STATISTICAL REVIEW OF FACTORS THAT IMPACT NBA SCORED POINTS

Bortiorkor Nii Tsui Alabi

Anirudh Giridhar

Won Hyun Kim

Andrew Lang

Ernest Lee

Andrew Lim

STAT 512: Applied Linear Regression Analysis

Section 03: Team 06 Project Final Report

Introduction

Basketball is one of the prominent sports in the world today and some of the world's greatest athletes are basketball players (Akers, Wolff & Buttross, 1992). Since its invention in 1891 by Jim Naismith, basketball has progressed from being an indoor fun game, an official collegiate sport, a national professional sport, and finally an Olympic sport in 1936 (Oliver, 2005). The National Basketball Association (NBA) as the main men's professional basketball league organization in North America ensures that the sport of basketball is competitive and upheld to the highest performance standard while rightfully observing all protocols, ethics and fair dealings (Hofler & Payne, 1997). Consequently, the NBA as a corporation has fostered several systematic tactical and strategic measures to prevent money-losing/devaluing of franchises, low game patronage, television rating decline, and overall encouraged favorable national appeal towards basketball (Csataljay, O'Donoghue, Hughes & Dancs, 2009). These tactical strategies have included quantitative and scientific measures for interpretation of key game outputs such as scorings and free throws earned. These factors are required to compute sound logical and analytical metrices like the player efficiency ratings (PER) to explore gaming opportunities and monitor performance across games and individual players (Sampaio, & Janeira, 2003).

Consequently, exploring efficiency to gaming opportunities and player performance implies that the more teams win, the higher the game event patronage (more fans of teams pay to attend games, buy jerseys, and other paraphernalia) and associated increase in allied financial remunerations. Therefore, it is important to players, coaches, and owners to develop winning strategies.

Teams win by scorning more points than the opposing team. The point system works as follows: three points are awarded for shots made from behind the three-point line, two points are awarded for shots made inside the three-point line, and one point is awarded for free throws. However, basketball is more than just point scoring; coaches need to know what teamwork strategies lead to more points. Fortunately, various metrics regarding team performance are recorded for every NBA game (Mikołajec, Maszczyk & Zając, 2013).

We propose a statistical analysis on nba.games.stats.csv which can be found from https://www.kaggle.com/ionaskel/nba-games-stats-from-2014-to-2018 (Kelepouris, 2018). We would like to find out which team performance metrics are the most important considerations for coaches trying to develop a winning strategy. Due to the substantial amount of entries in the data set, we choose to analyze the data presented for the Indiana Pacers. Though analysis we present a linear model that represents the best combination of metrics to predict points scored.

Methods

Data Description

The aim of this project is to investigate the factors that are significantly vital in impacting the amount of points obtained by NBA teams. Specifically, we are interested in how the amount of points scored by an NBA team is influenced by the following seven variables categorized below. By researching and analyzing these factors, we hope to understand which traits NBA scouts should be looking for to increase their points scored for the next game season.

The following table provides the description, range and type of variables we used:

Variable	Description	Range	Variable Type
X3PointShots	Percentage of 3-point shots made	(0, 0.75)	Continuous
FieldGoals	Percentage of field goals made	(19, 58)	Continuous
Assists	Number of assists	(6, 47)	Discrete
Rebounds	Number of rebounds	(19, 58)	Discrete
Turnovers	Number of turnovers	(2, 29)	Discrete
TotalFouls	Number of fouls	(7, 42)	Discrete
Steals	Number of steals	(0, 21)	Discrete
TeamPoints / Point differential	Total number of points scored	(64, 149)	Discrete

Yellow: Predictor Variable

Green: Response Variable

Preliminary Analyses

The first step we took in analyzing our data was to better understand which predictor variables had significant linear impact on our response variable. This could be achieved through data visualization: generating boxplots, scatterplots and histograms for each of our predictors (Appendix A). Our scatter plots indicate a linear relationship between Team Points and variables Field Goals, Assists, and 3-Point Shots. It does not show any signs of multicollinearity as there are no patterns to the intersecting variables. However, it must be addressed that our data has a large sample size (n=328), which could potentially hinder the observation of a linear pattern in our scatterplots.

As for our boxplots, we can see that the distributions of our variables are not heavily skewed to either side, which indicates normality. The mean of the variables is also relatively close to the median, with the presence of only a few outliers. The only exception to this would be the variable FieldGoals as it has a large range, creating several outliers.

Finally, our histograms help visualize the distribution of each variable and we can notice the general pattern of unimodality with a normal-shaped distribution. The one variable that does not quite follow this trend is Fouls, which has a relatively right skewed distribution with an outlier on the far end.

Model Selection

Full model $\mathbf{TeamPoints} \sim \mathbf{X3PointShots.} + \mathbf{FieldGoals.} + \mathbf{Assists} + \mathbf{TotalRebounds} + \mathbf{Turnovers} + \mathbf{TotalFouls} + \mathbf{Steals}$

After conducting a Type I ANOVA and Type II ANOVA test on our data set, we can see from the summary output (Appendix A) that the p-values for all of the predictor variables are significantly lower than 0.05. Thus, we can conclude that all the predictors have a significant linear impact on the response variable (Team Points). Therefore, we do not need to drop any insignificant variable and can keep our full model as it is with all the predictors (X3PointShots., FieldGoals., Assists , TotalRebounds, Turnovers, TotalFouls, Steals). See below for summary of the full model along with an ANOVA table.

```
Residuals:
     Min       1Q   Median       3Q      Max 
-14.6775  -4.2318  -0.5025   3.4969  23.7520 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)     7.84553    5.29149   1.483  0.13915    
X3PointShots.  20.59708    3.70982   5.552 5.94e-08 ***
FieldGoals.   135.23332    8.96011  15.093  < 2e-16 ***
Assists         0.24897    0.09240   2.694  0.00742 ** 
TotalRebounds   0.35405    0.06303   5.618 4.21e-08 ***
Turnovers      -0.59684    0.10456  -5.708 2.60e-08 ***
TotalFouls      0.37106    0.08293   4.474 1.07e-05 ***
Steals          0.65348    0.11391   5.737 2.24e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.313 on 320 degrees of freedom
Multiple R-squared:  0.6863,	Adjusted R-squared:  0.6794 
F-statistic:   100 on 7 and 320 DF,  p-value: < 2.2e-16
```

A anova: 8 × 5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
X3PointShots.	1	9632.7414	9632.74142	241.67614	5.452860e-41
FieldGoals.	1	14064.0145	14064.01453	352.85249	1.395704e-53
Assists	1	699.2289	699.22893	17.54298	3.636157e-05
TotalRebounds	1	729.3381	729.33808	18.29839	2.497178e-05
Turnovers	1	632.3312	632.33118	15.86458	8.427994e-05
TotalFouls	1	831.6185	831.61854	20.86450	7.044535e-06
Steals	1	1311.7100	1311.70998	32.90953	2.237686e-08
Residuals	320	12754.5783	39.85806	NA	NA

Figure 1 Model Summary and Type I ANOVA Results

```
Anova Table (Type II tests)

Response: TeamPoints
                Sum Sq  Df  F value    Pr(>F)    
X3PointShots.   1228.6   1  30.8251 5.937e-08 ***
FieldGoals.     9079.4   1 227.7931 < 2.2e-16 ***
Assists          289.4   1   7.2596  0.007424 ** 
TotalRebounds   1257.8   1  31.5566 4.212e-08 ***
Turnovers       1298.8   1  32.5854 2.603e-08 ***
TotalFouls       798.0   1  20.0199 1.067e-05 ***
Steals          1311.7   1  32.9095 2.238e-08 ***
Residuals      12754.6 320                       
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 2 Full Model Type II ANOVA Results

Continuing with our model selection, we ran a BestSub analysis to identify predictors that could be dropped, while simultaneously supporting our ANOVA results. The criteria for our model selection are listed in the BestSub table output (Appendix A). From the output, we can see that Model 7 (with all parameters) has the best values for every criterion. Therefore, we do not drop any predictors and keep our full model as it is, as supported by the ANOVA output.

Diagnostics

As for the MLM diagnostics of our data, we ran several procedures to see if our model satisfies the assumption of normal error distribution, linearity and constant error variance (Appendix A).

Starting with the linearity assumption, we formulated a Pearson Residual plot, which indicated that our data had a linear pattern for each predictor apart from the variables FieldGoals and Fitted Values which have a slightly upward curving best fit line. Therefore, the Pearson Residual plot confirms that the linearity assumption was met.

To test the normality, we ran a Shapiro-Wilk normality test on our full model and concluded that our data is not normal. Therefore, we did a boxcox transformation and got a new transformed response variable (TeamPointsnew). The next two figures show the results of Type I and Type II ANOVA for the new model.

```
Analysis of Variance Table

Response: TeamPointsnew
                Df  Sum Sq Mean Sq F value    Pr(>F)    
X3PointShots.    1 0.92888 0.92888 241.121 < 2.2e-16 ***
FieldGoals.      1 1.35050 1.35050 350.567 < 2.2e-16 ***
Assists          1 0.06073 0.06073  15.765 8.859e-05 ***
TotalRebounds    1 0.06314 0.06314  16.390 6.471e-05 ***
Turnovers        1 0.06953 0.06953  18.049 2.826e-05 ***
TotalFouls       1 0.07687 0.07687  19.955 1.101e-05 ***
Steals           1 0.12954 0.12954  33.627 1.602e-08 ***
Residuals      320 1.23274 0.00385                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 3 ANOVA Results for Model with Transformed Response Variable

```
Anova Table (Type II tests)

Response: TeamPointsnew
                Sum Sq  Df  F value    Pr(>F)    
X3PointShots. 0.12160   1  31.5650 4.196e-08 ***
FieldGoals.   0.87442   1 226.9857 < 2.2e-16 ***
Assists       0.02434   1   6.3191   0.01244 *  
TotalRebounds 0.11301   1  29.3351 1.198e-07 ***
Turnovers     0.13720   1  35.6143 6.377e-09 ***
TotalFouls    0.07366   1  19.1207 1.662e-05 ***
Steals        0.12954   1  33.6267 1.602e-08 ***
Residuals     1.23274 320                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 4 Type II ANOVA Results for Model with Transformed Response Variable

After conducting a Type I ANOVA and Type II ANOVA test on our new model, we can see from the summary output (Appendix A) that the p-values for all of the predictor variables are significantly lower than 0.05. Thus, we can conclude that all the predictors have a significant linear impact on the response variable (ln(Team Points)). Therefore, we do not need to drop any insignificant variable and can keep our full model as it is with all the predictors (X3PointShots., FieldGoals., Assists , TotalRebounds, Turnovers, TotalFouls, Steals).

Lastly, we conducted several Brown-Forsythe Tests on each of our variables against our new response variable. From the output (Appendix A), we can see that the p-values for all the outputs are significantly higher than the alpha, indicating the constant variance assumption was met. This can also be supported by our observation of the normality QQ-plot where the fitted values do not increase in variance from left to right.

Outliers/Cook's Distance

To test for any extreme outliers that could be detrimental to our regression model, we ran an Influence plot highlighting studentized residuals vs. Hat-values. From our graph, we can see that there were 2 data values that were X outliers (Cases 8323 & 3365). There were no Y outliers present. Although these 2 cases were outliers, our Cook's Distance (Appendix A) values showed that they were not high enough to warrant them being dropped from our data set. Therefore, there were no cases dropped from our data set due to outliers.

Variance Inflation Factor (VIF) on Final Model

Once our final model met the necessary underlying assumptions, we ran VIFs to evaluate whether our predictors had any multicollinearity issues. Looking at our output (Appendix A), the maximum VIFs did not exceed 3, except for just one. Our results showed that there was not much of a collinearity issue between our predictors.

Ridge Regression

Although our VIF values were relatively small, we ran Ridge Regression tests to further reduce the multicollinearity of our model. From our plots and output, the best value of K to use is 0.04. At this value, we can see that the VIF values for every predictor is approximately close to 1.

Cross Validation

Since we do not need to drop any predictors from our model, we ran a Cross-Validation test on the final model. We obtain a rather small standard deviation of Rsquared and RMSE. We can also see from the output that our Rsquared adjusted is around 0.68, the same as our summary results, which further prove that we do not need to drop any predictors.

Results

The final model is shown below:

$$\begin{aligned}\textbf{TeamPointsnew} \\ &= \mathbf{3.707 + 0.204X3PointShots + 1.237FieldGoals + 0.002Assists} \\ &\quad \mathbf{+ 0.003TotalRebounds - 0.006Turnovers + 0.003TotalFouls + 0.006Steals}\end{aligned}$$

Where $\text{TeamPointsnew} = \ln(\text{Teampoints})$

To make a prediction for team points based on this model, we find the confidence interval for the prediction of the TeamPointsnew variable and complete a back-transformation shown below.

$$Y' = \text{team points new}$$

$$\text{CI for } Y' = (\widehat{Y}'_{lower}, \widehat{Y}'_{upper})$$

$$\text{CI for } Y = (e^{\widehat{Y}_{lower}}, e^{\widehat{Y}_{upper}})$$

The final model shows that FieldGoals has the most influence on the number of points earned by a team. The second most influential variable is X3PointShots. Thus, teams with a high number of Field goals and three-point shots will score the most points. In addition, teams should avoid turnovers whenever possible, as the model shows that turnovers lead to a slight decrease in the number of points scored by a team. In general, basketball teams should employ strategies that maximize the amount of shots taken and minimize their rate of turnovers to the opposing team. The model makes intuitive sense; more shots leads to higher points while surrendering control of the ball to the opposing team leads to lower points. In addition, analysis of the final model shows promising results. Table 1 states that the final model has $R^2_{adj} = 0.678$ and the highest $p_{value} =$

0.012 (for the Assists parameter). For a comprehensive view of the parameter estimates and confidence intervals see Table 2.

Table 1 ANOVA for Final Model

```
Call:
lm(formula = TeamPointsnew ~ X3PointShots. + FieldGoals. + Assists +
    TotalRebounds + Turnovers + TotalFouls + Steals, data = bball)

Coefficients:
  (Intercept)  X3PointShots.    FieldGoals.        Assists  TotalRebounds
     3.707243       0.204908       1.327138       0.002284       0.003356
    Turnovers     TotalFouls         Steals
    -0.006134       0.003565       0.006494
```

A anova: 8 × 4

	Sum Sq	Df	F value	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>
X3PointShots.	0.12159855	1	31.564981	4.195510e-08
FieldGoals.	0.87442250	1	226.985670	3.854588e-39
Assists	0.02434320	1	6.319094	1.243506e-02
TotalRebounds	0.11300833	1	29.335100	1.198261e-07
Turnovers	0.13719797	1	35.614333	6.377391e-09
TotalFouls	0.07365933	1	19.120749	1.661645e-05
Steals	0.12954106	1	33.626725	1.602261e-08
Residuals	1.23274390	320	NA	NA

Table 2 Summary of Final Model

```
Analysis of Variance Table

Response: TeamPointsnew
               Df  Sum Sq Mean Sq F value    Pr(>F)
X3PointShots.   1 0.92888 0.92888 241.121 < 2.2e-16 ***
FieldGoals.     1 1.35050 1.35050 350.567 < 2.2e-16 ***
Assists         1 0.06073 0.06073  15.765 8.859e-05 ***
TotalRebounds   1 0.06314 0.06314  16.390 6.471e-05 ***
Turnovers       1 0.06953 0.06953  18.049 2.826e-05 ***
TotalFouls      1 0.07687 0.07687  19.955 1.101e-05 ***
Steals          1 0.12954 0.12954  33.627 1.602e-08 ***
Residuals     320 1.23274 0.00385
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Discussion

The goal of this research was to investigate which factors significantly contribute to the amount of points scored by NBA teams. Specifically, the research concentrated on finding the factors that contributed to the amount of points scored by the Indiana Pacers over many seasons. The number of factors considered was 7; Percentage of 3-point shots, percentage of field goals made, number of assists, number of rebounds, number of turnovers, number of fouls and number of steals. Diagnostics showed that all the variables were linearly related to the response variable (team points), and upon analyzing their individual contribution, it was revealed that all the factors have a significant linear impact on the response. The response was transformed to a logarithmic function using the boxcox transformation method as the untransformed Y did not pass a normality test. Several other diagnostics were also run, and analysis was performed on the model.

The model shows great results. Analysis showed that percentage of field goals has the most influence on the number of points, followed by percentage of 3-point shots. The model therefore gives evidence for the importance of field goals and three-point shots to win games. Furthermore, number of turnovers is the only inversely related variable, which is expected as teams should avoid turnovers to retain control over the ball in a game. This research should be helpful to the NBA scouts looking to scout prospective players for their respective teams – they can look for offense players who score a lot of points and defense players who increase turnovers for the enemy team by stealing the ball and maintaining ball control. However, some limitations of the model exist – the effectiveness of the model is subject to research on a single NBA team, and so more research should be done on all NBA teams to come up with a global statistic.

The conclusions presented by the research can be supported by the prioritization of players like the ones suggested above, who can maintain a defensive zone through a point guard, while simultaneously controlling the flow of the game by giving more opportunities to the top scorers.

References

Akers, M. D., Wolff, S., & Buttross, T. E. (1992). An empirical examination of the factors affecting the success of NCAA Division I College Basketball teams. *Journal of Business and Economic Studies*.

Csataljay, G., O’Donoghue, P., Hughes, M., & Dancs, H. (2009). Performance indicators that distinguish winning and losing teams in basketball. International *Journal of Performance Analysis in Spor*t, 9(1), 60-66.

Oliver, D. (2005). What wins basketball games, a review of „Basketball on paper: Rules and tools for performance analysis”. Polomac Books, 26-85.

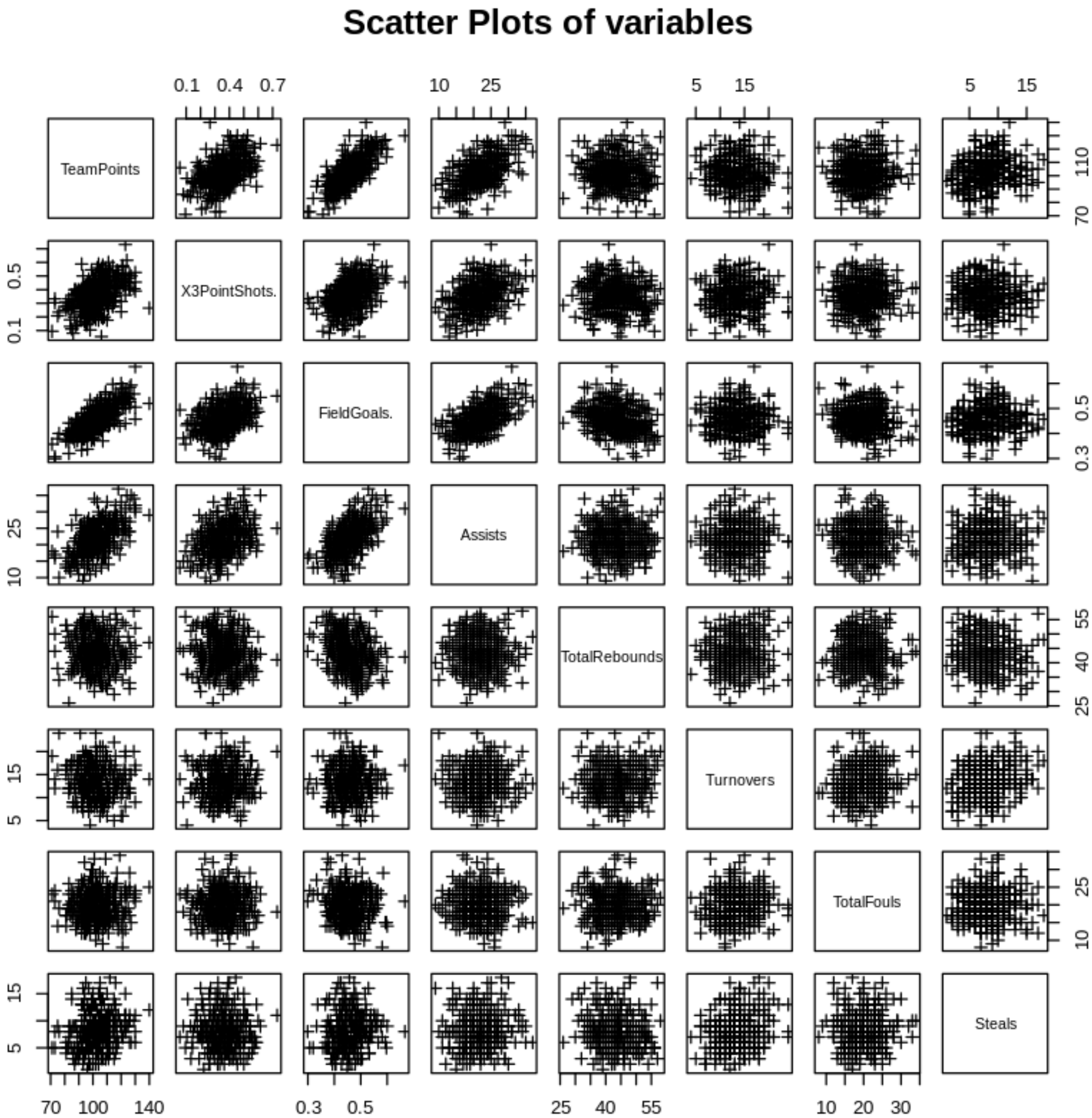
Sampaio, J., & Janeira, M. (2003). Statistical analyses of basketball team performance: understanding teams’ wins and losses according to a different index of ball possessions. *International Journal of Performance Analysis in Sport*, 3(1), 40-49.

Hofler, R. A., & Payne, J. E. (1997). Measuring efficiency in the national basketball association. *Economics letters*, 55(2), 293-299.

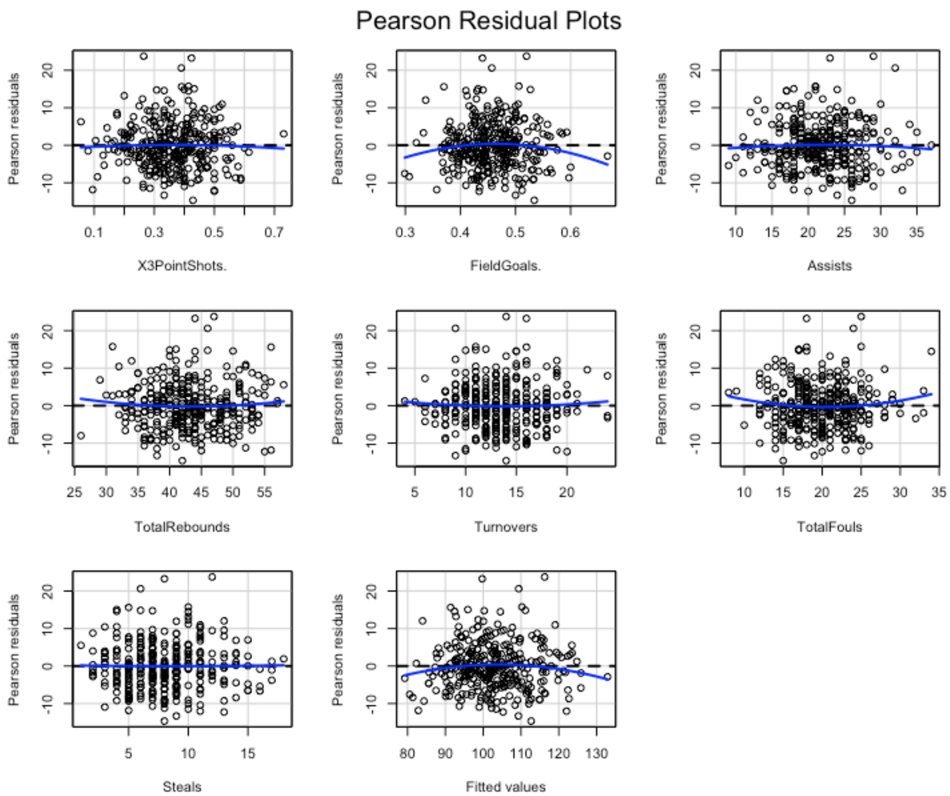
Kelepouris, I. (2018). NBA Team Game Stats from 2014 to 2018. Retrieved from https://www.kaggle.com/ionaskel/nba-games-stats-from-2014-to-2018

Mikołajec, K., Maszczyk, A., & Zając, T. (2013). Game Indicators Determining Sports Performance in the NBA. Journal of human kinetics, 37, 145–151

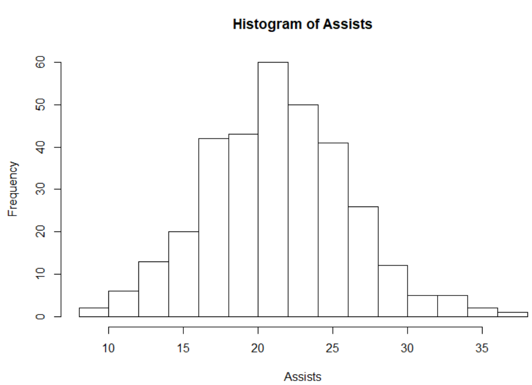
Appendix A

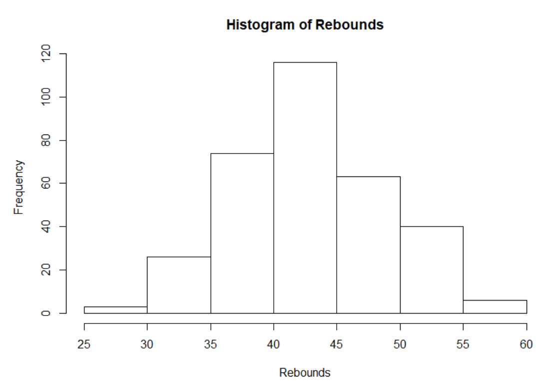

Scatter Plots of variables
TeamPoints
X3PointShots.
FieldGoals.
Assists
TotalRebounds
Turnovers
TotalFouls
Steals
0.1 0.4 0.7
10 25
5 15
5 15
110
70
0.5
0.1
0.5
0.3
25
10
55
40
25
15
5
25
10
15
5
70 100 140
0.3 0.5
25 40 55
10 20 30

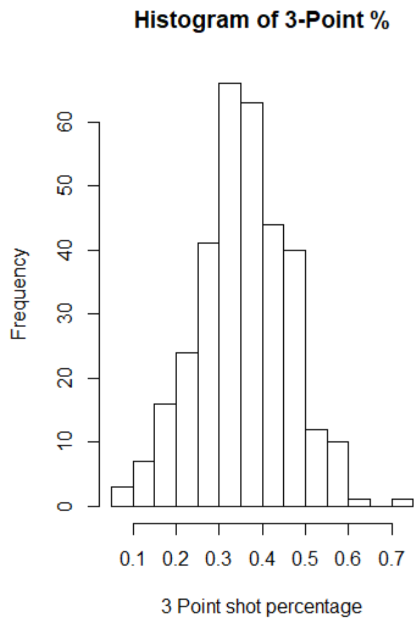
Residual Plots

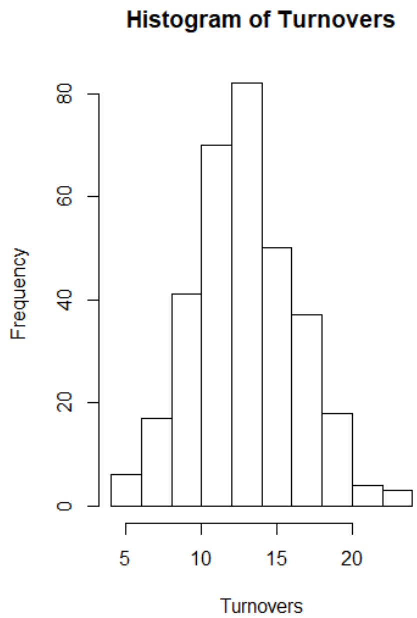


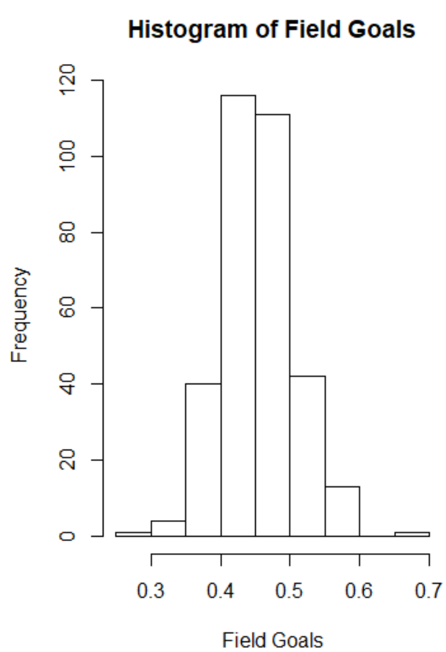
Histograms

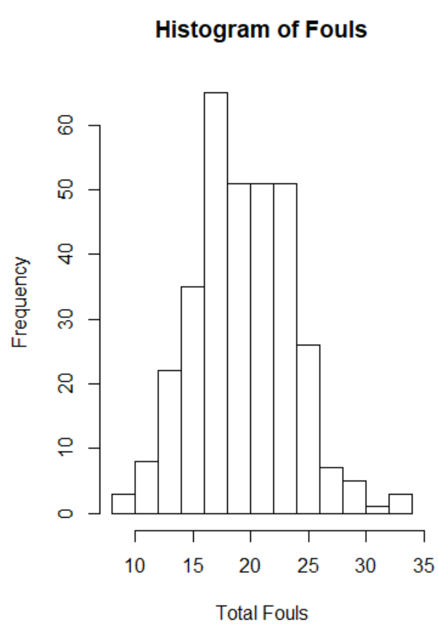


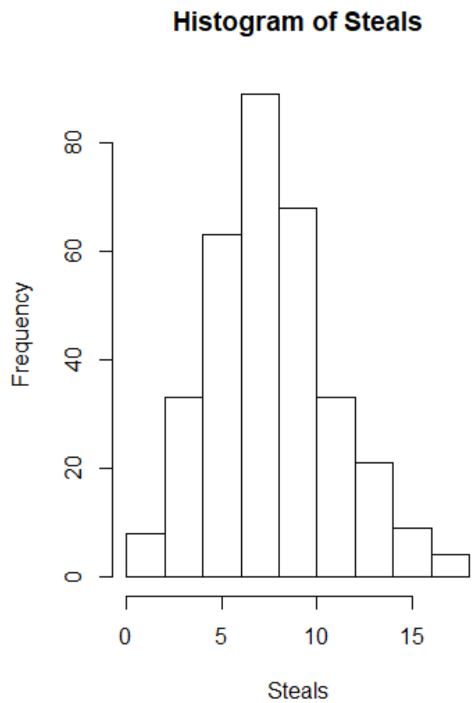



Histogram of 3-Point %
Frequency
0
10
20
30
40
50
60
0.1
0.2
0.3
0.4
0.5
0.6
0.7
3 Point shot percentage

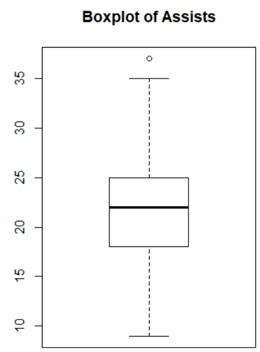

Histogram of Turnovers
Frequency
0
20
40
60
80
5
10
15
20
Turnovers

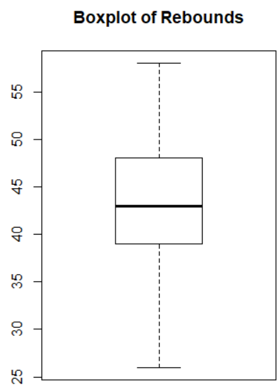

Histogram of Field Goals
Frequency
0
20
40
60
80
100
120
0.3
0.4
0.5
0.6
0.7
Field Goals

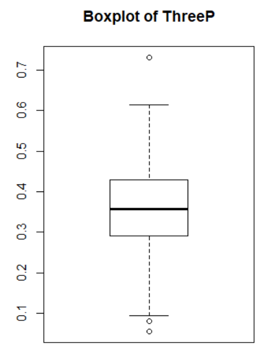

Histogram of Fouls
Frequency
0
10
20
30
40
50
60
10
15
20
25
30
35
Total Fouls

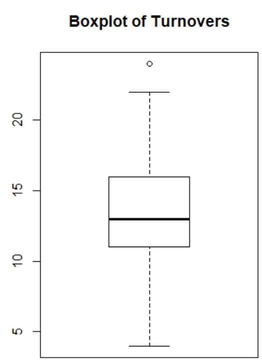

Histogram of Steals
Frequency
0
20
40
60
80
0
5
10
15
Steals

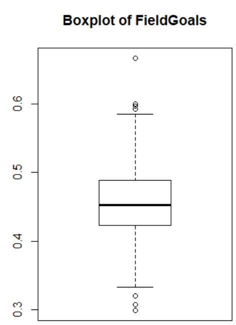
BoxPlots (Pre-Transform)

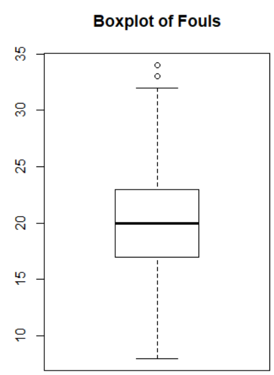


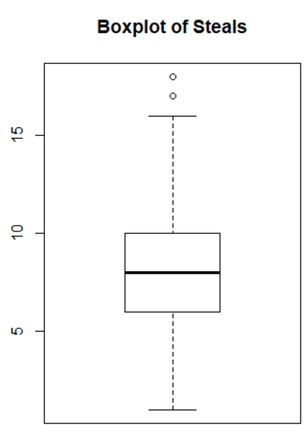


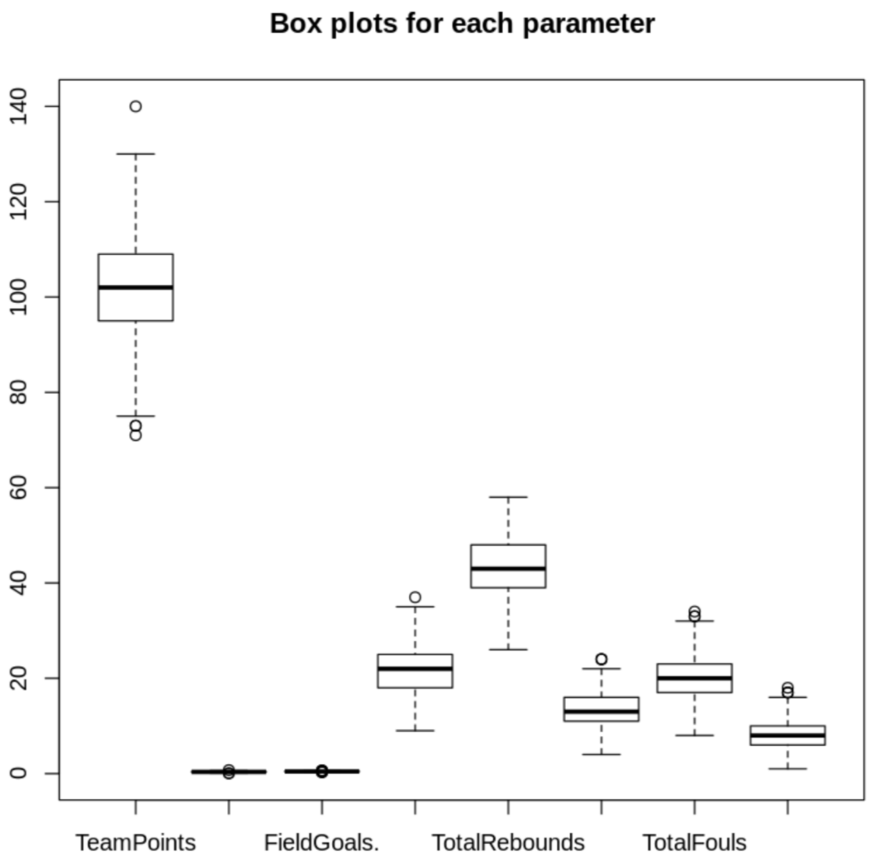






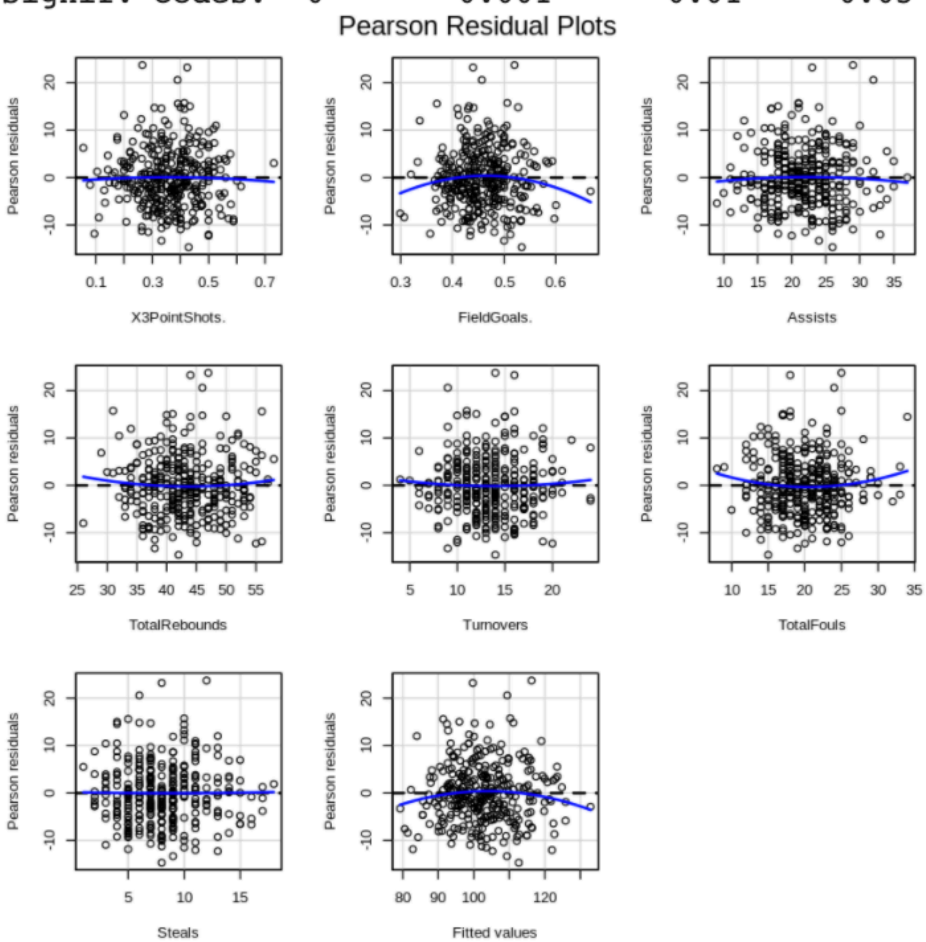





Box plots for each parameter
0
20
40
60
80
100
120
140
TeamPoints
FieldGoals.
TotalRebounds
TotalFouls

Residual Plots Pre-transform

```
               Test stat Pr(>|Test stat|)
X3PointShots.    -0.3360          0.73710
FieldGoals.      -1.7380          0.08318 .
Assists          -0.5295          0.59683
TotalRebounds     0.8612          0.38977
Turnovers         0.6049          0.54566
TotalFouls        1.4240          0.15542
Steals            0.1139          0.90941
Tukey test       -1.6366          0.10171
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Brown Forsythe Test Results (Pre-Transform)

```
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : resid and AssistsGroup

 statistic  : 0.2259788
 num df     : 4
 denom df   : 61.05583
 p.value    : 0.9228547

 Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

```
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : resid and FieldGoals.Group

 statistic  : 0.8707002
 num df     : 4
 denom df   : 61.81626
 p.value    : 0.4866983

 Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

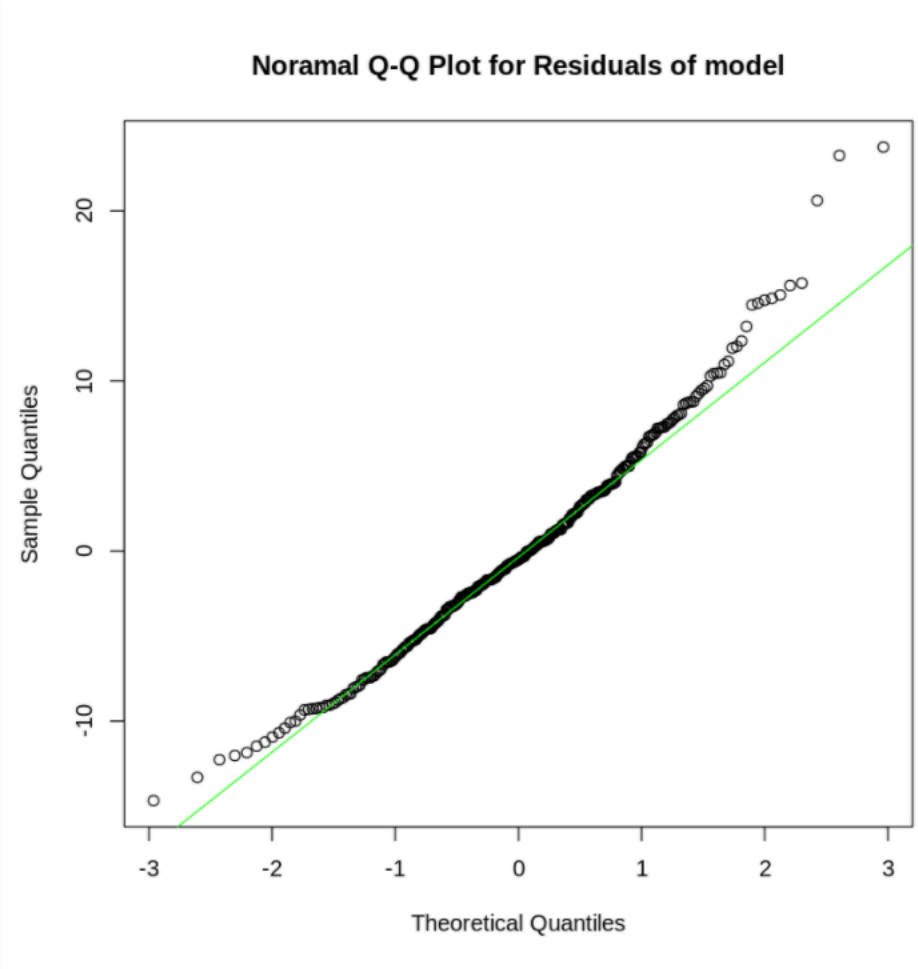
```
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : resid and TotalFoulsGroup

 statistic  : 1.404006
 num df     : 4
 denom df   : 114.3023
 p.value    : 0.2371068

 Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

```
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : resid and TotalReboundsGroup

 statistic  : 1.221902
 num df     : 4
 denom df   : 62.26163
 p.value    : 0.3106521

 Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

```
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : resid and TurnoversGroup

 statistic  : 0.9993378
 num df     : 4
 denom df   : 98.28106
 p.value    : 0.4117563

 Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

```
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : resid and StealsGroup

 statistic  : 0.9472937
 num df     : 4
 denom df   : 157.2211
 p.value    : 0.4383236

 Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

QQNorm Plots for residual (Pre transform)



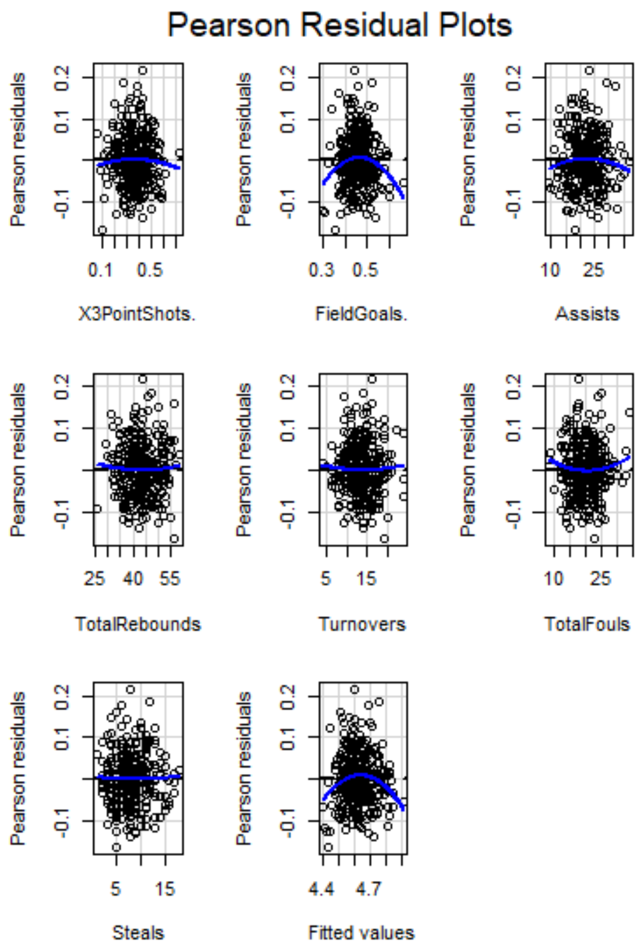
Shapiro Test Results (Pre-transform)

```
        Shapiro-Wilk normality test

data:  bball.fullmod$residual
W = 0.97969, p-value = 0.0001361
```

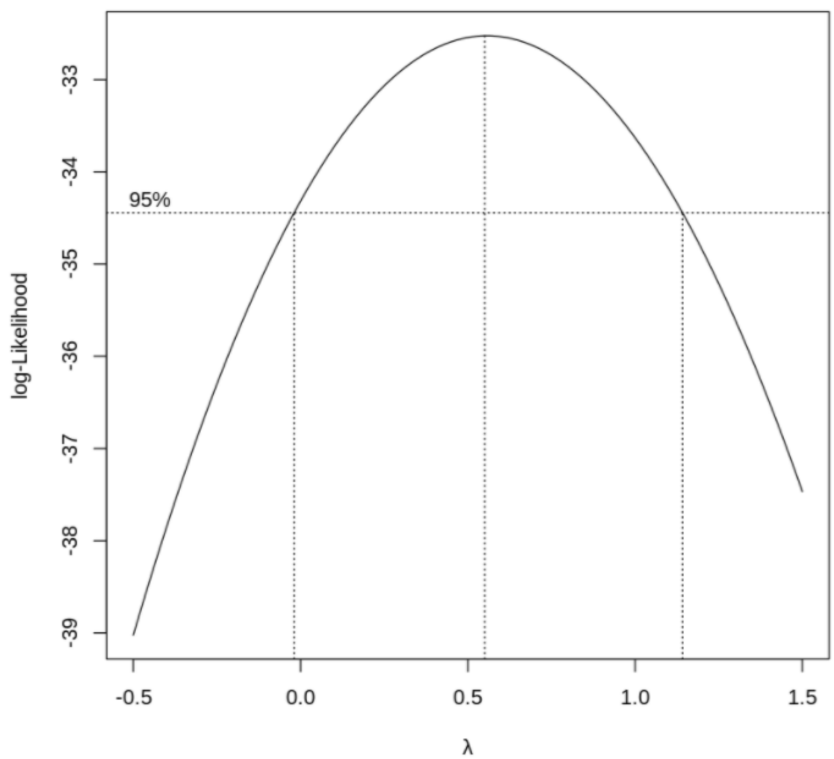
Shapiro Test (Post-transform)

```
        Shapiro-Wilk normality test

data:  residuals(bball.newmod)
W = 0.99178, p-value = 0.06593
```

Person Residua Plots (Post Transform)



Boxcox Transformation

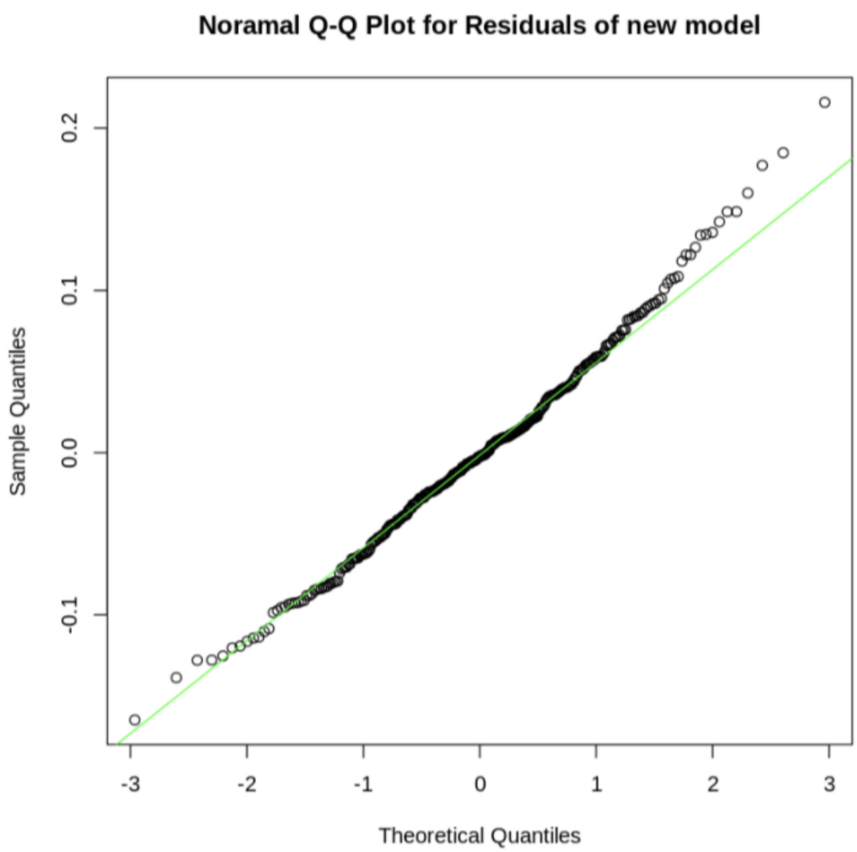


Brown Forsythe Test Results (Post-Transform)

```
  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and AssistsGroup

  statistic  : 0.4727327
  num df     : 4
  denom df   : 75.41799
  p.value    : 0.7555842

  Result     : Difference is not statistically significant.
-------------------------------------------------------------


  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and X3PointShots.Group

  statistic  : 0.1685243
  num df     : 4
  denom df   : 119.2134
  p.value    : 0.9540368

  Result     : Difference is not statistically significant.
-------------------------------------------------------------


  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and FieldGoals.Group

  statistic  : 1.772825
  num df     : 4
  denom df   : 41.04364
  p.value    : 0.1528177

  Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

```
  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and TotalFoulsGroup

  statistic  : 1.544203
  num df     : 4
  denom df   : 126.5893
  p.value    : 0.1933904

  Result     : Difference is not statistically significant.
-------------------------------------------------------------


  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and TotalReboundsGroup

  statistic  : 1.162524
  num df     : 4
  denom df   : 65.59652
  p.value    : 0.3354623

  Result     : Difference is not statistically significant.
-------------------------------------------------------------


  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and TurnoversGroup

  statistic  : 0.8779067
  num df     : 4
  denom df   : 72.88746
  p.value    : 0.4815052

  Result     : Difference is not statistically significant.
-------------------------------------------------------------


  Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
  data : residual and StealsGroup

  statistic  : 1.179605
  num df     : 4
  denom df   : 170.2343
  p.value    : 0.3215753

  Result     : Difference is not statistically significant.
-------------------------------------------------------------
```

QQNorm Plots for residual (Post transform)

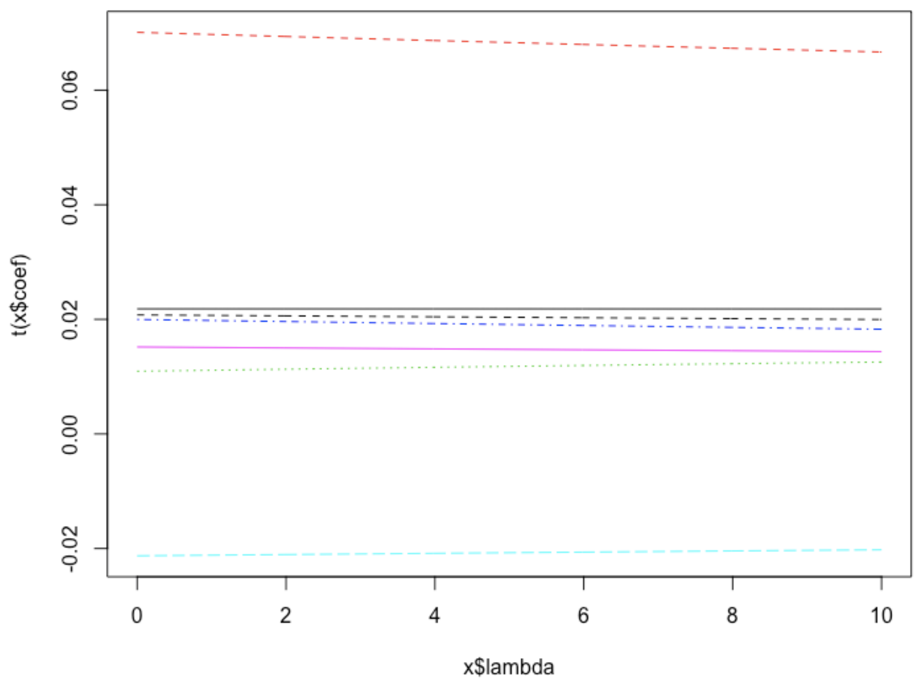


VIF Analysis

```
> VIF(lm(TeamPointsnew~X3PointShots.+FieldGoals.+Assists+TotalRebounds+Turnovers+TotalFouls+Steals, data = bball))
[1] 3.173359
> VIF(lm(TeamPointsnew~X3PointShots., data = bball))
[1] 1.311383
> VIF(lm(TeamPointsnew~FieldGoals., data = bball))
[1] 2.215877
> VIF(lm(TeamPointsnew~Assists, data = bball))
[1] 1.444195
> VIF(lm(TeamPointsnew~TotalRebounds, data = bball))
[1] 1.003771
> VIF(lm(TeamPointsnew~Turnovers, data = bball))
[1] 1.012345
> VIF(lm(TeamPointsnew~TotalFouls, data = bball))
[1] 1.004073
> VIF(lm(TeamPointsnew~Steals, data = bball))
[1] 1.040086
> VIF(lm(TeamPointsnew~FieldGoals.+X3PointShots.+Assists+Turnovers+TotalFouls+TotalRebounds+Steals, data = bball))
[1] 3.173359
> VIF(lm(TeamPointsnew~Assists+FieldGoals.+X3PointShots., data=bball))
[1] 2.488778
> VIF(lm(TeamPointsnew~FieldGoals.+X3PointShots.+Assists, data=bball))
[1] 2.488778
> VIF(lm(TeamPointsnew~X3PointShots.+Assists+FieldGoals., data=bball))
[1] 2.488778
> VIF(lm(TeamPointsnew~X3PointShots.+FieldGoals.+Assists, data=bball))
[1] 2.488778
```

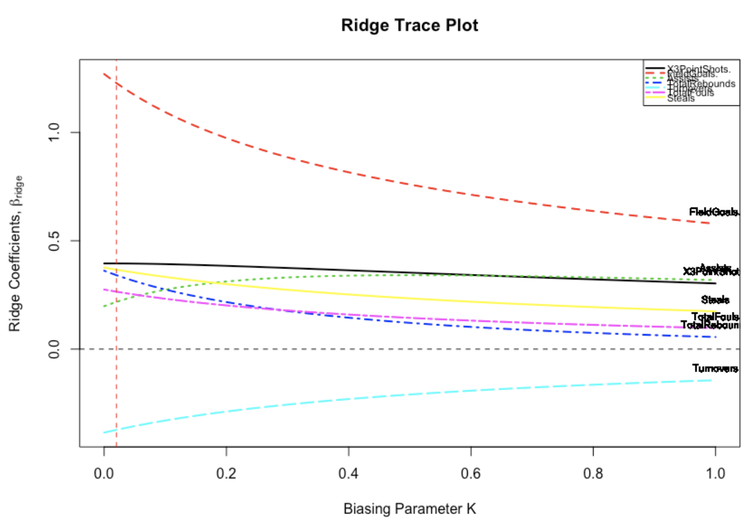
Ridge Regression (MASS Library)

```
> library(MASS)
> bball.ridgelm = lm.ridge(TeamPointsnew~X3PointShots.+FieldGoals.+Assists+To
talRebounds+Turnovers+TotalFouls+Steals, data = bball, lambda=seq(0,10,0.02))
> plot(bball.ridgelm, main="Ridge Trace Plot")
> select(bball.ridgelm)
modified HKB estimator is 2.742669
modified L-W estimator is 2.358101
smallest value of GCV  at 3.34
```



Ridge Regression (lmridge Library)

```
> library(lmridge)
> bball.ridgemod2 = lmridge(TeamPointsnew~X3PointShots.+FieldGoals.+Assists+TotalRebounds+Turnovers+TotalFouls+Steal
s, data = bball, K=seq(0,1,0.02))
> plot(bball.ridgemod2)
> vif(bball.ridgemod2)
       X3PointShots. FieldGoals. Assists TotalRebounds Turnovers TotalFouls  Steals
k=0          1.28553     1.84297 1.61134       1.15817   1.08233    1.02577 1.09242
k=0.02       1.21262     1.67295 1.48001       1.09494   1.03312    0.98344 1.04218
k=0.04       1.14615     1.52720 1.36609       1.03798   0.98747    0.94384 0.99558
k=0.06       1.08536     1.40123 1.26647       0.98633   0.94501    0.90671 0.95227
k=0.08       1.02960     1.29153 1.17874       0.93925   0.90543    0.87184 0.91192
k=0.1        0.97831     1.19537 1.10097       0.89610   0.86846    0.83903 0.87424
k=0.12       0.93100     1.11054 1.03163       0.85640   0.83385    0.80810 0.83901
k=0.14       0.88727     1.03528 0.96948       0.81972   0.80140    0.77892 0.80599
k=0.16       0.84674     0.96817 0.91350       0.78572   0.77092    0.75134 0.77500
k=0.18       0.80910     0.90804 0.86285       0.75410   0.74225    0.72524 0.74587
k=0.2        0.77407     0.85392 0.81684       0.72461   0.71524    0.70051 0.71844
k=0.22       0.74142     0.80502 0.77488       0.69704   0.68975    0.67706 0.69258
k=0.24       0.71091     0.76065 0.73648       0.67121   0.66567    0.65479 0.66817
k=0.26       0.68237     0.72026 0.70123       0.64695   0.64289    0.63363 0.64508
k=0.28       0.65562     0.68337 0.66877       0.62413   0.62132    0.61350 0.62324
k=0.3        0.63051     0.64956 0.63879       0.60261   0.60086    0.59433 0.60253
k=0.32       0.60690     0.61850 0.61104       0.58229   0.58145    0.57606 0.58289
k=0.34       0.58468     0.58988 0.58528       0.56308   0.56299    0.55863 0.56424
k=0.36       0.56373     0.56343 0.56131       0.54488   0.54545    0.54200 0.54650
k=0.38       0.54395     0.53894 0.53896       0.52763   0.52874    0.52611 0.52963
k=0.4        0.52526     0.51620 0.51809       0.51124   0.51281    0.51091 0.51356
k=0.42       0.50757     0.49506 0.49854       0.49566   0.49763    0.49638 0.49823
k=0.44       0.49082     0.47534 0.48022       0.48083   0.48313    0.48246 0.48361
k=0.46       0.47493     0.45693 0.46300       0.46671   0.46928    0.46913 0.46965
k=0.48       0.45985     0.43970 0.44680       0.45323   0.45603    0.45635 0.45630
k=0.5        0.44551     0.42354 0.43154       0.44037   0.44336    0.44409 0.44354
k=0.52       0.43187     0.40838 0.41713       0.42807   0.43123    0.43232 0.43132
k=0.54       0.41889     0.39411 0.40351       0.41632   0.41960    0.42102 0.41962
k=0.56       0.40652     0.38068 0.39063       0.40506   0.40845    0.41017 0.40840
k=0.58       0.39472     0.36800 0.37842       0.39428   0.39776    0.39973 0.39765
k=0.6        0.38345     0.35603 0.36683       0.38395   0.38749    0.38968 0.38732
k=0.62       0.37269     0.34471 0.35583       0.37404   0.37762    0.38002 0.37740
k=0.64       0.36240     0.33399 0.34537       0.36452   0.36814    0.37071 0.36788
k=0.66       0.35255     0.32382 0.33541       0.35538   0.35902    0.36174 0.35871
k=0.68       0.34312     0.31416 0.32592       0.34659   0.35024    0.35310 0.34990
k=0.7        0.33409     0.30499 0.31687       0.33814   0.34179    0.34477 0.34141
k=0.72       0.32542     0.29626 0.30823       0.33000   0.33365    0.33673 0.33324
k=0.74       0.31711     0.28795 0.29997       0.32217   0.32580    0.32897 0.32537
k=0.76       0.30913     0.28002 0.29207       0.31462   0.31823    0.32147 0.31778
k=0.78       0.30146     0.27246 0.28451       0.30734   0.31093    0.31423 0.31046
k=0.8        0.29409     0.26524 0.27727       0.30032   0.30389    0.30724 0.30339
k=0.82       0.28700     0.25833 0.27033       0.29355   0.29708    0.30048 0.29657
k=0.84       0.28018     0.25173 0.26368       0.28701   0.29051    0.29394 0.28998
k=0.86       0.27361     0.24540 0.25728       0.28070   0.28416    0.28761 0.28362
k=0.88       0.26728     0.23934 0.25114       0.27459   0.27801    0.28149 0.27746
k=0.9        0.26118     0.23352 0.24524       0.26869   0.27207    0.27556 0.27151
k=0.92       0.25529     0.22794 0.23956       0.26298   0.26632    0.26981 0.26576
k=0.94       0.24961     0.22258 0.23409       0.25746   0.26076    0.26425 0.26018
k=0.96       0.24413     0.21743 0.22883       0.25212   0.25537    0.25885 0.25479
k=0.98       0.23884     0.21248 0.22375       0.24694   0.25015    0.25363 0.24956
k=1          0.23372     0.20771 0.21886       0.24193   0.24509    0.24855 0.24450
```



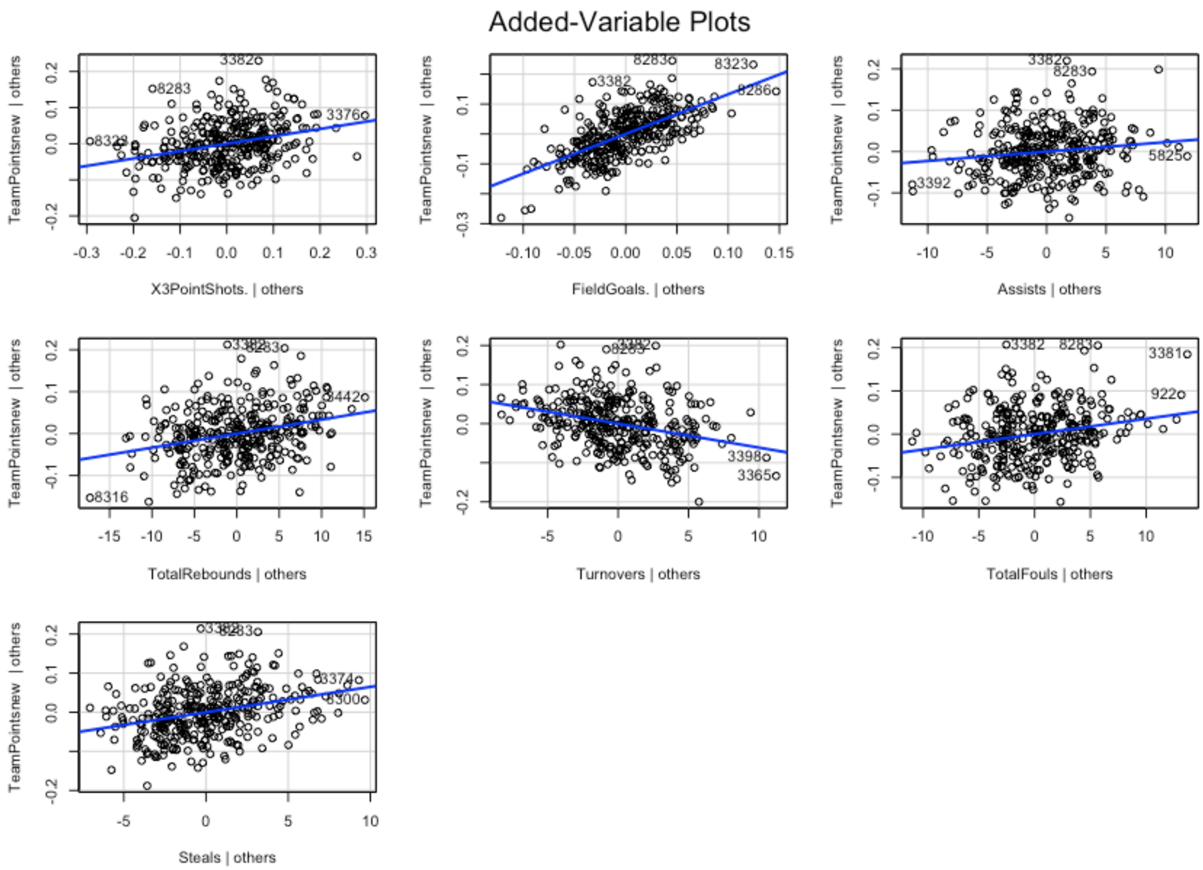
Best Subset Results

```
> bs
  p 1 2 3 4 5 6 7     SSEp        r2    r2.adj        Cp      AICp      SBCp   PRESSp
1 2 0 1 0 0 0 0 0 1.765413 0.5487115 0.5473271 134.27218 -1709.678 -1702.092 1.788633
2 3 1 1 0 0 0 0 0 1.632565 0.5826711 0.5801030 101.78698 -1733.338 -1721.959 1.663727
3 4 1 1 0 1 0 0 0 1.547479 0.6044216 0.6007588  81.69993 -1748.895 -1733.723 1.588621
4 5 1 1 0 1 0 0 1 1.456633 0.6276441 0.6230329  60.11801 -1766.739 -1747.773 1.502996
5 6 1 1 0 1 1 0 1 1.333824 0.6590377 0.6537432  30.23868 -1793.628 -1770.870 1.383960
6 7 1 1 0 1 1 1 1 1.257087 0.6786537 0.6726472  12.31909 -1811.063 -1784.512 1.311566
7 8 1 1 1 1 1 1 1 1.232744 0.6848765 0.6779832   8.00000 -1815.477 -1785.133 1.294454
```

Cross Validation

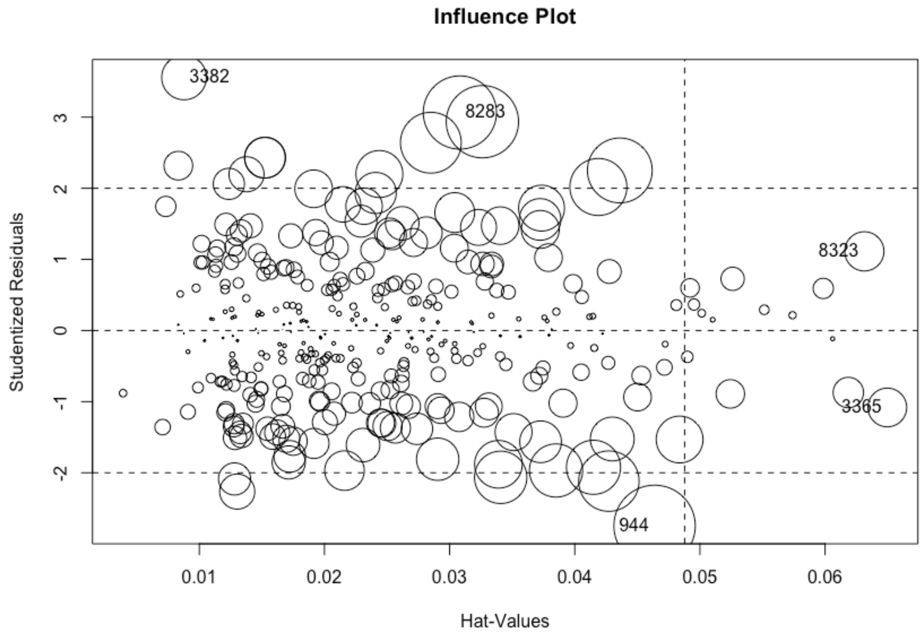
```
> library(leaps)
> library(caret)
> train.control<-trainControl(method="cv", number=5)
> step.model1<-train(TeamPointsnew~X3PointShots.+FieldGoals.+Assists+TotalRebounds+
Turnovers+TotalFouls+Steals, data=bball, method="leapBackward",
+                     tuneGrid=data.frame(nvmax=7),
+                     trControl=train.control)
>
> step.model1$results
  nvmax      RMSE  Rsquared        MAE     RMSESD RsquaredSD       MAESD
1     7 0.0623032 0.6818976 0.04890644 0.01004391 0.07507033 0.007000752
```

AV-Plots



Studentized Deleted Residuals

```
> # Studentized deleted residuals
> qt(1-0.05/(2*328), 328-1-8)
[1] 3.833088
> influencePlot(bball.newmod)
        StudRes         Hat      CookD
944  -2.746173 0.046377453 0.04492704
3365 -1.082044 0.064971720 0.01016407
3382  3.557915 0.008779809 0.01352305
8283  3.064737 0.030819700 0.03638120
8323  1.112924 0.063136943 0.01042618
```




```
> # Hati
> hati = lm.influence(bball.newmod)$hat
> bound = 2*mean(hati)
> bound
[1] 0.04878049
```

```
> #Cook threshold
> qf(0.5,8,320)  #major
[1] 0.9199458
```