such problem occurs, our best method is to drop one of the variables that shows problems of multicollinearity.

Results

Full Model

Predicted Model:

$$BF = -9.76 + 1.47MA + 4.087OA + 0.29BMI - 0.58N - 0.13C + 0.22T + 0.88AB - 0.35H - 0.14K + 0.09AN + 0.08B + 0.46F - 1.98W$$

Individual T-test:

$H_0: B_i = 0 \qquad H_a: B_i \neq 0$

Performing our individual t-test, we see that some of our p-values of the explanatory variables seems to be big which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis (Appendix 2.1).

Overall F-test:

$H_0: B_1 = B_2 = B_3 \ldots B_n = 0 \qquad H_a: B_1 = B_2 = B_3 \ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 52.48 (Appendix 2.1), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7438 (Appendix 2.1), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Residual Plots and Normality Plots:

Based on the Residual Plots (Appendix 2.3), we do see an influential point in BMI, Neck, Chest etc. However overall, we do not see any patterns of concern.

- The plot of e vs our predicted y looks random and bounded by a horizontal band, and these suggest that the linear model is fine, and the random errors have mean 0 and constant variance are not violated.
- Rstudent vs predicted y suggests that there might be some outliers
- The QQ plot looks relatively straight, and this suggest there is no significant violations of normality.
- The plot of e vs each regressor suggests the model is fine and the random errors seem to have mean 0 and constant variance are not violated.

Overall, there is no strong evidence that the model assumptions are not satisfied.

Multicollinearity Check:

Based off VIF, we have multicollinearity issues with BMI (11.501), abdomen (11.74) and hip (10.59) (Appendix 2.4).

Influential Points:

Observation 39 seems to be a heavy influential point. We also clearly identify an outlier based on the residual plots. Now we are going to consider a model without the influential point (Appendix 2.5).

Full Model without Influential Point

Predicted Model:

$$\boldsymbol{BF} = -9.76 + 1.47\boldsymbol{MA} + 4.087\boldsymbol{OA} + 0.29\boldsymbol{BMI} - 0.58\boldsymbol{N} - 0.13\boldsymbol{C} + 0.22\boldsymbol{T} + 0.88\boldsymbol{AB} - 0.35\boldsymbol{H} - 0.14\boldsymbol{K} + 0.09\boldsymbol{AN} + 0.08\boldsymbol{B} + 0.46\boldsymbol{F} - 1.98\boldsymbol{W}$$

Individual T-test:

H_0: $B_i = 0$ H_a: $B_i \neq 0$

Performing our individual t-test (Appendix 3.2), we see that some of our p-values of the explanatory variables seems to be pretty big which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis.

Overall F-test:

$H_0 : B_1 = B_2 = B_3 \ldots\ldots B_n = 0$ $H_a : B_1 = B_2 = B_3 \ldots\ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 52.56 (Appendix 3.2), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7519 (Appendix 3.2), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Residual Plots and Normality Plots:

By removing the influential point (observation 39), we do see a bigger improvement on our residual plots (Appendix 3.3). Overall assumptions have been met.

Check for Multicollinearity:

Based off VIF, we have multicollinearity issues with BMI (10.473) and abdomen (10.726) (-Appendix 3.4).

Influential Points:

Observation 84 seems to be a heavy influential point. We also clearly identify an outlier based on the residual plots (Appendix 3.5).

1st Recommended Model

Predicted Model (Based on Forward Selection):

$$\boldsymbol{BF} = 6.50 + 1.44\boldsymbol{MA} + 3.99\boldsymbol{OA} + 0.82\boldsymbol{AB} + 0.23\boldsymbol{BMI} - 0.61\boldsymbol{N} + 0.24\boldsymbol{T} - 0.37\boldsymbol{H} + 0.43\boldsymbol{F} - 2.02\boldsymbol{W}$$

Individual T-test:

H_0: $B_i = 0$ H_a: $B_i \neq 0$

Performing our individual t-test (Appendix 6.1), we see that some of our p-values of the explanatory variables seems to be big, which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis.

Overall F-test:

$H_0: B_1 = B_2 = B_3 \ldots B_n = 0$ $\quad\quad$ $H_a: B_1 = B_2 = B_3 \ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 75.96(Appendix 6.1), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.741 (Appendix 6.1), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Check for Multicollinearity:

Based off VIF, we have no multicollinearity (Appendix 6.1).

Influential Points:

Once again, we can see observation 39 be a possible influential point (Appendix 6.1).

2nd Recommended Model

Predicted Model (Based off Backward and Stepwise Selection):

$$\boldsymbol{BF} = 3.18 + 1.42\boldsymbol{MA} + 4.07\boldsymbol{OA} + 0.86\boldsymbol{AB} - 0.56\boldsymbol{N} + 0.26\boldsymbol{T} - 0.35\boldsymbol{H} + 0.45\boldsymbol{F} - 2.06\boldsymbol{W}$$

Individual T-test:

$H_0: B_i = 0$ $\quad\quad$ $H_a: B_i \neq 0$

Performing our individual t-test (Appendix 6.2), we see one of our p-values pretty big which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis.

Overall F-test:

$H_0: B_1 = B_2 = B_3 \ldots B_n = 0$ $\quad\quad$ $H_a: B_1 = B_2 = B_3 \ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 85.27 (Appendix 6.2), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7397 (Appendix 6.2), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Check for Multicollinearity:

Based off VIF, we have no multicollinearity issues (Appendix 6.2).

Influential Points:

Once again, we can see observation 39 be a possible influential point (Appendix 6.2).

3rd Recommended Model (without observation 39)

Predicted Model (Based off Forward Selection):

$$BF = 6.65 + 0.82\mathbf{AB} - 2.03\mathbf{W} + 1.70\mathbf{MA} + 4.03\mathbf{OA} + 0.62\mathbf{BMI} - 0.20\mathbf{C} - 0.25\mathbf{H} - 0.40\mathbf{N} + 0.28\mathbf{F} + 0.15\mathbf{T}$$

Individual T-test:

$H_0: B_i = 0 \qquad H_a: B_i \neq 0$

Performing our individual t-test (Appendix 6.3), we see some of our p-values pretty big which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis.

Overall F-test:

$H_0 : B_1 = B_2 = B_3 \ldots B_n = 0 \qquad H_a : B_1 = B_2 = B_3 \ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 71.41 (Appendix 6.3), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7508 (Appendix 6.3), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Check for Multicollinearity:

Based off VIF, we have multicollinearity issues with hip (10.42) (Appendix 6.3).

Influential Points:

Without observation 39, we can see that there are some possible influential points, however they are not that drastic. Hence, they may be potential influential points.

4th Recommended Model (without observation 39)

Predicted Model (Based off Backward Selection)

$$BF = 0.07 + 0.74\mathbf{AB} - 2.36\mathbf{W} + 1.92\mathbf{MA} + 4.32\mathbf{OA} + 0.62\mathbf{BMI} - 0.23\mathbf{C}$$

Individual T-test:

$H_0: B_i = 0 \qquad H_a: B_i \neq 0$

Performing our individual t-test (Appendix 6.4), we see that all of our p-values are small which means that there is enough evidence that the model is significant and hence we reject the null hypothesis.

Overall F-test:

$H_0 : B_1 = B_2 = B_3 \ldots B_n = 0 \qquad H_a : B_1 = B_2 = B_3 \ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 115.9 (Appendix 6.4), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7426 (Appendix 6.4), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Check for Multicollinearity:

Based off VIF, we have no multicollinearity issues (Appendix 6.4).

Influential Points:

Without observation 39, we can see that there are some possible influential points, however they are not that drastic. Hence, they may be potential influential points (Appendix 6.4).

5th Recommended Model (without observation 39)

Predicted Model (Based off Stepwise Selection):

$$\boldsymbol{BF} = 5.12 + 0.79\boldsymbol{AB} - 2.18\boldsymbol{W} + 1.62\boldsymbol{MA} + 3.57\boldsymbol{OA} + 0.70\,\boldsymbol{BMI} - 0.23\boldsymbol{C} - 0.15\boldsymbol{H}$$

Individual T-test:

H_0: $B_i = 0$ $\qquad$ H_a: $B_i \neq 0$

Performing our individual t-test (Appendix 6.5), we see that one of our p-value of the explanatory variable seems to be pretty big which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis.

Overall F-test:

$H_0 : B_1 = B_2 = B_3 \ldots B_n = 0$ $\qquad$ $H_a : B_1 = B_2 = B_3 \ldots B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 100.1 (Appendix 6.5), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7448 (Appendix 6.5), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Check for Multicollinearity:

Based off VIF, we have multicollinearity issues with abdomen (10.114110) (Appendix 6.5).

Influential Points:

Without observation 39, we can see that there are some possible influential points, however there are not that drastic. Hence, they may be potential influential points (Appendix 6.5).

6th Recommended Model (without observation 39):

Predicted Model (Alternate Stepwise Regression code):

$$\boldsymbol{BF} = 6.52 + 0.82\boldsymbol{AB} - 2.04\boldsymbol{W} + 1.55\boldsymbol{MA} + 3.64\boldsymbol{OA} + 0.68\,\boldsymbol{BMI} - 0.22\boldsymbol{C} - 0.17\boldsymbol{H} - 0.36\boldsymbol{N} + 0.30\boldsymbol{F}$$

Individual T-test:

H_0: B_i =0 H_a: $B_i \neq 0$

Performing our individual t-test (Appendix 6.6), we see that one of our p-value of the explanatory variable seems to be big which means that there is no evidence that the model is insignificant and hence we fail to reject the null hypothesis.

Overall F-test:

$H_0 : B_1 = B_2 = B_3 \ldots.. B_n = 0$ $H_{a\,:}\ B_1 = B_2 = B_3 \ldots.. B_n \neq 0$ (not all zero)

Our overall model seems to be significant based on our F-statistic, 79.03 (Appendix 6.6), with a p-value roughly zero. This entails that we have strong evidence to reject the null hypothesis. However even though our R squared is 0.7493 (Appendix 6.6), some of our independent variable p-values are insignificant, which is a contradiction of our observation.

Multicollinearity check:

Based off VIF, we have multicollinearity issues with Abdomen (10.423726) (Appendix 6.6).

Influential Points:

Without observation 39, we can see that there are some possible influential points, however they are not that drastic. Hence, they may be potential influential points (Appendix 6.6).

Summarized Table of Criterions of for each Recommended Model (Appendix 7):

Model	Reg	R squared	Adjusted R squared	AIC	BIC	VIF	C_P	PRESS
1	9	0.741	0.7312	1439.219	1477.911	NA	6.585873	4807.058
2	8	0.7397	0.7311	1438.38	1473.555	NA	5.69308	4740.067
3(ob39rem)	10	0.7508	0.7403	1423.09	1465.251	Hip (10.42)	7.049047	4740.067
4(ob39rem)	6	0.7426	0.7362	1423.151	1451.258	NA	6.814494	4469.812
5(ob39rem)	7	0.7448	0.7374	1422.947	1454.567	Abdomen (10.11)	6.666045	4471.351
6(ob39rem)	9	0.7493	0.7398	1422.59	1461.237	Abdomen (10.42)	6.474756	4453.317

Table 1

Discussion of Results

Based on Table 1 in the results section, we see certain models are more significant than other ones. In this discussion we will recommend 2 models where each would come from the full dataset, and the same dataset without observation 39. In choosing the "best" model, we would recommend Model 2 and Model 4 (observation 39 removed).

Firstly, Model 2 seems the best model based off of the AIC, BIC, the C_p and the Press residuals. Other than C_p, we see that all those criterions are smaller than the criterions of Model 1, which shows that this accurately describes body fat percentage. Furthermore, we have one less regressor than Model 1, which simplifies our model. When looking at the adjusted R squared, although it is lower than Model 1, it is a 0.0001 change, which has little to none effect in significance. Hence, we believe that Model 2 best describes Body fat percentage with the full dataset.

Lastly, we recommend Model 4 from the dataset without observation 39. By removing such observation, one must be aware that there might not be enough data on this certain influential point. We highly recommend finding more data to explore whether this could be an error in the data, or whether this is a true outlier. With more data related to this observation, we could have a more accurate understanding of body fat percentage. However, based on Table 1, we see that multicollinearity does not exist while the other 3 models had multicollinearity issues. Furthermore, with less regressors we get a simpler model to explain body fat percentage with a decent R squared value. Finally looking at Mallow's C_p, we see that this is the closet number with respect to the number of betas in the model, which shows that it is quite significant.

Conclusion

After a thorough analysis of different models including the full model, we personally would recommend a subset of the full model, as it is much simpler while still being as effective as the full model. When using the full dataset, we recommend a model using the following regressors, categorical ages, abdomen, neck, thigh, hip, forearm, and wrist. With these regressors, we can better predict the body fat percentage based on a set of measurements of a human body and thus can be generalised to any population. There is no multicollinearity present between all regressors. Although model 4 is also predicting satisfactorily, it is doing so because an observation has been removed. Removing data to fix an issue is always questionable in statistics as it might be a valid observed data. Thus, we recommend that for our project, model 2 from the result section of the report, is best suited to predict the body fat percentage and it can be used as a general model for any given population.

References

1. Neporent, Liz. “Why Legendary Bodybuilder Who Died With Almost Zero Body Fat Lives On.” ABC News, ABC News Network, 25 Mar. 2015, abcnews.go.com/Health/legendary-bodybuilder-died-body-fat-lives/story?id=29899438

2. Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis*. Oxford: Wiley-Blackwell, 2013.

3. Collinsakal. “Body Fat Eda + Prediction in R.” Kaggle. Kaggle, September 16, 2021. https://www.kaggle.com/collinsakal/body-fat-eda-prediction-in-r/data.

Appendix 1: Functions, Imported Libraries and Datasets in R

1. 1 Dataset:

Dataset

Note in R, this dataset variable is known as ***df0***

1. 2 Manual Functions used in R:

```
PRESS <- function(linear.model) {
  #' calculate the predictive residuals
  pr <- residuals(linear.model)/(1-lm.influence(linear.model)$hat)
  #' calculate the PRESS
  PRESS <- sum(pr^2)
  return(PRESS)
}
```

1. 3 Imported Libraries:

- library(tidyverse)
- library(car)
- library(matlib)
- library(leaps)
- library(MASS)
- library(ggplot2)
- library(qqplotr)
- library(SignifReg)
- library(corrplot)
- library(coefplot)
- library(moments)
- library(olsrr)
- library(qpcR)

1. 4 Removal of impossible observations:

```
df0 %>% filter(BodyFat <= 2)
df0 %>% filter(Height == 29.5)
df1 = df0 %>% filter(Height != 29.5) %>% filter(BodyFat >=2)
```

1. 5 Creating BMI and Age categories Dataset:

```
BMI = (df1$Weight/(df1$Height)^2)*703
BMI = round(BMI, digits = 2)
catAge = cut(df1$Age, breaks = c(20,40,60,90), labels = c("Young", "Middle", "Older"), right =
FALSE)
levels(catAge)
df2 = cbind(df1, BMI, catAge)
df2 = subset(df2, select= -c(Weight, Height, Age))
view(df2)
```

*Note that **df2** still includes all 249 observations, which means this includes all influential points and outliers*

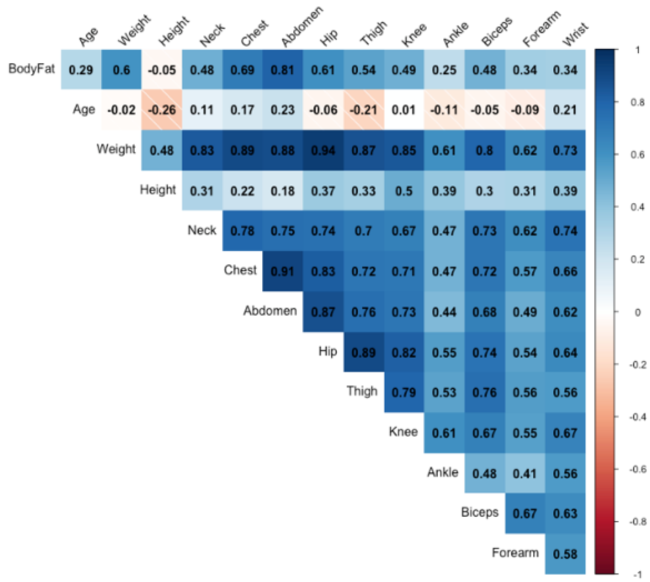
1. 6 Removal of influential point, observation 39 dataset:

```
df2 %>% filter(df2$BMI > 48)
df3_ob39rem = df2 %>% filter(BMI <= 41)
view(df3_ob39rem)
```

1. 7 Correlation Matrix:

```
corrplot(cor(df0), type = "upper",
         order = "original",
         method = "shade",
         tl.col = "black",
         tl.srt = 45,
         addCoef.col = "black",
         diag = FALSE)
```

Output:



Appendix 2: Full Model

2. 1 Computing the Estimated Model for a dataset:

fit0 = lm(BodyFat ~ ., data = df2)

2. 2 Summary of Estimates, standard Error t test and p values:

summary(fit0)
anova(fit0)

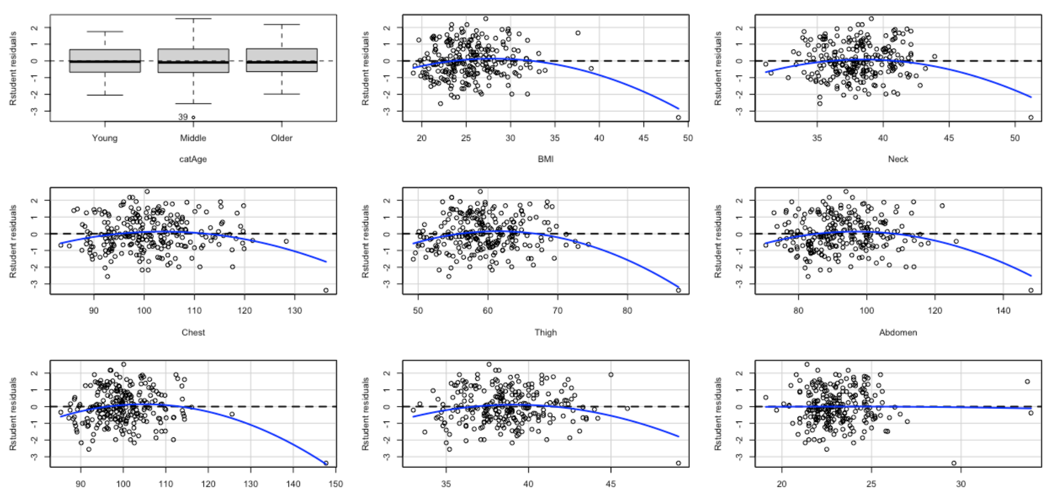
Output:

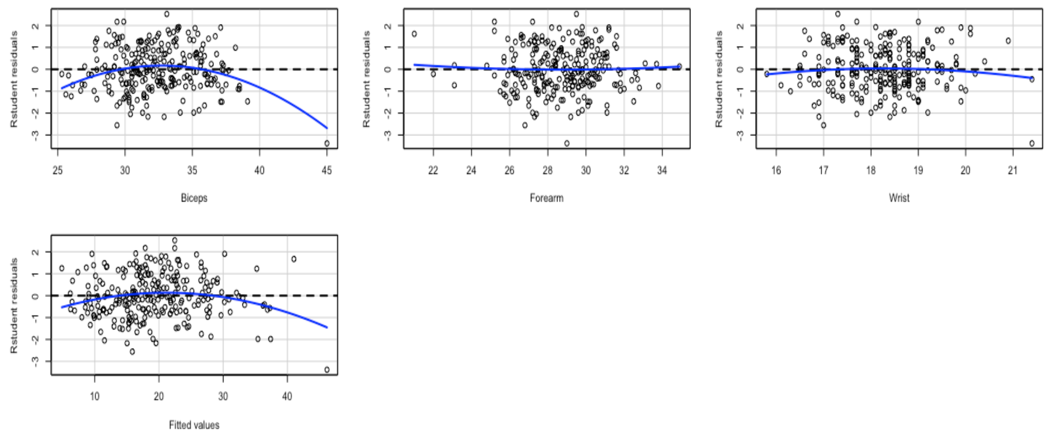
```
Call:
lm(formula = BodyFat ~ ., data = df2)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.9607  -2.8610  -0.2731   2.9353  10.4398 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   9.75501    7.44445   1.310 0.191349    
Neck         -0.57515    0.22161  -2.595 0.010045 *  
Chest        -0.13221    0.09643  -1.371 0.171675    
Abdomen       0.88154    0.08701  10.131  < 2e-16 ***
Hip          -0.35112    0.12515  -2.806 0.005444 ** 
Thigh         0.21912    0.14308   1.531 0.126991    
Knee         -0.14276    0.23729  -0.602 0.548004    
Ankle         0.08747    0.21604   0.405 0.685930    
Biceps        0.08413    0.16958   0.496 0.620291    
Forearm       0.45733    0.19737   2.317 0.021356 *  
Wrist        -1.97939    0.51761  -3.824 0.000168 ***
BMI           0.29223    0.25303   1.155 0.249293    
catAgeMiddle  1.47187    0.69079   2.131 0.034153 *  
catAgeOlder   4.08692    1.16324   3.513 0.000531 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.265 on 235 degrees of freedom
Multiple R-squared:  0.7438,    Adjusted R-squared:  0.7296
```

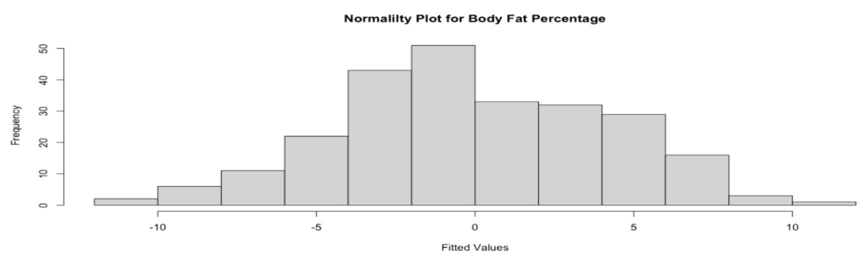
```
F-statistic: 52.48 on 13 and 235 DF,  p-value: < 2.2e-16

Analysis of Variance Table
Response: BodyFat
           Df Sum Sq Mean Sq  F value    Pr(>F)    
Neck        1 3878.6  3878.6 213.2639 < 2.2e-16 ***
Chest       1 4319.0  4319.0 237.4802 < 2.2e-16 ***
Abdomen     1 3284.3  3284.3 180.5906 < 2.2e-16 ***
Hip         1  354.3   354.3  19.4814 1.549e-05 ***
Thigh       1   53.5    53.5   2.9442  0.087502 .  
Knee        1   25.9    25.9   1.4234  0.234043    
Ankle       1    6.5     6.5   0.3587  0.549799    
Biceps      1   17.9    17.9   0.9857  0.321816    
Forearm     1   34.9    34.9   1.9192  0.167261    
Wrist       1  159.9   159.9   8.7931  0.003335 ** 
BMI         1   47.2    47.2   2.5953  0.108526    
catAge      2  224.5   112.3   6.1724  0.002440 ** 
Residuals 235 4273.9    18.2                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

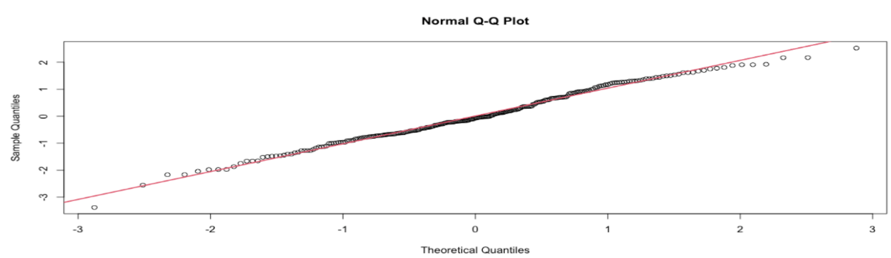
2. 3 Residual Plots

residualPlots(fit0, type = "rstudent")
par(mfrow = c(2,1))
hist(fit0$resid, main = "Normalilty Plot for Body Fat Percentage", xlab = "Fitted Values")
t_i = rstudent(fit0)
qqnorm(t_i)
qqline(t_i, col = 2, lwd = 2)









2. 4 Variance Inflation Factor

vif(newFit0)

Output:

```
             GVIF Df GVIF^(1/(2*Df))
catAge   1.886652  2        1.171987
BMI     10.472630  1        3.236144
Neck     3.624668  1        1.903856
Chest    8.659464  1        2.942697
Thigh    6.636095  1        2.576062
Abdomen 10.725570  1        3.274992
Hip      9.320025  1        3.052872
Knee     4.005767  1        2.001441
Ankle    1.723674  1        1.312887
Biceps   3.249949  1        1.802761
Forearm  2.369913  1        1.539452
Wrist    3.007311  1        1.734160
```

2. 5 Influential Points

summary(influence.measures(fit0))

Output:

```
Potentially influential observations of
```

```
          lm(formula = BodyFat ~ ., data = df2) :

dfb.1_ dfb.Neck dfb.Chst dfb.Abdm dfb.Hip dfb.Thgh dfb.Knee dfb.Ankl  dfb.Bcps dfb.Frrm
5    0.04  -0.07    -0.05     0.07    -0.05    0.00     0.06     0.00     0.02    -0.01
31   0.04  -0.04    -0.02    -0.03    -0.01    0.03     0.06    -0.24    -0.02     0.03
39   0.37  -0.79     0.82     0.54    -0.96    0.49    -0.01    -0.22    -0.16     1.10_*
41   0.06   0.06    -0.03     0.03    -0.09   -0.02     0.15    -0.02     0.08    -0.03
53  -0.01   0.11     0.00    -0.04     0.01    0.11    -0.03     0.01    -0.35     0.13
85   0.05  -0.02    -0.08     0.01    -0.02   -0.09    -0.21     1.05_*   0.13     0.00
105 -0.02   0.07     0.02    -0.01     0.01    0.02    -0.02    -0.01    -0.01    -0.02
158  0.00   0.00    -0.01     0.01     0.01   -0.01     0.00     0.00    -0.02     0.06
173 -0.09   0.03     0.21    -0.09    -0.01    0.06     0.07    -0.18     0.14    -0.89
203  0.00   0.02     0.00    -0.01    -0.02    0.01     0.00     0.00     0.04    -0.08
204  0.19   0.20    -0.16    -0.07     0.05   -0.13     0.04    -0.11     0.04     0.12
213  0.29  -0.02    -0.14     0.11     0.06   -0.26    -0.22     0.05    -0.03    -0.02
221 -0.19   0.02     0.00    -0.01     0.04   -0.06     0.14    -0.15     0.00     0.01
244  0.01   0.00     0.00     0.01    -0.01    0.02    -0.03     0.02    -0.01    -0.02
    dfb.Wrst dfb.BMI dfb.ctAM dfb.ctAO dffit   cov.r   cook.d hat
5    0.01     0.01   -0.06    -0.05     0.13    1.19_*  0.00   0.11
31   0.06     0.05    0.03     0.01    -0.26    1.53_*  0.00   0.31_*
39   0.22    -1.05_* -0.41    -0.10    -2.74_*  0.90    0.51   0.40_*
41  -0.10    -0.03   -0.03     0.00    -0.24    1.35_*  0.00   0.22_*
53  -0.07     0.08    0.00     0.04    -0.40    1.21_*  0.01   0.16
85  -0.31     0.06    0.13     0.35     1.12_*  1.45_*  0.09   0.36_*
105 -0.04    -0.04    0.00     0.02    -0.09    1.28_*  0.00   0.18_*
158 -0.02     0.00   -0.01     0.00     0.06    1.31_*  0.00   0.19_*
173  0.38    -0.03   -0.26    -0.30     0.99_*  1.25_*  0.07   0.27_*
203  0.02     0.00    0.01    -0.01     0.09    1.29_*  0.00   0.18_*
204 -0.28     0.21    0.14     0.05     0.50    0.76_*  0.02   0.04
213 -0.03     0.38   -0.02    -0.19     0.74_*  1.08    0.04   0.16
221  0.10     0.00   -0.15    -0.06    -0.37    0.74_*  0.01   0.02
244  0.02     0.00    0.00    -0.02    -0.06    1.18_*  0.00   0.10
Potentially influential observations of
          lm(formula = BodyFat ~ ., data = df2) :

    dfb.1_ dfb.Neck dfb.Chst dfb.Abdm dfb.Hip dfb.Thgh dfb.Knee dfb.Ankl dfb.Bcps dfb.Frrm dfb.Wrst dfb.BMI
dfb.ctAM dfb.ctAO dffit   cov.r   cook.d hat
5    0.04  -0.07    -0.05     0.07    -0.05    0.00     0.06     0.00     0.02    -0.01     0.01     0.01   -
0.06    -0.05     0.13    1.19_*  0.00   0.11
31   0.04  -0.04    -0.02    -0.03    -0.01    0.03     0.06    -0.24    -0.02     0.03     0.06     0.05
0.03     0.01    -0.26    1.53_*  0.00   0.31_*
39   0.37  -0.79     0.82     0.54    -0.96    0.49    -0.01    -0.22    -0.16     1.10_*   0.22    -1.05_* -
0.41    -0.10    -2.74_*  0.90    0.51   0.40_*
41   0.06   0.06    -0.03     0.03    -0.09   -0.02     0.15    -0.02     0.08    -0.03    -0.10    -0.03   -
0.03     0.00    -0.24    1.35_*  0.00   0.22_*
53  -0.01   0.11     0.00    -0.04     0.01    0.11    -0.03     0.01    -0.35     0.13    -0.07     0.08
0.00     0.04    -0.40    1.21_*  0.01   0.16
85   0.05  -0.02    -0.08     0.01    -0.02   -0.09    -0.21     1.05_*   0.13     0.00    -0.31     0.06
0.13     0.35     1.12_*  1.45_*  0.09   0.36_*
105 -0.02   0.07     0.02    -0.01     0.01    0.02    -0.02    -0.01    -0.01    -0.02    -0.04    -0.04
0.00     0.02    -0.09    1.28_*  0.00   0.18_*
158  0.00   0.00    -0.01     0.01     0.01   -0.01     0.00     0.00    -0.02     0.06    -0.02     0.00   -
0.01     0.00     0.06    1.31_*  0.00   0.19_*
173 -0.09   0.03     0.21    -0.09    -0.01    0.06     0.07    -0.18     0.14    -0.89     0.38    -0.03   -
0.26    -0.30     0.99_*  1.25_*  0.07   0.27_*
203  0.00   0.02     0.00    -0.01    -0.02    0.01     0.00     0.00     0.04    -0.08     0.02     0.00
0.01    -0.01     0.09    1.29_*  0.00   0.18_*
204  0.19   0.20    -0.16    -0.07     0.05   -0.13     0.04    -0.11     0.04     0.12    -0.28     0.21
0.14     0.05     0.50    0.76_*  0.02   0.04
213  0.29  -0.02    -0.14     0.11     0.06   -0.26    -0.22     0.05    -0.03    -0.02    -0.03     0.38   -
0.02    -0.19     0.74_*  1.08    0.04   0.16
221 -0.19   0.02     0.00    -0.01     0.04   -0.06     0.14    -0.15     0.00     0.01     0.10     0.00   -
0.15    -0.06    -0.37    0.74_*  0.01   0.02
244  0.01   0.00     0.00     0.01    -0.01    0.02    -0.03     0.02    -0.01    -0.02     0.02     0.00
0.00    -0.02    -0.06    1.18_*  0.00   0.10
```

Appendix 3: Full Model without observation 39

3. 1 Computing the Estimated Model for a dataset:

alt_fit0 = lm(BodyFat ~ ., data = df3_ob39rem)

3. 2 Summary of Estimates, standard error, t and f test, p values, ANOVA:

summary(alt_fit0)
anova(alt_fit0)

Output:

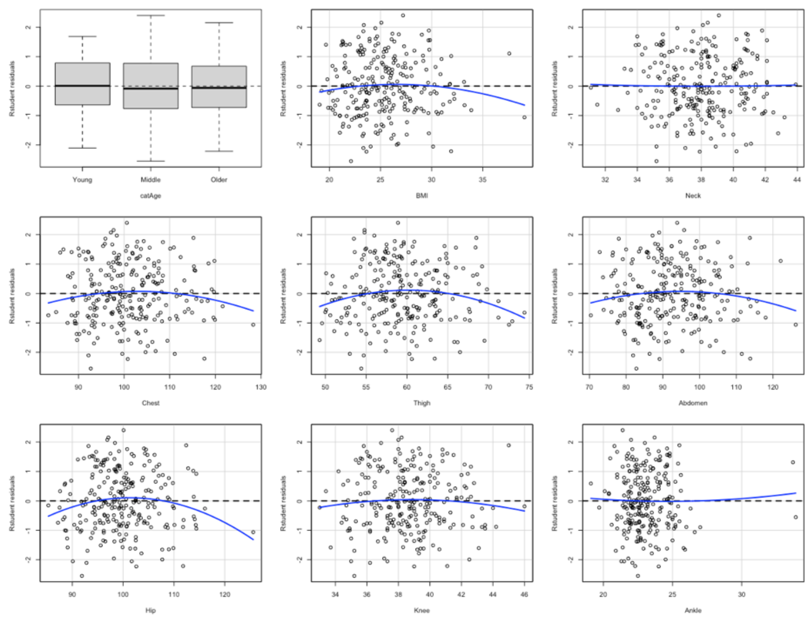
```
Call:
lm(formula = BodyFat ~ ., data = df3_ob39rem)

Residuals:
    Min      1Q  Median      3Q     Max
```

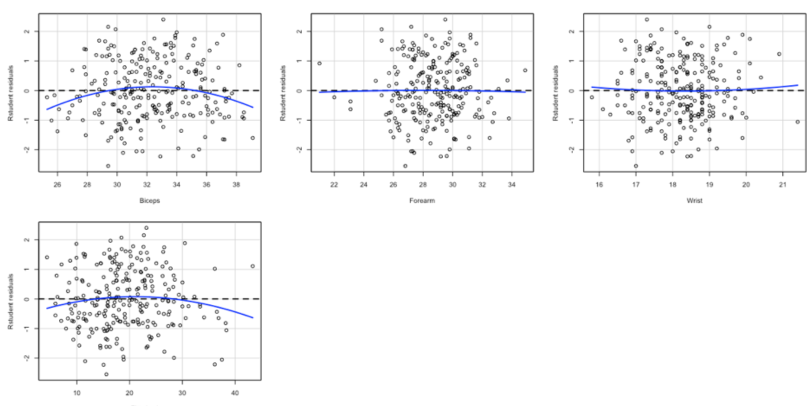
```
-10.408  -3.038  -0.236   3.130   9.720

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.05630    7.32808   0.963 0.336586
Neck         -0.40444    0.22265  -1.817 0.070573 .
Chest        -0.20954    0.09709  -2.158 0.031929 *
Abdomen       0.83520    0.08624   9.685  < 2e-16 ***
Hip          -0.23336    0.12732  -1.833 0.068088 .
Thigh         0.15098    0.14144   1.067 0.286876
Knee         -0.14062    0.23219  -0.606 0.545340
Ankle         0.13444    0.21185   0.635 0.526314
Biceps        0.11091    0.16613   0.668 0.505048
Forearm       0.24472    0.20310   1.205 0.229446
Wrist        -2.09191    0.50758  -4.121 5.23e-05 ***
BMI           0.55204    0.25924   2.129 0.034262 *
catAgeMiddle  1.75172    0.68100   2.572 0.010721 *
catAgeOlder   4.19732    1.13872   3.686 0.000283 ***
---
Signif. codes:  0 ‚Äò***‚Äô 0.001 ‚Äò**‚Äô 0.01 ‚Äò*‚Äô 0.05 ‚Äò.‚Äô 0.1 ‚Äò ‚Äô 1

Residual standard error: 4.173 on 234 degrees of freedom
Multiple R-squared:  0.7519,    Adjusted R-squared:  0.7381
F-statistic: 54.56 on 13 and 234 DF,  p-value: < 2.2e-16

Analysis of Variance Table

Response: BodyFat
           Df Sum Sq Mean Sq  F value    Pr(>F)
Neck        1 3658.9  3658.9 210.1166 < 2.2e-16 ***
Chest       1 4315.6  4315.6 247.8316 < 2.2e-16 ***
Abdomen     1 3538.9  3538.9 203.2296 < 2.2e-16 ***
Hip         1  208.7   208.7  11.9872 0.0006367 ***
Thigh       1   33.1    33.1   1.9028 0.1690864
Knee        1   41.0    41.0   2.3555 0.1261950
Ankle       1    3.9     3.9   0.2231 0.6371479
Biceps      1   15.5    15.5   0.8885 0.3468547
Forearm     1    6.3     6.3   0.3633 0.5472734
Wrist       1  177.8   177.8  10.2082 0.0015906 **
BMI         1  110.2   110.2   6.3257 0.0125715 *
catAge      2  240.2   120.1   6.8978 0.0012283 **
Residuals 234 4074.8    17.4
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

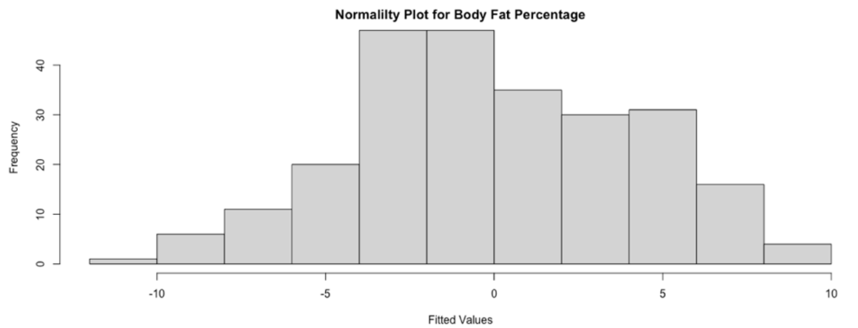
3. 3 Residual Plots

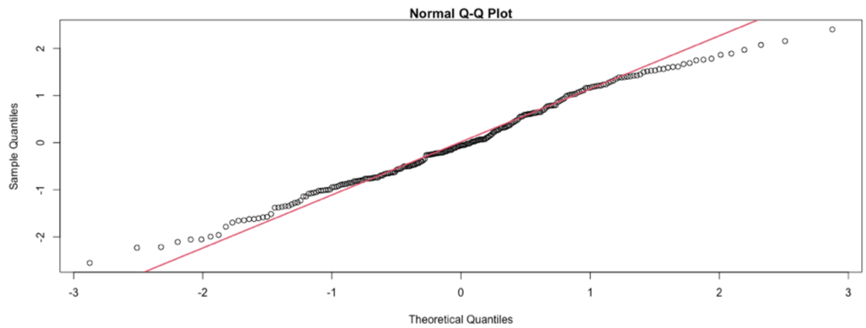
```
residualPlots(alt_fit0, type = "rstudent")
par(mfrow = c(2,1))
hist(alt_fit0$resid, main = "Normalilty Plot for Body Fat Percentage", xlab = "Fitted Values")
alt_t_i = rstudent(alt_fit0)
qqnorm(alt_t_i)
qqline(alt_t_i, col = 2, lwd = 2)
```

Output:


Rstudent residuals
Young
Middle
Older
catAge
BMI
Neck
Chest
Thigh
Abdomen
Hip
Knee
Ankle


Rstudent residuals
Biceps
Forearm
Wrist
Fitted values





3. 4 Variance Inflation Factor:

vif(alt_Fit0)

```
              GVIF Df GVIF^(1/(2*Df))
Neck      3.624668  1        1.903856
Chest     8.659464  1        2.942697
Abdomen  10.725570  1        3.274992
Hip       9.320025  1        3.052872
Thigh     6.636095  1        2.576062
Knee      4.005767  1        2.001441
Ankle     1.723674  1        1.312887
Biceps    3.249949  1        1.802761
Forearm   2.369913  1        1.539452
Wrist     3.007311  1        1.734160
BMI      10.472630  1        3.236144
catAge    1.886652  2        1.171987
```

3. 5 Influential Points

summary(influence.measures(alt_fit0))

Output:

```
Potentially influential observations of
     lm(formula = BodyFat ~ ., data = df3_ob39rem) :

    dfb.1_ dfb.Neck dfb.Chst dfb.Abdm dfb.Hip dfb.Thgh dfb.Knee dfb.Ankl dfb.Bcps dfb.Frrm dfb.Wrst
5    0.06  -0.09    -0.06     0.10    -0.07    0.00     0.09     0.00     0.02     0.00     0.01
31   0.06  -0.06    -0.02    -0.03    -0.02    0.05     0.08    -0.35    -0.03     0.05     0.09
40   0.17   0.09    -0.02     0.09    -0.25   -0.01     0.35    -0.06     0.17    -0.01    -0.23
52  -0.01   0.10    -0.01    -0.04     0.02    0.08    -0.02     0.01    -0.29     0.09    -0.06
84   0.03   0.00    -0.09     0.00     0.00   -0.09    -0.19     0.92     0.12    -0.03    -0.27
104 -0.01   0.02     0.01     0.00     0.00    0.01     0.00     0.00     0.00    -0.01    -0.01
157  0.01   0.00    -0.02     0.04     0.02   -0.06    -0.03     0.01    -0.13     0.32    -0.08
172 -0.07   0.06     0.07    -0.08     0.05    0.00     0.04    -0.09     0.09    -0.56     0.21
200  0.23  -0.01    -0.15     0.13    -0.22   -0.07     0.13    -0.08    -0.01    -0.04     0.04
202  0.00  -0.06     0.00     0.03     0.02    0.00    -0.01     0.00    -0.08     0.17    -0.03
203  0.17   0.21    -0.17    -0.09     0.08   -0.14     0.03    -0.10     0.05     0.07    -0.27
212  0.17   0.03    -0.14     0.04     0.10   -0.21    -0.15     0.05    -0.01    -0.08    -0.03
220 -0.20   0.03    -0.01    -0.02     0.05   -0.07     0.14    -0.14     0.00    -0.01     0.10
```

```
    dfb.BMI dfb.ctAM dfb.ctAO dffit   cov.r   cook.d hat
5    0.01   -0.08    -0.07     0.18    1.18_*  0.00   0.12
31   0.07    0.03     0.01    -0.38    1.52_*  0.01   0.32_*
40  -0.13   -0.10     0.00    -0.60    1.31_*  0.03   0.25_*
52   0.08    0.00     0.04    -0.33    1.23_*  0.01   0.17
84   0.08    0.12     0.31     0.99_*  1.50_*  0.07   0.36_*
104 -0.01    0.00     0.00    -0.02    1.29_*  0.00   0.18_*
157 -0.03   -0.07    -0.02     0.36    1.31_*  0.01   0.21_*
172  0.05   -0.12    -0.17     0.62    1.46_*  0.03   0.31_*
200  0.22   -0.20    -0.16    -0.41    0.82_*  0.01   0.03
202 -0.02   -0.02     0.02    -0.18    1.32_*  0.00   0.20_*
203  0.23    0.15     0.05     0.50    0.79_*  0.02   0.04
212  0.31    0.01    -0.12     0.54    1.22_*  0.02   0.19_*
220  0.01   -0.14    -0.06    -0.37    0.74_*  0.01   0.02
```

Appendix 6: Recommended Models

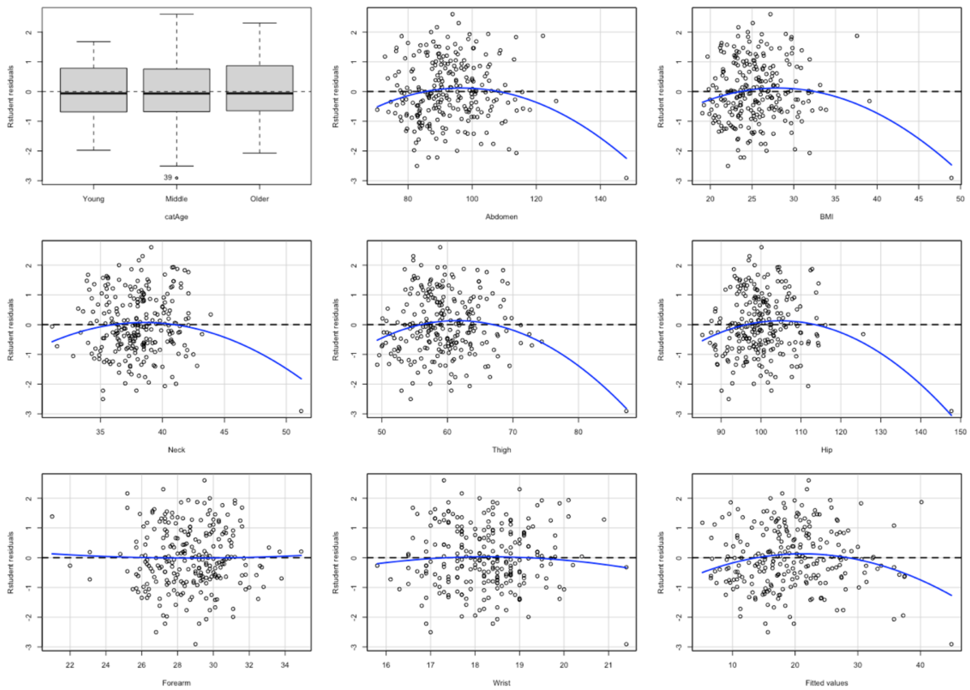
6. 1 Recommended Model 1

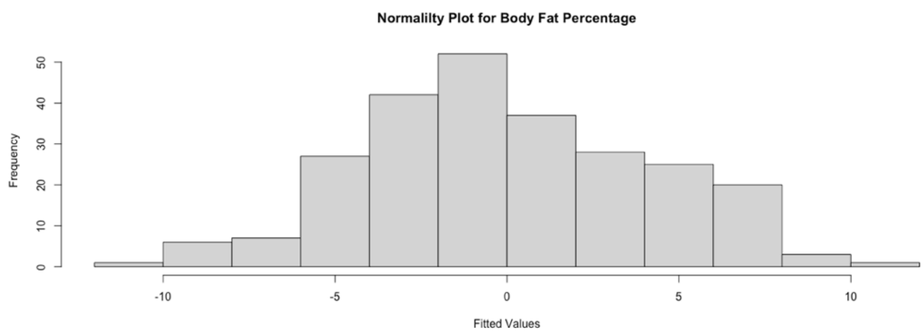
```
ReducFit = lm(BodyFat ~ catAge + Abdomen + BMI + Neck +
          Thigh + Hip + Forearm + Wrist, data = df2)
summary(ReducFit)
residualPlots(ReducFit3, type = "rstudent")
par(mfrow = c(2,1))
hist(ReducFit$resid, main = "Normality Plot for Body Fat Percentage", xlab ="Fitted Values")
ReducFit_t_i = rstudent(ReducFit)
qqnorm(ReducFit_t_i)
qqline(ReducFit_t_i, col = 2, lwd = 2)
vif(ReducFit)
summary(influence.measures(ReducFit))
```

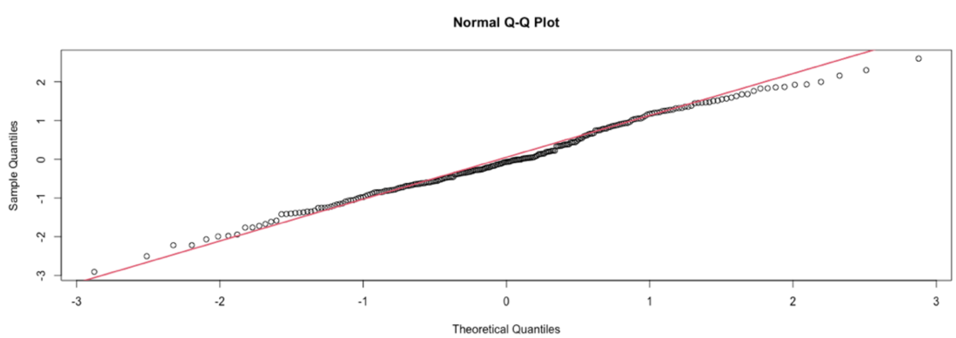
Output:

```
Call:
lm(formula = BodyFat ~ catAge + Abdomen + BMI + Neck + Thigh + 
    Hip + Forearm + Wrist, data = df2)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.4453  -2.8103  -0.2783   3.2649  10.7471 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   6.49618    7.01854   0.926 0.355601    
catAgeMiddle  1.44909    0.68080   2.129 0.034317 *  
catAgeOlder   3.98735    1.13297   3.519 0.000518 ***
Abdomen       0.81824    0.07698  10.629  < 2e-16 ***
BMI           0.23583    0.22313   1.057 0.291623    
Neck         -0.60645    0.21644  -2.802 0.005496 ** 
Thigh         0.23988    0.12823   1.871 0.062610 .  
Hip          -0.37065    0.11920  -3.109 0.002102 ** 
Forearm       0.42834    0.18376   2.331 0.020587 *  
Wrist        -2.02257    0.47965  -4.217 3.52e-05 ***
---
Signif. codes:  0 ,Äò***,Äô 0.001 ,Äò**,Äô 0.01 ,Äò*,Äô 0.05 ,Äò.,Äô 0.1 ,Äò ,Äô 1

Residual standard error: 4.252 on 239 degrees of freedom
Multiple R-squared:  0.741,     Adjusted R-squared:  0.7312 
F-statistic: 75.96 on 9 and 239 DF,  p-value: < 2.2e-16

           Test stat Pr(>|Test stat|)  
Abdomen      -1.5281          0.12783  
Wrist         0.3763          0.70700  
catAge                                 
BMI          -1.2510          0.21217  
Chest        -1.5499          0.12251  
Hip          -2.4134          0.01657 *
```

```
Neck          0.1711          0.86429
Forearm      -0.2238          0.82314
Thigh        -2.3551          0.01934 *
Tukey test   -1.6998          0.08917 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

            GVIF Df GVIF^(1/(2*Df))
catAge  1.767235  2        1.152985
Abdomen 9.243844  1        3.040369
BMI     8.996752  1        2.999459
Neck    3.750560  1        1.936636
Thigh   5.961570  1        2.441633
Hip     9.668531  1        3.109426
Forearm 1.868766  1        1.367028
Wrist   2.713806  1        1.647363
Potentially influential observations of
         lm(formula = BodyFat ~ catAge + Abdomen + BMI + Neck + Thigh +      Hip + Forearm +
Wrist, data = df2) :

    dfb.1_ dfb.ctAM dfb.ctAO dfb.Abdm dfb.BMI dfb.Neck dfb.Thgh dfb.Hip dfb.Frrm
15  -0.01   0.01     0.00    -0.01     0.00   -0.01    -0.05     0.04    0.00
39   0.53  -0.36    -0.09     0.87    -0.75   -0.58     0.29    -0.85    1.06_*
41   0.04  -0.01     0.02     0.02    -0.05    0.04     0.04    -0.03    0.00
80  -0.03  -0.07     0.14     0.09    -0.06    0.02    -0.08    -0.02   -0.11
105  0.00   0.00     0.00     0.00    -0.01    0.03     0.00     0.00   -0.01
158  0.00  -0.02    -0.01     0.00    -0.01    0.00    -0.03     0.01    0.08
173 -0.05  -0.20    -0.24     0.01     0.04    0.08     0.07    -0.01   -0.72
201  0.20  -0.18    -0.13     0.10     0.16   -0.03    -0.02    -0.21   -0.06
203  0.00   0.01    -0.01    -0.01     0.01    0.03     0.02    -0.02   -0.07
204  0.14   0.17     0.07    -0.15     0.18    0.20    -0.11     0.05    0.12
213  0.29  -0.05    -0.25     0.02     0.48   -0.06    -0.39     0.00   -0.10
221 -0.21  -0.12    -0.02     0.01    -0.05    0.03    -0.03     0.06    0.02
239  0.00   0.02    -0.03     0.00    -0.06    0.09     0.11    -0.09   -0.03
    dfb.Wrst dffit   cov.r   cook.d hat
15   0.01    -0.06    1.13_*  0.00   0.08
39   0.18    -2.17_*  1.15_*  0.46   0.36_*
41  -0.05    -0.12    1.17_*  0.00   0.11
80   0.13     0.44    0.87_*  0.02   0.04
105 -0.01    -0.03    1.23_*  0.00   0.15_*
158 -0.03     0.09    1.24_*  0.00   0.16_*
173  0.33     0.79_*  1.27_*  0.06   0.24_*
201  0.03    -0.35    0.87_*  0.01   0.02
203  0.02     0.08    1.21_*  0.00   0.14_*
204 -0.35     0.48    0.81_*  0.02   0.03
213 -0.09     0.76_*  1.05    0.06   0.14_*
221  0.09    -0.32    0.82_*  0.01   0.02
239 -0.01    -0.21    1.14_*  0.00   0.10
```







6. 2 Recommended Model 2

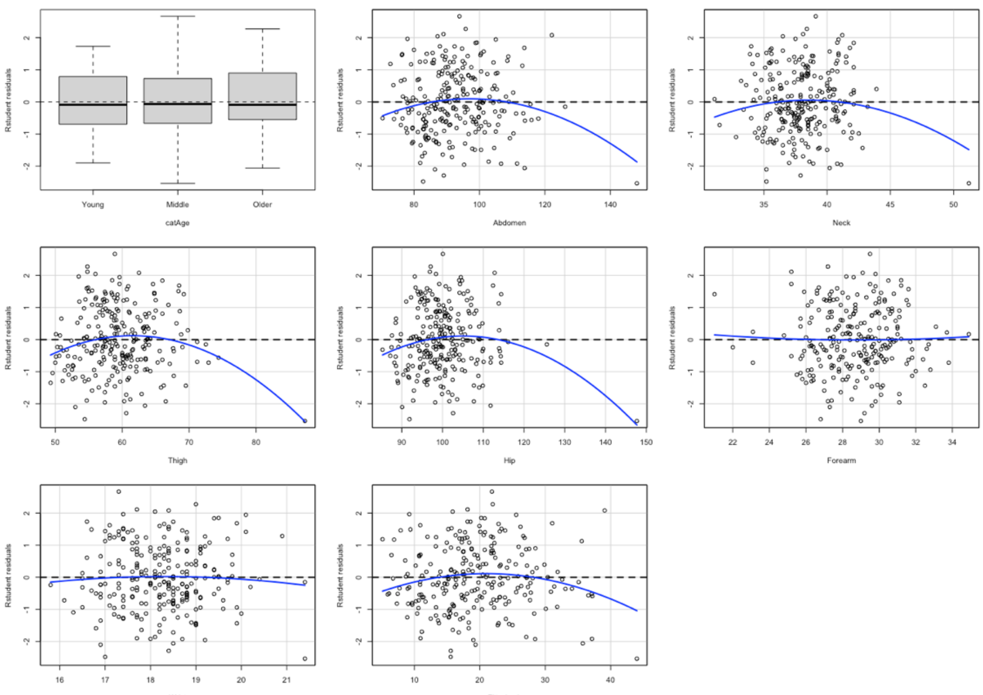
```
ReducFit2 = lm(BodyFat ~ catAge + Abdomen + Neck + Thigh + Hip + Forearm + Wrist, data = df2)
summary(ReducFit2)
residualPlots(ReducFit2, type = "rstudent")
par(mfrow = c(2,1))
hist(ReducFit2$resid, main = "Normalilty Plot for Body Fat Percentage", xlab ="Fitted Values")
```

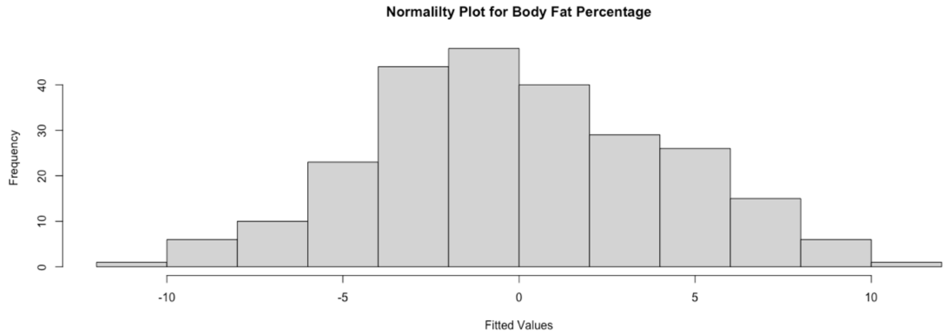
ReducFit2_t_i = rstudent(ReducFit2)
qqnorm(ReducFit2_t_i)
qqline(ReducFit2_t_i, col = 2, lwd = 2)
vif(ReducFit2)
summary(influence.measures(ReducFit2))

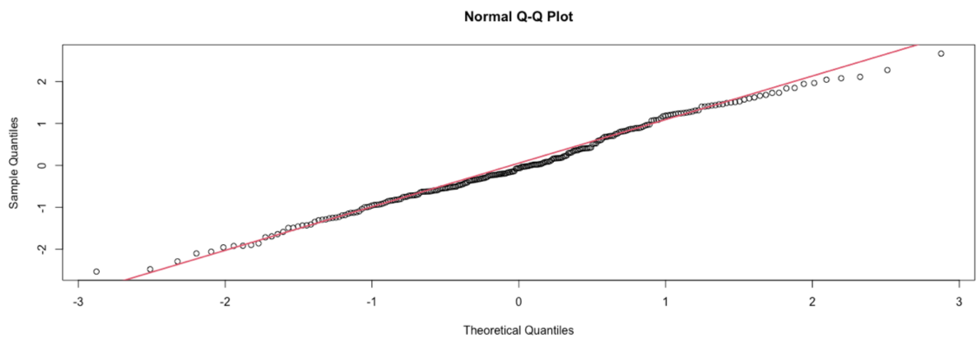
Output:

```
Call:
lm(formula = BodyFat ~ catAge + Abdomen + Neck + Thigh + Hip +
    Forearm + Wrist, data = df2)

Residuals:
     Min       1Q   Median       3Q      Max
-10.3485  -2.7146  -0.2677   3.1878  11.0470

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.17664    6.27809   0.506 0.613330
catAgeMiddle  1.41598    0.68024   2.082 0.038441 *
catAgeOlder   4.06608    1.13079   3.596 0.000393 ***
Abdomen       0.86359    0.06393  13.507  < 2e-16 ***
Neck         -0.56225    0.21241  -2.647 0.008659 **
Thigh         0.26213    0.12652   2.072 0.039350 *
Hip          -0.34943    0.11753  -2.973 0.003248 **
Forearm       0.44963    0.18270   2.461 0.014556 *
Wrist        -2.05617    0.47871  -4.295 2.54e-05 ***
---
Signif. codes:  0 ‚Äò***‚Äô 0.001 ‚Äò**‚Äô 0.01 ‚Äò*‚Äô 0.05 ‚Äò.‚Äô 0.1 ‚Äò ‚Äô 1

Residual standard error: 4.253 on 240 degrees of freedom
Multiple R-squared:  0.7397,    Adjusted R-squared:  0.7311
F-statistic: 85.27 on 8 and 240 DF,  p-value: < 2.2e-16

           Test stat Pr(>|Test stat|)
catAge
Abdomen      -2.4433         0.015278 *
Neck         -1.6755         0.095146 .
Thigh        -3.3128         0.001067 **
Hip          -3.1427         0.001885 **
Forearm       0.1943         0.846119
Wrist        -0.4545         0.649851
Tukey test   -2.3705         0.017762 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

            GVIF Df GVIF^(1/(2*Df))
catAge  1.740157  2        1.148543
Abdomen 6.372559  1        2.524393
Neck    3.610538  1        1.900142
Thigh   5.800879  1        2.408501
Hip     9.394407  1        3.065030
Forearm 1.846302  1        1.358787
Wrist   2.701888  1        1.643742
Potentially influential observations of
         lm(formula = BodyFat ~ catAge + Abdomen + Neck + Thigh + Hip +      Forearm + Wrist,
data = df2) :

    dfb.1_ dfb.ctAM dfb.ctAO dfb.Abdm dfb.Neck dfb.Thgh dfb.Hip dfb.Frrm dfb.Wrst
5    0.05  -0.05    -0.04     0.06    -0.07     0.03    -0.03    0.00     0.02
15  -0.01   0.01     0.00    -0.01    -0.01    -0.05     0.05    0.00     0.01
36   0.05   0.03     0.01     0.05    -0.02    -0.21     0.23    0.10    -0.23
39   0.82  -0.27    -0.12     0.45    -0.62     0.14    -0.84    0.83     0.19
41   0.03   0.00     0.01     0.00     0.02     0.01    -0.02    0.00    -0.02
105  0.00   0.00     0.00     0.01    -0.03     0.00     0.00    0.01     0.02
158  0.00  -0.02    -0.01     0.00     0.00    -0.03     0.01    0.06    -0.02
168 -0.03   0.05     0.03    -0.09     0.05    -0.10     0.10   -0.02     0.00
173 -0.08  -0.20    -0.24     0.05     0.09     0.08     0.00   -0.74     0.34
201  0.14  -0.19    -0.12     0.24     0.01     0.01    -0.19   -0.04     0.02
203 -0.01   0.01    -0.01    -0.01     0.04     0.03    -0.02   -0.09     0.03
204  0.07   0.16     0.09    -0.06     0.24    -0.08     0.09    0.14    -0.37
213  0.08  -0.08    -0.24     0.38     0.04    -0.34     0.08   -0.05    -0.13
221 -0.20  -0.12    -0.02    -0.02     0.02    -0.04     0.05    0.02     0.09
```

```
239  0.02   0.02    -0.03    -0.04     0.07     0.09    -0.08   -0.03    -0.01
    dffit   cov.r   cook.d hat
5    0.10    1.11_*  0.00   0.07
15  -0.07    1.13_*  0.00   0.08
36   0.40    1.11    0.02   0.11_*
39  -1.72_*  1.20_*  0.32   0.32_*
41  -0.05    1.14_*  0.00   0.09
105  0.04    1.20_*  0.00   0.14_*
158  0.07    1.23_*  0.00   0.16_*
168 -0.17    1.12_*  0.00   0.08
173  0.80_*  1.27_*  0.07   0.24_*
201 -0.32    0.87_*  0.01   0.02
203  0.10    1.20_*  0.00   0.14_*
204  0.46    0.82_*  0.02   0.03
213  0.63_*  0.96    0.04   0.08
221 -0.31    0.84_*  0.01   0.02
239 -0.17    1.14_*  0.00   0.10
```







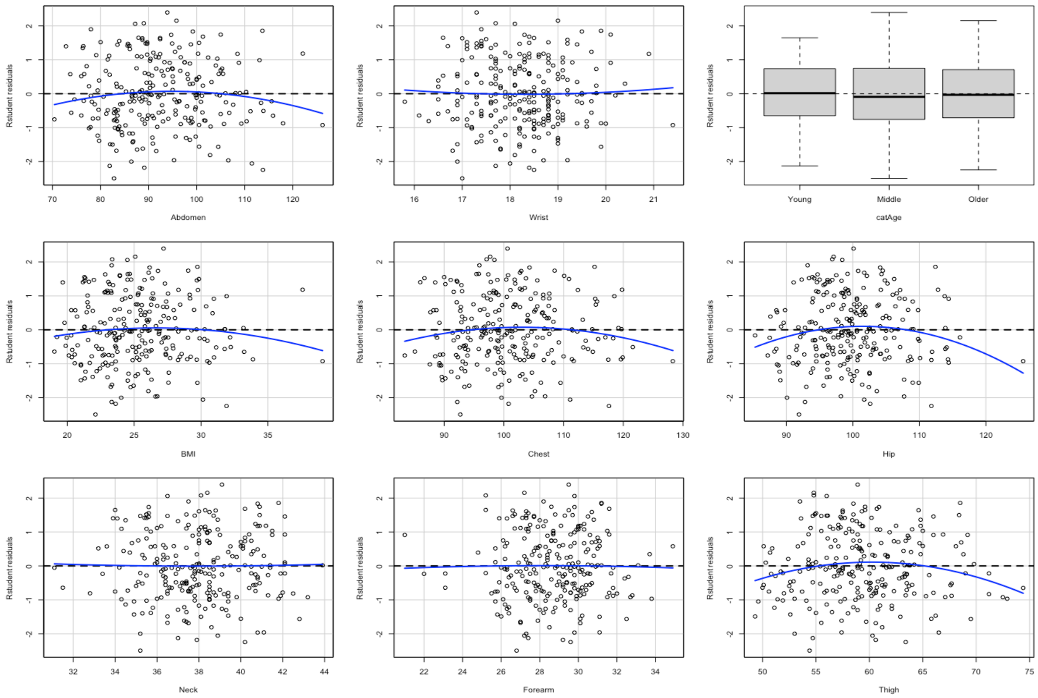
6. 3 Recommended Model 3

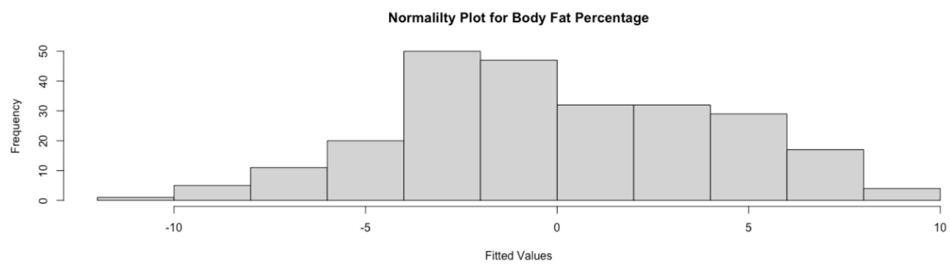
ReducFit3 = lm(BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest + Hip + Neck + Forearm + Thigh,
data = df3_ob39rem)
summary(ReducFit3)
vif(ReducFit3)
summary(influence.measures(ReducFit3))
residualPlots(ReducFit3, type = "rstudent")
par(mfrow = c(2,1))
hist(ReducFit3$resid, main = "Normalilty Plot for Body Fat Percentage", xlab ="Fitted Values")
ReducFit3_t_i = rstudent(ReducFit3)
qqnorm(ReducFit3_t_i)
qqline(ReducFit3_t_i, col = 2, lwd = 2)

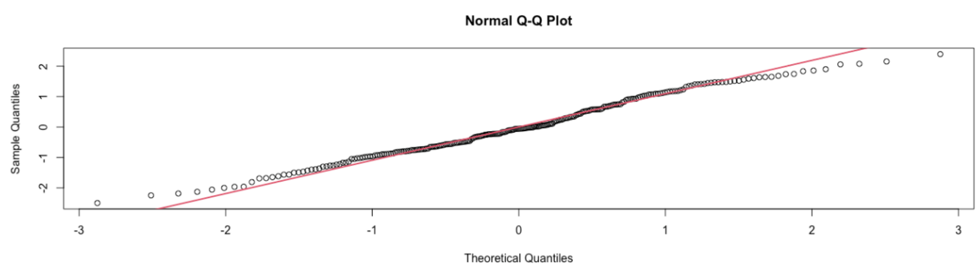
Output:

```
Call:
lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest + 
    Hip + Neck + Forearm + Thigh, data = df3_ob39rem)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.1811  -2.9862  -0.2251   3.0399   9.6613 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   6.65308    7.19581   0.925 0.356126    
Abdomen       0.82046    0.08466   9.691  < 2e-16 ***
Wrist        -2.02985    0.47111  -4.309 2.41e-05 ***
catAgeMiddle  1.69846    0.67038   2.534 0.011937 *  
catAgeOlder   4.03075    1.10817   3.637 0.000338 ***
BMI           0.61254    0.24577   2.492 0.013375 *  
Chest        -0.20395    0.09491  -2.149 0.032661 *  
Hip          -0.24828    0.12186  -2.037 0.042717 *  
```

```
Neck         -0.39687    0.21939  -1.809 0.071725 .
Forearm       0.27615    0.19206   1.438 0.151807
Thigh         0.15352    0.12805   1.199 0.231735
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.156 on 237 degrees of freedom
Multiple R-squared:  0.7508,    Adjusted R-squared:  0.7403
F-statistic: 71.41 on 10 and 237 DF,  p-value: < 2.2e-16

             GVIF Df GVIF^(1/(2*Df))
Abdomen 10.423728  1        3.228580
Wrist    2.612171  1        1.616221
catAge   1.794898  2        1.157471
BMI      9.490523  1        3.080669
Chest    8.344075  1        2.888611
Hip      8.608649  1        2.934050
Neck     3.548627  1        1.883780
Forearm  2.136872  1        1.461805
Thigh    5.483474  1        2.341682
Potentially influential observations of
         lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest +      Hip + Neck +
Forearm + Thigh, data = df3_ob39rem) :

    dfb.1_ dfb.Abdm dfb.Wrst dfb.ctAM dfb.ctAO dfb.BMI dfb.Chst dfb.Hip dfb.Neck
15  -0.01  -0.02     0.01     0.01     0.00     0.00    0.01     0.04   -0.01
36   0.02  -0.05    -0.21     0.04     0.01     0.03    0.08     0.21   -0.02
40   0.12   0.08    -0.14    -0.04     0.06    -0.19    0.04    -0.13    0.08
104 -0.01   0.00    -0.01     0.00     0.00    -0.01    0.01     0.00    0.02
157  0.03   0.05    -0.08    -0.07    -0.02    -0.03   -0.04     0.01   -0.02
172 -0.09  -0.08     0.20    -0.11    -0.15     0.05    0.08     0.05    0.08
202  0.01   0.02    -0.03    -0.01     0.01    -0.02   -0.01     0.01   -0.04
203  0.15  -0.08    -0.31     0.17     0.07     0.23   -0.17     0.09    0.23
212  0.19   0.03    -0.06    -0.01    -0.16     0.39   -0.17     0.06    0.03
220 -0.21   0.00     0.09    -0.11    -0.02    -0.03    0.00     0.07    0.04
235  0.02   0.05     0.02    -0.05    -0.14     0.04   -0.03     0.04   -0.01
    dfb.Frrm dfb.Thgh dffit cov.r   cook.d hat
15   0.00    -0.05    -0.07  1.14_*  0.00   0.09
36   0.04    -0.18     0.37  1.14_*  0.01   0.13
40   0.06     0.13    -0.37  1.17_*  0.01   0.14_*
104 -0.01     0.00    -0.02  1.26_*  0.00   0.17_*
157  0.24    -0.09     0.28  1.27_*  0.01   0.19_*
172 -0.55     0.04     0.59  1.43_*  0.03   0.30_*
202  0.09    -0.02    -0.10  1.25_*  0.00   0.16_*
203  0.09    -0.14     0.48  0.84_*  0.02   0.04
212 -0.11    -0.30     0.55  1.19_*  0.03   0.18_*
220  0.00    -0.03    -0.32  0.80_*  0.01   0.02
235 -0.02    -0.13    -0.18  1.14_*  0.00   0.09
           Test stat Pr(>|Test stat|)
Abdomen      -1.5281          0.12783
Wrist         0.3763          0.70700
catAge
BMI          -1.2510          0.21217
Chest        -1.5499          0.12251
Hip          -2.4134          0.01657 *
Neck          0.1711          0.86429
Forearm      -0.2238          0.82314
Thigh        -2.3551          0.01934 *
Tukey test   -1.6998          0.08917 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```







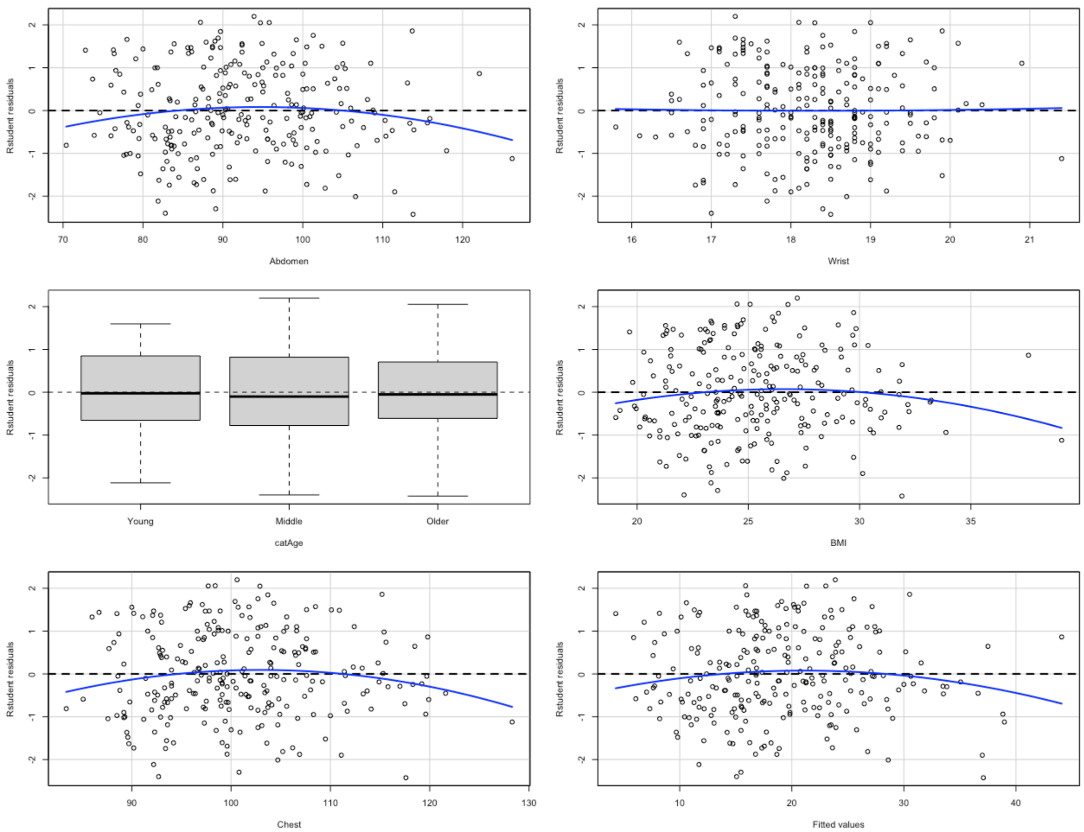
6. 4 Recommended Model 4

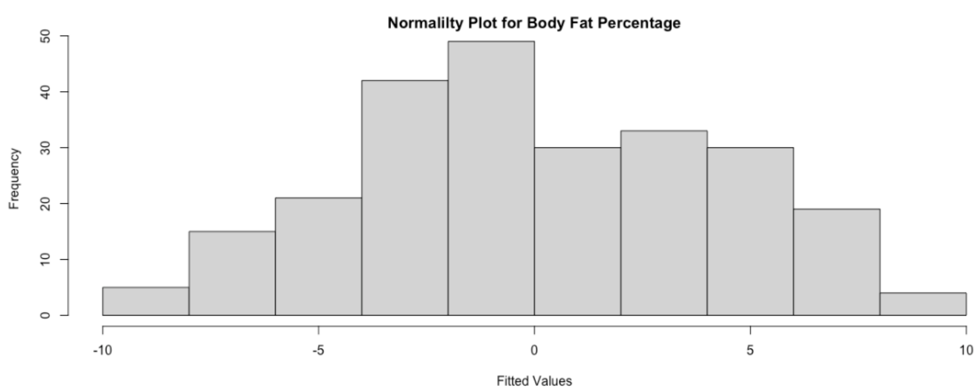
```
ReducFit4 = lm(BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest, data = df3_ob39rem)
summary(ReducFit4)
vif(ReducFit4)
par(mfrow = c(2,1))
hist(ReducFit4$resid, main = "Normalilty Plot for Body Fat Percentage", xlab ="Fitted Values")
ReducFit4_t_i = rstudent(ReducFit4)
qqnorm(ReducFit4_t_i)
qqline(ReducFit4_t_i, col = 2, lwd = 2)
```

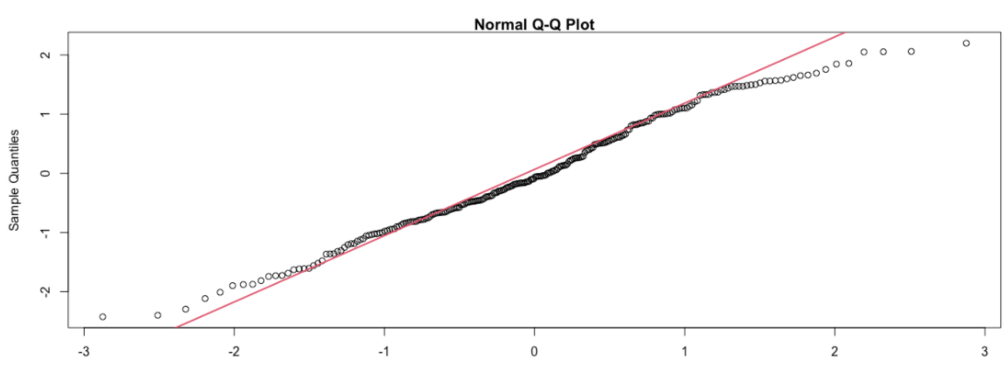
summary(influence.measures(ReducFit4))

Output:

```
Call:
lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest,
    data = df3_ob39rem)

Residuals:
    Min      1Q  Median      3Q     Max
-9.8737 -2.8357 -0.3217  3.4168  9.0188

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.07349    6.15123   0.012  0.99048
Abdomen       0.73728    0.07513   9.813  < 2e-16 ***
Wrist        -2.35973    0.38627  -6.109 4.00e-09 ***
catAgeMiddle  1.91584    0.62646   3.058  0.00248 **
catAgeOlder   4.32447    0.90352   4.786 2.97e-06 ***
BMI           0.62398    0.22972   2.716  0.00708 **
Chest        -0.23347    0.09210  -2.535  0.01188 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 4.189 on 241 degrees of freedom
Multiple R-squared:  0.7426,    Adjusted R-squared:  0.7362
F-statistic: 115.9 on 6 and 241 DF,  p-value: < 2.2e-16

            GVIF Df GVIF^(1/(2*Df))
Abdomen 8.080886  1        2.842690
Wrist   1.728666  1        1.314787
catAge  1.162083  2        1.038268
BMI     8.161827  1        2.856891
Chest   7.734906  1        2.781170


Potentially influential observations of
         lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest,      data =
df3_ob39rem) :

    dfb.1_ dfb.Abdm dfb.Wrst dfb.ctAM dfb.ctAO dfb.BMI dfb.Chst dffit   cov.r
36   0.06   0.00    -0.15     0.03     0.00     0.01    0.07     0.19    1.10_*
40   0.16   0.06    -0.12    -0.08     0.04    -0.19    0.06    -0.37    1.10_*
94  -0.03   0.00     0.02     0.00    -0.01    -0.02    0.02     0.04    1.10_*
200  0.11   0.04    -0.06    -0.10     0.01     0.13   -0.14    -0.27    0.90_*
201 -0.02   0.01     0.10    -0.02     0.00     0.00   -0.07    -0.13    1.11_*
212  0.09  -0.01    -0.11     0.07     0.01     0.20   -0.10     0.30    1.13_*
220 -0.19   0.03     0.14    -0.14    -0.05    -0.02    0.01    -0.30    0.89_*
234 -0.06  -0.03     0.27     0.01    -0.31    -0.04   -0.12    -0.56_*  0.92
236  0.03  -0.05     0.00     0.02     0.08     0.10   -0.04     0.13    1.09_*
237  0.04  -0.02    -0.05     0.00     0.07    -0.01    0.03     0.10    1.09_*
241  0.01   0.02     0.03     0.00    -0.03     0.01   -0.04    -0.06    1.10_*
    cook.d hat
36   0.01   0.08
40   0.02   0.10_*
94   0.00   0.07
200  0.01   0.01
201  0.00   0.08
212  0.01   0.11_*
220  0.01   0.01
234  0.04   0.05
236  0.00   0.06
237  0.00   0.06
241  0.00   0.07
           Test stat Pr(>|Test stat|)
Abdomen      -1.6909          0.09216 .
Wrist         0.1322          0.89491
catAge
BMI          -1.5445          0.12378
Chest        -1.8356          0.06766 .
Tukey test   -1.6750          0.09394 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


Rstudent residuals
Abdomen
Wrist
catAge
Young
Middle
Older
BMI
Chest
Fitted values


Normalilty Plot for Body Fat Percentage
Frequency
Fitted Values


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles

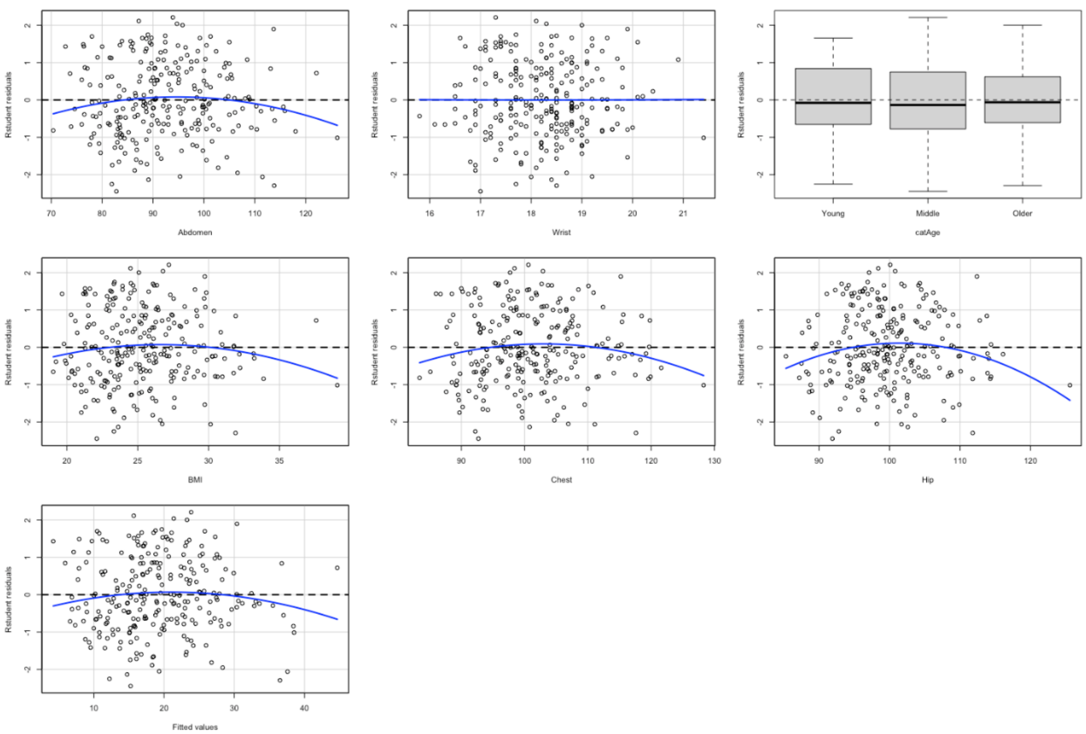
6. 5 Recommended Model 5

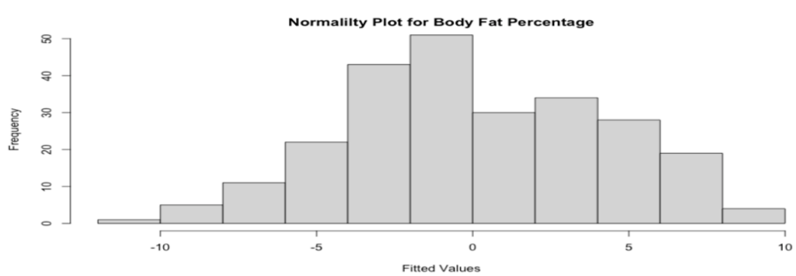
ReducFit5 = lm(BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest + Hip, data = df3_ob39rem)
summary(ReducFit5)
vif(ReducFit5)
residualPlots(ReducFit5, type = "rstudent")
par(mfrow = c(2,1))
hist(ReducFit5$resid, main = "Normalilty Plot for Body Fat Percentage", xlab ="Fitted Values")
ReducFit5_t_i = rstudent(ReducFit5)
qqnorm(ReducFit5_t_i)
qqline(ReducFit5_t_i, col = 2, lwd = 2)
summary(influence.measures(ReducFit5))

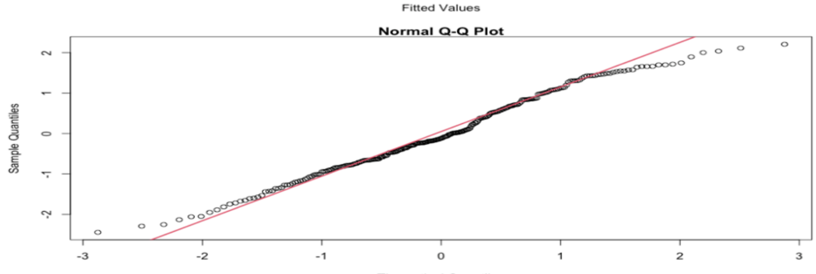
Output:

```
Call:
lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest +
    Hip, data = df3_ob39rem)

Residuals:
    Min      1Q  Median      3Q     Max
-10.036  -2.850  -0.495   3.212   9.040

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.11537    7.03732   0.727 0.468001
Abdomen       0.79231    0.08386   9.448  < 2e-16 ***
Wrist        -2.17676    0.40513  -5.373 1.83e-07 ***
catAgeMiddle  1.61648    0.65759   2.458 0.014672 *
catAgeOlder   3.57465    1.03678   3.448 0.000667 ***
BMI           0.70194    0.23529   2.983 0.003145 **
Chest        -0.23614    0.09191  -2.569 0.010792 *
Hip          -0.14933    0.10202  -1.464 0.144576
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.179 on 240 degrees of freedom
Multiple R-squared:  0.7448,    Adjusted R-squared:  0.7374
F-statistic: 100.1 on 7 and 240 DF,  p-value: < 2.2e-16

             GVIF Df GVIF^(1/(2*Df))
Abdomen 10.114110  1        3.180269
Wrist    1.910558  1        1.382229
catAge   1.540525  2        1.114082
BMI      8.602718  1        2.933039
Chest    7.737962  1        2.781719
Hip      5.968022  1        2.442953
           Test stat Pr(>|Test stat|)
Abdomen      -1.6651          0.09721 .
Wrist         0.0345          0.97252
catAge
BMI          -1.5285          0.12772
Chest        -1.8031          0.07264 .
Hip          -2.5617          0.01103 *
Tukey test   -1.4976          0.13424
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Potentially influential observations of
         lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest +      Hip, data = 
df3_ob39rem) :

    dfb.1_ dfb.Abdm dfb.Wrst dfb.ctAM dfb.ctAO dfb.BMI dfb.Chst dfb.Hip dffit
36   0.01  -0.04    -0.22     0.07     0.06    -0.01    0.09     0.11    0.27
40   0.16   0.08    -0.08    -0.09    -0.01    -0.15    0.05    -0.08   -0.34
57   0.05   0.06    -0.01    -0.03    -0.08    -0.02    0.10    -0.15    0.21
94  -0.05  -0.01     0.03     0.01     0.00    -0.04    0.04     0.02    0.06
169 -0.20  -0.03    -0.07     0.22     0.18    -0.15    0.11     0.19   -0.37
201 -0.04  -0.01     0.10    -0.01     0.02     0.00   -0.08     0.04   -0.17
212  0.10   0.03    -0.07     0.03    -0.03     0.18   -0.08    -0.07    0.26
```

```
220 -0.20   0.00      0.12     -0.11     -0.02     -0.04     0.01      0.07    -0.31
234  0.06   0.08      0.31     -0.06     -0.36      0.02    -0.11     -0.23    -0.58_*
236  0.03  -0.02      0.01      0.01      0.04      0.08    -0.04     -0.03     0.11
241  0.00   0.01      0.03      0.01     -0.03      0.01    -0.06      0.02    -0.09
    cov.r   cook.d hat
36   1.11_*  0.01   0.09
40   1.11_*  0.01   0.10_*
57   1.11_*  0.01   0.08
94   1.11_*  0.00   0.07
169  0.90_*  0.02   0.03
201  1.12_*  0.00   0.08
212  1.15_*  0.01   0.12_*
220  0.86_*  0.01   0.02
234  0.92    0.04   0.06
236  1.10_*  0.00   0.07
241  1.11_*  0.00   0.07
```







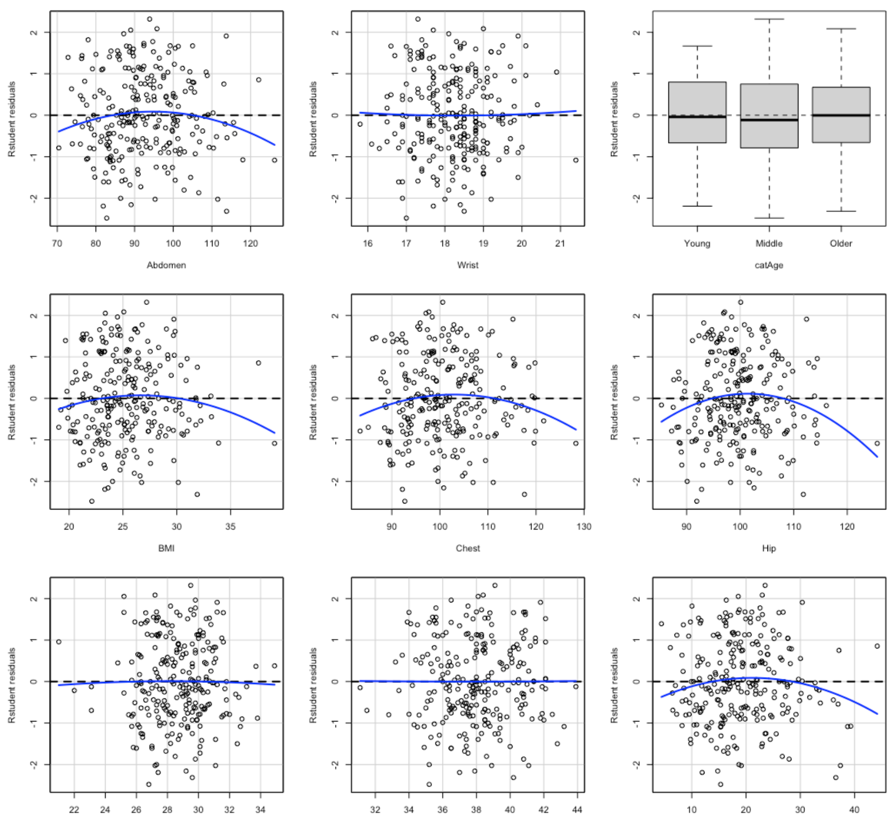
6. 6 Recommended Model 6

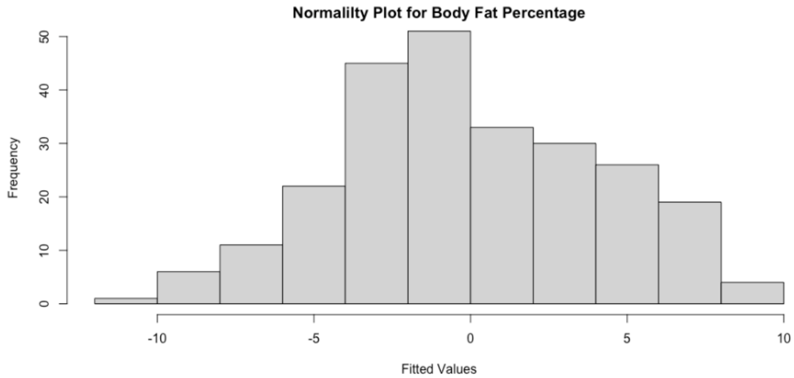
ReducFit6 = lm(BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest + Hip + Forearm + Neck, data = df3_ob39rem)
summary(ReducFit6)
vif(ReducFit6)
summary(influence.measures(ReducFit6))
residualPlots(ReducFit6, type = "rstudent")
par(mfrow = c(2,1))
hist(ReducFit6$resid, main = "Normalilty Plot for Body Fat Percentage", xlab ="Fitted Values")
ReducFit6_t_i = rstudent(ReducFit6)
qqnorm(ReducFit6_t_i)
qqline(ReducFit6_t_i, col = 2, lwd = 2)

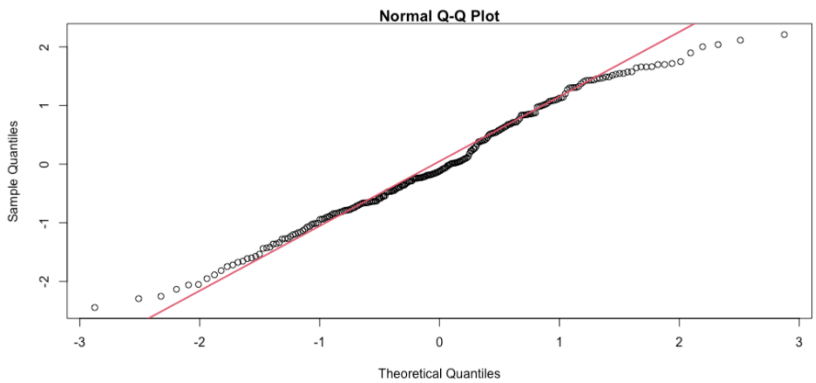
Output:

```
Call:
lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest +
    Hip + Forearm + Neck, data = df3_ob39rem)

Residuals:
     Min       1Q   Median       3Q      Max
-10.1191  -3.0024  -0.3033   3.1304   9.3799

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.51697    7.20152   0.905 0.366410
Abdomen       0.82041    0.08474   9.681  < 2e-16 ***
Wrist        -2.03968    0.47147  -4.326 2.23e-05 ***
catAgeMiddle  1.54615    0.65884   2.347 0.019758 *
catAgeOlder   3.63903    1.05989   3.433 0.000703 ***
BMI           0.68240    0.23898   2.855 0.004678 **
Chest        -0.22441    0.09345  -2.401 0.017104 *
Hip          -0.16812    0.10198  -1.649 0.100538
Forearm       0.29705    0.19145   1.552 0.122083
Neck         -0.36350    0.21782  -1.669 0.096465 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.16 on 238 degrees of freedom
Multiple R-squared:  0.7493,    Adjusted R-squared:  0.7398
F-statistic: 79.03 on 9 and 238 DF,  p-value: < 2.2e-16

             GVIF Df GVIF^(1/(2*Df))
Abdomen 10.423726  1        3.228580
Wrist    2.611379  1        1.615976
catAge   1.638170  2        1.131331
BMI      8.957173  1        2.992854
Chest    8.074422  1        2.841553
Hip      6.017613  1        2.453082
Forearm  2.119270  1        1.455771
Neck     3.491532  1        1.868564
Potentially influential observations of
         lm(formula = BodyFat ~ Abdomen + Wrist + catAge + BMI + Chest +      Hip + Forearm +
Neck, data = df3_ob39rem) :

    dfb.1_ dfb.Abdm dfb.Wrst dfb.ctAM dfb.ctAO dfb.BMI dfb.Chst dfb.Hip dfb.Frrm
40   0.14   0.09    -0.16    -0.08     0.02    -0.19    0.02    -0.08    0.09
104 -0.02  -0.01    -0.04     0.00     0.01    -0.02    0.01     0.01   -0.02
157  0.02   0.03    -0.05    -0.03     0.00    -0.03   -0.01    -0.03    0.15
172 -0.10  -0.09     0.21    -0.12    -0.18     0.06    0.08     0.09   -0.57
201 -0.03  -0.01     0.10    -0.01     0.01     0.00   -0.07     0.04   -0.02
202  0.01   0.01    -0.01    -0.01     0.01    -0.01    0.00     0.00    0.05
203  0.15  -0.08    -0.30     0.19     0.11     0.20   -0.14     0.01    0.07
212  0.13   0.02    -0.04     0.03    -0.05     0.23   -0.08    -0.08   -0.09
220 -0.20   0.00     0.09    -0.11    -0.01    -0.04    0.00     0.07    0.00
    dfb.Neck dffit   cov.r   cook.d hat
40   0.12    -0.41    1.13_*  0.02   0.12_*
104  0.06    -0.06    1.24_*  0.00   0.16_*
157 -0.02     0.17    1.24_*  0.00   0.16_*
```

```
172  0.09     0.62_*  1.42_*  0.04   0.29_*
201 -0.02    -0.17    1.13_*  0.00   0.09
202 -0.02    -0.05    1.24_*  0.00   0.16_*
203  0.20     0.45    0.86_*  0.02   0.04
212  0.00     0.32    1.16_*  0.01   0.13_*
220  0.04    -0.31    0.82_*  0.01   0.02
           Test stat Pr(>|Test stat|)
Abdomen      -1.7968         0.073647 .
Wrist         0.2133         0.831253
catAge
BMI          -1.5784         0.115817
Chest        -1.8245         0.069342 .
Hip          -2.6090         0.009658 **
Forearm      -0.2681         0.788827
Neck          0.0465         0.962933
Tukey test   -1.8876         0.059081 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```




Normalilty Plot for Body Fat Percentage
Frequency
0
10
20
30
40
50
-10
-5
0
5
10
Fitted Values


Normal Q-Q Plot
Sample Quantiles
-2
-1
0
1
2
-3
-2
-1
0
1
2
3
Theoretical Quantiles

Appendix 7: Criterions

7. 1 Full Model: AIC, BIC, C_p, PRESS

AIC(fit0, k = 2)
BIC(fit0)
ols_mallows_cp(fit0, fit0)
PRESS(fit0)

Output:

```
[1] 1444.494
[1] 1497.256
[1] 22.43372
[1] 4984.599
```

7. 2 Full Model (without observation 39): AIC, BIC, C_p, PRESS

AIC(alt_fit0, k = 2)
BIC(alt_fit0)
ols_mallows_cp(alt_fit0, alt_fit0)
PRESS(alt_fit0)

Output:

```
[1] 1444.494
[1] 1497.256
[1] 22.43372
[1] 4984.599
```

7. 3 Recommended Model 1: AIC, BIC, C_p, PRESS

AIC(ReducFit, k = 2)
BIC(ReducFit)
ols_mallows_cp(ReducFit, fit0)
PRESS(ReducFit)

Output:

```
[1] 1439.219
[1] 1477.911
[1] 6.585873
[1] 4807.058
```

7. 4 Recommended Model 2: AIC, BIC, C_p, PRESS

AIC(ReducFit2, k = 2)
BIC(ReducFit2)
ols_mallows_cp(ReducFit2, fit0)
PRESS(ReducFit2)

Output:

```
[1] 1438.38
[1] 1473.555
[1] 5.696308
[1] 4740.067
```

7. 5 Recommended Model 3: AIC, BIC, C_p, PRESS

```
AIC(ReducFit3, k = 2)
BIC(ReducFit3)
ols_mallows_cp(ReducFit3, alt_fit0)
PRESS(ReducFit3)
```

Output:

```
[1] 1423.09
[1] 1465.251
[1] 7.049047
[1] 4465.648
```

7. 6 Recommended Model 4: AIC, BIC, C_p, PRESS

```
AIC(ReducFit4, k = 2)
BIC(ReducFit4)
ols_mallows_cp(ReducFit4, alt_fit0)
PRESS(ReducFit4)
```

Output:

```
[1] 1423.151
[1] 1451.258
[1] 6.814494
[1] 4469.812
```

7. 7 Recommended Model 5: AIC, BIC, C_p, PRESS

```
AIC(ReducFit5, k = 2)
BIC(ReducFit5)
ols_mallows_cp(ReducFit5, alt_fit0)
PRESS(ReducFit5)
```

Output:

```
[1] 1422.947
[1] 1454.567
[1] 6.666045
[1] 4471.351
```

7. 8 Recommended Model 6: AIC, BIC, C_p, PRESS

```
AIC(ReducFit6, k = 2)
BIC(ReducFit6)
ols_mallows_cp(ReducFit6, alt_fit0)
PRESS(ReducFit6)
```

Output:

```
[1] 1422.59
[1] 1461.237
[1] 6.474756
[1] 4456.317
```