1. **Problem description**

As a super basketball fan, I began watch NBA games since kid as a statistics major student I enjoy doing some research on my interesting based on my professional knowledges. The dataset contains all regular season games statistics for all teams.

30 teams: 15 form east conference 15 from west conference

6 divisions: Atlantic, Central, Southeast, Northwest, Pacific and Southwest.

Top 8 from each conference will advanced in to playoffs.

Since the data contains the regular season statistics only without playoffs so I will predict the total game won before the playoffs by given 12 variables.

**Summary:**

The dataset I researched is 14-15 season NBA data. There are total 30 teams in the league, 15 from each conference 5 from each division. Each team participate all 82 games their overall game stats are collected which are my regressors, total game won for regular season is the dependent variable. My job is to create a model to predict the total games won based past or future statistics.

After I briefly review the whole dataset, I found this is full of details something like the post-game report we saw from TV. However, some of the variables are close to another and their relationship are very clear:

1.    FG=P3+P2

2.    FGP=FG/FGA

3.    FGA=P3A+P2A

4.    P3P=P3/P3A

5.    P2P=P2/P2A

6.    TRB=DRB+ORB

Question: How to address those combinations delete them first or later?

Besides, based on my personal experience I assumed assist, field goal percentage and 3p% are better predictors and FT%, minutes played and total number of rebounds are bad predictors.

Based on my chart, the trend is kind clear for my assumptions. All good teams outperform bad teams via 3 points, assists and filed goal %. For minutes played, total number of rebounds and free throw percentage I do not see any clear difference among good or bad teams.

Then I did the all scatter plots and loess plots to assume on how to do transformations. (1/x sqrt(x) will be good transformation during plots)

Then, I begin my prediction constructions. I put all regressions in the initial model first which is a multiple regression model.

1.    Plots are very good, assumptions are violated

2.    Linear combinations multicollinearity

3.    Multicollinearity high VIF

4.    High CV

5.    Some variables are not significant from to T-test

So, I began fix those problems by few transformations base on my residuals vs predicted plots from initial model. Then, I found that there were basically two different patterns open tunnel or close tunnel. So, I made the transformation to my old model by their plots: 1/x for open tunnel and sqrt (x) for close tunnel. After that the result is not quite good I still found patterns in residuals vs precited ones. So, I will try each transformation separately for all 3 transformations (1/x sqrt x and y^2). As results, I found the sqrt(x) one outperforms the other 2 for all aspects R2 CV assumptions QQ plots etc. So, I choose sqrt(X).

When I review my model I found some influential, leverage points and potential outliers list in my model as well however I do not think it’s a problem because all data are collected from professional crew should be no problem at all I cannot delete the team if it played too good or bad.

After I get my transformation model, solved the plots and assumptions problems. I began fix the multicollinearity I tried ridge and PCA however those two methods are not very good because ridge and PCA did not have very good interpretability. Besides, there are still too many variables in the model I will use subset selection try to reduce the numbers of variables and solve multicollinearity problems.

In the end, I tried run subsets selection by using Cp as criteria. Then I found that 3 subsets that satisfy my needs Cp=p+1, by compare all of them I choose the one without multicollinearity. All plot number assumptions are looks very good.

Finally, I try to predict 16-17 season games won by my model However although the number is not very close but the rank prediction is quite good.

If I have more advanced repressors the model could be better.

1. **Investigation of the data**

By checking the data, I found some of the variables are overlapping with others such as total rebounds equal offense rebounds plus defense rebounds and field goal percent equals field goals divide by field goal attempts So, I may delete them in the model to make the prediction looks neat and simple.

First, I calculate the summary statistics for all teams based on the given variables by their mean, stand deviation, minimum, maximum and quantiles. After that I have a big picture that what the kind of data I am dealing like which teams are doing better or worse on 1 or more aspects. Basically, good team do better on all aspects and bad team do worse on all the aspects vice versa.

Second, I sort the origin data by games won instead of alphabetical order for better demonstration of the whole data thought 30 teams. The best team GSW was on the first spot, worst team MIN on the last spot. Based on my own experience I will makes few bar graphs on some of the variables get their summary statistics before the real analysis begins.

I assumed FG%, 3p% and assists will contribute more for a team, so I make 3 bar charts to show their relationships. The x-axis represents each team from best to worst based on games won the y-axis represents the corresponding variables.

Based on those charts, my prediction gets confirmed all those 3 charts share the same pattern that from the left to right (best team to worst team) their FG%, 3p% and assists are generally decreased. By that, I did not mean that for team, its 3 aspects are always better than the team has less games won, however it is easy to see the trend here. I marked them with a straight line.

Meanwhile, I thought following factors are not significant according to my experience (total number of rebounds, free throw percentage and minutes played) I do the same bar charts for them as well. As a result, my assumption gets confirmed again there is not such pattern for these 3 factors. There is no such obvious pattern however couple extreme values are observed I will do research for them in the later section.

Then I do all scatter plots to see their general distribution for games won vs all 12 variables, unlike other scientific experiments, it should be no such outlier exist in this data set since all data are collected from each single game by professional technical statistics crews. Besides, delete one variable for a team sounds unfair because if I delete the factor that a team is good or bad at it (that is why observed as outlier) the model get biased because the best/bad factor could have contributed the most to the model. Based on the results, few outliers, influential points, and leverage points are observed. (Through outlier, influential point and leverage point analysis will be conducted on section 5).

After the scatter plots, I will do the loess plots see are there any transformation needed for those 12 variables, I found that some lines are still not too straight (with some curves) when loess close to 1.0 so combined with results from last part those variables are needing to be transformed.

1. **Specification the model**

1. I will do the simplest model first: the **multiple linear model**.

Y=β0+β1x1+β2x2+…β22x22+∈

2. Transformed 1/X model:

3. Transformed sqrt(X) model:

1. Transformed Y^2 model:

Regressors and response variables are list below:

|  |  |  |
| --- | --- | --- |
|  | **Glossary** |  |
| **Games Won -- See separate list below-----** | | |
| **MP -- Minutes Played-----x1** | |  |
| **FG -- Field Goals-----x2** | |  |
| **FGA -- Field Goal Attempts-----x3** | | |
| **FG% -- Field Goal Percentage-----x4** | | |
| **P3 -- 3-Point Field Goals-----x5** | | |
| **P3A -- 3-Point Field Goal Attempts-----x6** | | |
| **P3P -- 3-Point Field Goal Percentage-----x7** | | |
| **P2 -- 2-Point Field Goals-----x8** | | |
| **P2A -- 2-point Field Goal Attempts-----x9** | | |
| **P2P -- 2-Point Field Goal Percentage-----x10** | | |
| **FT -- Free Throws-----x11** | |  |
| **FTA -- Free Throw Attempts-----x12** | | |
| **FTP -- Free Throw Percentage-----x13** | | |
| **ORB -- Offensive Rebounds-----x14** | | |
| **DRB -- Defensive Rebounds-----x15** | | |
| **TRB -- Total Rebounds-----x16** | |  |
| **AST-- Assists** | **x17** |  |
| **STL --Steals--** | **X18** |  |
| **BLK --Blocks** | **X19** |  |
| **TOV – Turnovers x20** | |  |
| **PF -- Personal Fouls x21** | |  |
| **PTS --Points--** | **X22** |  |

The reason for that is multicollinearity (assumed). Some variables share the same features with others and they are basically mean the same thing for example field goal percentage, filed goal made and field goal attempts. The field goal percentage is equal to field goal made divided by the field goal attempts. So

Model assumptions:

1. Estimation of the Appropriate model

Full estimations are present by 4 different models: origin model, 2 transform X model and 1 transform Y model.

1. Original model:

ANOVA table

| **Analysis of Variance** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | | **Sum of Squares** | | **Mean Square** | | **F Value** | | **Pr > F** |
| **Model** | 18 | | 5151.98893 | | 286.22161 | | 15.53 | | <.0001 |
| **Error** | 11 | | 202.71107 | | 18.42828 | |  | |  |
| **Corrected Total** | 29 | | 5354.70000 | |  | |  | |  |
| **Root MSE** | | | 4.29282 | | **R-Square** | | 0.9621 | |
| **Dependent Mean** | | | 40.90000 | | **Adj R-Sq** | | 0.9002 | |
| **Coeff Var** | | | 10.49589 | |  | |  | |

|  |  |
| --- | --- |
| **Note:** | Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased. |

|  |  |
| --- | --- |
| **Note:** | The following parameters have been set to 0, since the variables are a linear combination of other variables as shown. |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **P2 =** | 3.05E-6 \* Intercept + 1 \* FG - 439E-12 \* FGA - 7.38E-6 \* FGP - 1 \* P3 + 111E-12 \* P3A + 8.02E-7 \* P3P - 4.87E-7 \* PP2 + 794E-13 \* FT - 689E-13 \* FTA - 1.69E-7 \* FTP - 316E-13 \* STL + 186E-13 \* TOV | | | | | | | | | |
| **P2P =** | -6.22E-8 \* Intercept - 582E-13 \* FG + FGA + 348E-12 \* P3 - 1 \* P3A - 6.08E-7 \* P3P - 2.74E-7 \* PP2 - 988E-13 \* FT + 81E-12 \* FTA + 1.98E-7 \* FTP | | | | | | | | | |
| **TRB =** | -4.21E-6 \* Intercept - 1.31E-9 \* FG + 587E-12 \* FGA + 0.00001 \* FGP + 368E-12 \* P3 - 101E-12 \* P3A - 9.35E-7 \* P3P - 9.43E-7 \* PP2 - 788E-13 \* FT + 501E-13 \* FTA + 1.29E-7 \* FTP + ORB + DRB + 406E-13 \* STL - 286E-13 \* BLK - 237E-13 \* TOV | | | | | | | | | |
| **PTS =** | -1.62E-6 \* Intercept + 2 \* FG + 219E-12 \* FGA + 4.37E-6 \* FGP + 1 \* P3 + 101E-12 \* P3A + 2.28E-7 \* P3P - 8.33E-7 \* PP2 + 1 \* FT + 147E-12 \* FTA + 3.7E-7 \* FTP | | | | | | | | | |
| **Parameter Estimates** | | | | | | | | | | |
| **Variable** | | | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Tolerance** | **Variance Inflation** |
| **Intercept** | | | Intercept | B | 1668.90255 | 1366.20016 | 1.22 | 0.2474 | . | 0 |
| **MP** | | | MP | 1 | 0.02041 | 0.01992 | 1.03 | 0.3274 | 0.30318 | 3.29840 |
| **FG** | | | FG | B | 0.64849 | 0.40938 | 1.58 | 0.1415 | 0.00022207 | 4503.16095 |
| **FGA** | | | FGA | B | -0.33215 | 0.18623 | -1.78 | 0.1021 | 0.00047847 | 2089.99121 |
| **FGP** | | | FGP | B | -5878.79687 | 3111.49462 | -1.89 | 0.0855 | 0.00027380 | 3652.27181 |
| **P3** | | | P3 | B | -0.09309 | 0.29252 | -0.32 | 0.7563 | 0.00040459 | 2471.64132 |
| **P3A** | | | P3A | B | 0.01565 | 0.10540 | 0.15 | 0.8847 | 0.00047532 | 2103.83100 |
| **P3P** | | | P3P | B | 481.93538 | 645.70741 | 0.75 | 0.4711 | 0.00451 | 221.78515 |
| **P2** | | | P2 | 0 | 0 | . | . | . | . | . |
| **P2P** | | | P2P | 0 | 0 | . | . | . | . | . |
| **PP2** | | | PP2 | B | 1329.61731 | 968.42960 | 1.37 | 0.1971 | 0.00193 | 518.66120 |
| **FT** | | | FT | B | 0.17319 | 0.25502 | 0.68 | 0.5111 | 0.00050551 | 1978.19585 |
| **FTA** | | | FTA | B | -0.12829 | 0.19414 | -0.66 | 0.5223 | 0.00050132 | 1994.74552 |
| **FTP** | | | FTP | B | -170.52589 | 478.38912 | -0.36 | 0.7282 | 0.00379 | 264.13021 |
| **ORB** | | | ORB | B | 0.05194 | 0.02153 | 2.41 | 0.0345 | 0.20731 | 4.82361 |
| **DRB** | | | DRB | B | 0.07008 | 0.01487 | 4.71 | 0.0006 | 0.19935 | 5.01628 |
| **TRB** | | | TRB | 0 | 0 | . | . | . | . | . |
| **AST** | | | AST | 1 | 0.01646 | 0.01178 | 1.40 | 0.1900 | 0.19797 | 5.05123 |
| **STL** | | | STL | B | 0.10645 | 0.02824 | 3.77 | 0.0031 | 0.11626 | 8.60158 |
| **BLK** | | | BLK | B | 0.00245 | 0.02233 | 0.11 | 0.9147 | 0.38050 | 2.62813 |
| **TOV** | | | TOV | B | -0.07654 | 0.01827 | -4.19 | 0.0015 | 0.16690 | 5.99154 |
| **PF** | | | PF | 1 | -0.00675 | 0.01144 | -0.59 | 0.5672 | 0.34841 | 2.87021 |
| **PTS** | | | PTS | 0 | 0 | . | . | . | . | . |

As you can see some variables are the linear combinations of every variables, so they are denoted by others and that is why they are not show in the full model.

CV: 10.49 R^2：96.21%

1. The prediction equation for original model is

Y-hat=1668.90-0.02041MP+0.64849FG-0.33215FGA--5878.79687FGP-0.09309P3+0.01565P3A+481.93538P3P+1329.61731PP2+0.17319FT-0.12829FTA-170.52589FTP+0.05194ORB+0.07008DRB+0.01646AST+0.10645STL+0.00245BLK-0.07654TOV-0.00675PF.

1. Testing hypothesis on F-Test (ANOVA table):

H0: this model is not fit H1: this model is fit α=0.05

Since the F value is 15.53 and the P-value is smaller than 0.001 which is also smaller than α so reject the null hypothesis, the model is adequate.

1. Testing hypothesis on T-test (each variables):

H0: at least βi (β1 to β22) is not 0 H1: all bi are 0 α=0.05

According to the parameter estimates tables above, for all 22 variables most of their P-value is larger than 0.05 except ORB DRB STL and TOV

1. Comment:

For the initial model, P-value is 96.21% which is pretty good, the CV is 10.49 which is litter bit large.

The real problem is linear combination and T-test. Since the model is not full rank and estimates are biased, so we have more than 1 linear combination for the model. Some of variables are omitted from the model (atomically by the SAS reg PROC) what if the omitted variables are important we may have lost the important information. Besides, only rebounds and turn overs are significant in the models which is make no sense, none of the field goal (2 points,3 points) factors are involved which is suspicious. Eventually, this model is not appropriate.

1. Transform on X: X=1/X

ANOVA table

| **Analysis of Variance** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | | **Sum of Squares** | | **Mean Square** | | **F Value** | | **Pr > F** |
| **Model** | 22 | | 5299.60497 | | 240.89113 | | 30.61 | | <.0001 |
| **Error** | 7 | | 55.09503 | | 7.87072 | |  | |  |
| **Corrected Total** | 29 | | 5354.70000 | |  | |  | |  |
| **Root MSE** | | | 2.80548 | | **R-Square** | | 0.9897 | |
| **Dependent Mean** | | | 40.90000 | | **Adj R-Sq** | | 0.9574 | |
| **Coeff Var** | | | 6.85936 | |  | |  | |

| **Parameter Estimates** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Tolerance** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | -16.41961 | 1083.68823 | -0.02 | 0.9883 | . | 0 |
| **MP1** |  | 1 | 3342240 | 6224591 | 0.54 | 0.6079 | 0.20498 | 4.87845 |
| **FG1** |  | 1 | -7004342 | 3148767 | -2.22 | 0.0615 | 0.00014077 | 7103.61076 |
| **FGA1** |  | 1 | 6222898 | 7581877 | 0.82 | 0.4388 | 0.00026525 | 3769.98740 |
| **FGP1** |  | 1 | -309.49474 | 771.84338 | -0.40 | 0.7004 | 0.00007367 | 13575 |
| **P31** |  | 1 | -98409 | 144013 | -0.68 | 0.5164 | 0.00010567 | 9463.77743 |
| **P3A1** |  | 1 | 332204 | 408325 | 0.81 | 0.4427 | 0.00013053 | 7661.17841 |
| **P3P1** |  | 1 | 212.68972 | 170.03293 | 1.25 | 0.2512 | 0.00042753 | 2339.04249 |
| **P21** |  | 1 | 2837743 | 1144662 | 2.48 | 0.0423 | 0.00031898 | 3134.94401 |
| **P2a1** |  | 1 | -1742904 | 2778286 | -0.63 | 0.5503 | 0.00017887 | 5590.59556 |
| **P2p1** |  | 1 | 273.29455 | 617.40971 | 0.44 | 0.6714 | 0.00010903 | 9172.09596 |
| **FT1** |  | 1 | 449446 | 417725 | 1.08 | 0.3176 | 0.00034413 | 2905.84891 |
| **FTA1** |  | 1 | -841113 | 527753 | -1.59 | 0.1550 | 0.00036786 | 2718.44002 |
| **FTP1** |  | 1 | -431.81372 | 217.80833 | -1.98 | 0.0879 | 0.00233 | 428.46713 |
| **ORB1** |  | 1 | -174347 | 111304 | -1.57 | 0.1612 | 0.00199 | 501.85396 |
| **DRB1** |  | 1 | -1657995 | 1040946 | -1.59 | 0.1552 | 0.00085144 | 1174.48378 |
| **TRB1** |  | 1 | 2495476 | 1830306 | 1.36 | 0.2150 | 0.00058230 | 1717.31457 |
| **AST1** |  | 1 | -82599 | 29544 | -2.80 | 0.0267 | 0.16616 | 6.01848 |
| **STL1** |  | 1 | -7700.96583 | 9812.03174 | -0.78 | 0.4583 | 0.06531 | 15.31103 |
| **BLK1** |  | 1 | -8842.12572 | 3824.25956 | -2.31 | 0.0540 | 0.13707 | 7.29550 |
| **TOV1** |  | 1 | 58226 | 18568 | 3.14 | 0.0165 | 0.13879 | 7.20515 |
| **PF1** |  | 1 | -9125.08558 | 31679 | -0.29 | 0.7816 | 0.14309 | 6.98884 |
| **PTS1** |  | 1 | 6460631 | 3205160 | 2.02 | 0.0837 | 0.00096223 | 1039.25751 |

CV: 6.86 R^2：98.97%

The prediction equation for original model is

Y-hat=-16.41963+540(1/MP)-7004323(1/FG)+6222898(1/FGA)-309.49474(1/FGP)-98409(1/P3)+332204(1/P3A)+212.689372(1/P3P)+2837743(1/P2)-1742904(1/P2A)+273.29455(1/P2P)+449446(1/FT)-841113(1/FTA)-431.81372(1/FTP)-174347(1/ORB)-1657995(1/DRB)+2495476(1/TRB)-82599(1/AST)-7700(1/STL)-8842(1/BLK)+582226(1/TOV)-9125(1/PF)+6460631(1/PTS)

Testing hypothesis on F-Test:

H0: this model is not fit H1: this model is fit α=0.05

Since the F value is 30.61 he P-value is smaller than 0.001 which is also smaller than α so reject the null hypothesis, the model is adequate.

Testing hypothesis on T-test (each variables):

According to the parameter estimates tables above, for all 22 variables most of their P-value is larger than 0.05 except TOV and AST.

Comment:

In this transformation(x=1/x), both CV and R^2 are getting better and no linear combinations issues, however for the T-test only have 2 variables in this model.

1. Transform on X: X=sqrt (X)

ANOVA table

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 22 | 5279.72535 | 239.98752 | 22.41 | 0.0002 |
| **Error** | 7 | 74.97465 | 10.71066 |  |  |
| **Corrected Total** | 29 | 5354.70000 |  |  |  |

CV: 8 R^2 98.6%：

The prediction equation for original model is

Y-hat= 87+1.29sqrt(MP)+274.88sqrt(FG)-51.72662sqrt(FGA)+2769sqrt(FGP)-108sqrt(P3) +55.63sqrt(P3A) +1333sqrt(P3P)-160.56sqrt(P2)+38.48sqrt(P2A)-1835.15(P2P)+80.293sqrt(FTA)+839.58sqrt(FTP)+55.67sqrt(ORB)+98.14sqrt(DRB)+98.14sqrt(TRB)+1.55sqrt(AST)+1.65sqrt(STL)+1.8676sqrt(BLK)-3.15sqrt(TOV)+0.19sqrt(PF)-76.15sqrt(PTS)

Testing hypothesis on F-Test:

H0: this model is not fit H1: this model is fit

Since the F value is 22.41 he P-value is smaller than 0.001 which is also smaller than α so reject the null hypothesis, the model is adequate.

Testing hypothesis on T-test (each variables)

According to the parameter estimates tables above, for all 22 variables most of their P-value is larger than 0.05 only AST is an exceptional.

Comment:

In this transformation(x=sqrt(x)) CV and R^2 are better than original model as well. Linear combination is not an issue. Only one variable has P-value smaller than 0.05 which is AST.

1. Transform on Y: Y=Y^2.

ANOVA table

| **Analysis of Variance** | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | | **Sum of Squares** | | **Mean Square** | **F Value** | | | **Pr > F** |
| **Model** | 18 | | 33171859 | | 1842881 | 21.28 | | | <.0001 |
| **Error** | 11 | | 952661 | | 86606 |  | | |  |
| **Corrected Total** | 29 | | 34124520 | |  |  | | |  |
| **Root MSE** | | | 294.28824 | | **R-Square** | | | 0.9721 |
| **Dependent Mean** | | | 1851.30000 | | **Adj R-Sq** | | | 0.9264 |
| **Coeff Var** | | | 15.89630 | |  | | |  |

|  |  |
| --- | --- |
| **Note:** | Model is not full rank. Least-squares solutions for the parameters are not unique. Some statistics will be misleading. A reported DF of 0 or B means that the estimate is biased. |
| **Note:** | The following parameters have been set to 0, since the variables are a linear combination of other variables as shown. |

|  |  |
| --- | --- |
| **P2 =** | 3.32E-6 \* Intercept + 1 \* FG - 424E-12 \* FGA - 8.36E-6 \* FGP - 1 \* P3 + 402E-13 \* P3A + 8.19E-7 \* P3P - 1.34E-6 \* P2P - 176E-12 \* FT + 127E-12 \* FTA + 3.11E-7 \* FTP - 163E-13 \* DRB - 331E-13 \* STL + 171E-13 \* TOV |
| **P2A =** | 2.87E-6 \* Intercept + 902E-12 \* FG + 1 \* FGA - 8.72E-6 \* FGP - 424E-12 \* P3 - 1 \* P3A + 1.41E-6 \* P3P - 1.84E-6 \* P2P + 474E-13 \* FT - 454E-13 \* FTA |
| **TRB =** | 1.01E-6 \* Intercept + 301E-12 \* FG - 129E-12 \* FGA - 3.22E-6 \* FGP - 133E-12 \* P3 + 218E-13 \* P3A + 5.35E-7 \* P3P + 8.47E-7 \* P2P + ORB + DRB |
| **PTS =** | -2.09E-6 \* Intercept + 2 \* FG + 308E-12 \* FGA + 8.58E-6 \* FGP + 1 \* P3 + 963E-13 \* P3A - 7.51E-7 \* P3P - 2.76E-6 \* P2P + 1 \* FT - 914E-13 \* FTA - 2.33E-7 \* FTP |

CV: 15.89 R^2：97.21%  
The prediction equation for original model is

Y-hat= sqrt (208613+1.128MP+72.466FG-35.34FGA-619668FGP+4.52P3-3.123P3A+22183P3P+106297P2P+10.94713FT+-7.8FTA-8423FTP+3.44ORB+5.06DRB+1.5AST+9.1STL-0.03BLK-5.737TOV-0.567PF)

Testing hypothesis on F-Test:

H0: this model is not fit H1: this model is fit

Since the F value is 21.28 p-value is smaller than 0.001 which is also smaller than α so reject the null hypothesis, the model is adequate.

Testing hypothesis on T-test (each variables)

According to the parameter estimates tables above, for all 22 variables most of their P-value is larger than 0.05 except FG FGA FGP ORB DRB AST STL and TOV.

Comment:

R2 is better than original model but CV is worse than original model. Linear combination problems still exist since I only do the transformation on X Instead Y. For the T-test I get more variables that are significant in this model than transformation on X models.

Conclusion：

Pros: all transformation has better CV and R2 values than original model except transformation on Y model.

Cons: Linear combination problems, Multicollinearity problem (large VIF number >1000), very few significant variables in original model and 2 transformations on X models.

For final decision, plots, influential and leverage point will be analysis in the next section

.

1. Assessment of the chosen Prediction Equation.

Based on the results from the part 4, I found both transform on X or Y improve the fit However some variables may not need to be transformed so I will take look at some plots for the original model.

**Original multiple linear regression:**

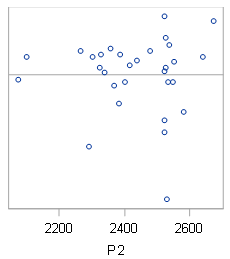
I will take look at those predicted vs residuals plots, points are equally distributed on the side of 0 line however the shape for the plots is obvious which is close tunnel. So, transformation is NECESSARY, I will try few transformations on X if none of them success I will then consider transformation on Y.

For normality, most of the point are close to the straight line with only little bit tail problem which is accepted.



For the Residual vs predictors plots, points are equally distributed by the 0 line for all plots which is good, however some patterns are found during these plots and they are mainly 2 patterns there: a close tunnel shape or an open tunnel shape. I list 2 typical ones to represent each of them respectively. (I will list all the corresponding partial plots in the Appendix part)





So, for those 2 typical plots, I will do transformation those 2 diverse groups: for open tunnel, I will go new x=1/x for close tunnel I will go x=sqrt(x), for plots with no obvious pattern no transformation is performed.

In summary, following variables are being transformed:

New x=(1/x): P2 P2A ORB AST PTS--------5

New x=sqrt (x): MP FG FGP P3 P3A P3P P2P FT FTP FTA BLK ---------5

No change: FGA DRB TRB STL TOV PF-----12

So, the new prediction equation is:

Y=β0+β1(sqrtx1)+β2(sqrtx2)+β3x3+β4(sqrtx4)+β5(sqrtx5)+β6(sqrtx6)+β7(sqrtx7)+β8(1/x8)+β9(1/x9)+β10(sqrtx10)+β11(sqrtx11)+β12(sqrtx12)+β13(sqrtx13)+β14(1/x14)+β15x15+β16x16+β17(1/x17)+β18x18+β19(sqrtx19)+β20x20+β21x21+β22(1/x22)+error.

Put this transformation in to SAS reg proc I get:

ANOVA table:

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 22 | 5218.69542 | 237.21343 | 12.21 | 0.0012 |
| **Error** | 7 | 136.00458 | 19.42923 |  |  |
| **Corrected Total** | 29 | 5354.70000 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Root MSE** | | 4.40786 | | | **R-Square** | | 0.9746 |
| **Dependent Mean** | | 40.90000 | | | **Adj R-Sq** | | 0.8948 |
| **Coeff Var** | | 10.77716 | | |  | |  |
| **Parameter Estimates** | | | | | | | | | | | | | |
| **Variable** | | **Label** | | **DF** | **Parameter Estimate** | | **Standard Error** | | | **t Value** | **Pr > |t|** | **Tolerance** | **Variance Inflation** |
| **Intercept** | | Intercept | | 1 | -1101.62514 | | 7408.52980 | | | -0.15 | 0.8860 | . | 0 |
| **MP2** | |  | | 1 | 7.23961 | | 6.57579 | | | 1.10 | 0.3073 | 0.23257 | 4.29978 |
| **FG2** | |  | | 1 | 20.72963 | | 74.20992 | | | 0.28 | 0.7881 | 0.00008745 | 11435 |
| **FGA** | | FGA | | 1 | 0.02578 | | 0.34566 | | | 0.07 | 0.9426 | 0.00014642 | 6829.64489 |
| **FGP2** | |  | | 1 | 7876.05956 | | 12297 | | | 0.64 | 0.5422 | 0.00003285 | 30441 |
| **P32** | |  | | 1 | -32.40570 | | 64.63408 | | | -0.50 | 0.6315 | 0.00002250 | 44444 |
| **P3A2** | |  | | 1 | 14.18204 | | 41.28177 | | | 0.34 | 0.7413 | 0.00002373 | 42132 |
| **P3P2** | |  | | 1 | -1096.15992 | | 2062.24448 | | | -0.53 | 0.6115 | 0.00065846 | 1518.70637 |
| **P21** | |  | | 1 | 542511 | | 2539407 | | | 0.21 | 0.8369 | 0.00015999 | 6250.27471 |
| **P2a1** | |  | | 1 | 4115021 | | 7972440 | | | 0.52 | 0.6216 | 0.00005362 | 18649 |
| **P2p2** | |  | | 1 | -6618.36232 | | 10442 | | | -0.63 | 0.5464 | 0.00003372 | 29653 |
| **FT2** | |  | | 1 | 50.47777 | | 41.68334 | | | 1.21 | 0.2652 | 0.00011614 | 8610.09267 |
| **FTA2** | |  | | 1 | -47.21028 | | 35.76289 | | | -1.32 | 0.2283 | 0.00011885 | 8413.98051 |
| **FTP2** | |  | | 1 | -2046.27227 | | 1759.74880 | | | -1.16 | 0.2830 | 0.00087406 | 1144.08507 |
| **ORB1** | |  | | 1 | 2431.39874 | | 92945 | | | 0.03 | 0.9799 | 0.00705 | 141.76187 |
| **DRB** | | DRB | | 1 | -0.00857 | | 0.12351 | | | -0.07 | 0.9466 | 0.00305 | 328.05599 |
| **TRB** | | TRB | | 1 | 0.04094 | | 0.12584 | | | 0.33 | 0.7544 | 0.00196 | 510.04782 |
| **AST1** | |  | | 1 | -53126 | | 51497 | | | -1.03 | 0.3366 | 0.13500 | 7.40728 |
| **STL** | | STL | | 1 | 0.01485 | | 0.06115 | | | 0.24 | 0.8152 | 0.02613 | 38.27252 |
| **BLK2** | |  | | 1 | 2.94831 | | 2.32905 | | | 1.27 | 0.2461 | 0.05945 | 16.81985 |
| **TOV** | | TOV | | 1 | -0.06578 | | 0.02196 | | | -3.00 | 0.0201 | 0.12175 | 8.21388 |
| **PF** | | PF | | 1 | 0.02439 | | 0.02628 | | | 0.93 | 0.3843 | 0.06961 | 14.36576 |
| **PTS1** | |  | | 1 | -2125801 | | 13837414 | | | -0.15 | 0.8822 | 0.00012744 | 7846.79764 |

CV: 10.72 R2:97.12%

**Original model summary:**

The CV and R2 are like the original which is accepted, QQ-plot looks very good, however Residual VS Predicted value shows some pattern there it looks like a rhombus.

For the residuals vs each regressor, I found some of them are getting better after transformation, but most of them still form obvious patterns that similar to the original plots so this particular transformation fails. I list some of the plots here. (All partial plots are in the Appendix section)

After I check the plots problems, next step is take look at the influential points, leverage points and outliers, by doing that I will implement the results from SAS reg proc diagnostic part.

| **Output Statistics** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Obs** | **Dependent Variable** | **Predicted Value** | **Std Error Mean Predict** | **Residual** | **Std Error Residual** | **Student Residual** | -2-1 0 1 2 | **Cook's D** | **RStudent** | **Hat Diag H** |
|
| **1** | 60.0000 | 59.7585 | 4.1978 | 0.2415 | 1.344 | 0.180 | | | | | 0.014 | 0.1667 | 0.9070 |
| **2** | 40.0000 | 38.8221 | 3.9448 | 1.1779 | 1.967 | 0.599 | | |\* | | 0.063 | 0.5693 | 0.8009 |
| **3** | 38.0000 | 37.1925 | 3.9490 | 0.8075 | 1.958 | 0.412 | | | | | 0.030 | 0.3865 | 0.8026 |
| **4** | 33.0000 | 31.8506 | 4.1044 | 1.1494 | 1.607 | 0.715 | | |\* | | 0.145 | 0.6877 | 0.8670 |
| **5** | 50.0000 | 50.7208 | 4.0499 | -0.7208 | 1.740 | -0.414 | | | | | 0.040 | -0.3883 | 0.8442 |
| **6** | 53.0000 | 50.4453 | 3.5452 | 2.5547 | 2.619 | 0.975 | | |\* | | 0.076 | 0.9714 | 0.6469 |
| **7** | 50.0000 | 48.3618 | 3.1695 | 1.6382 | 3.063 | 0.535 | | |\* | | 0.013 | 0.5055 | 0.5171 |
| **8** | 30.0000 | 31.3084 | 3.1844 | -1.3084 | 3.048 | -0.429 | | | | | 0.009 | -0.4028 | 0.5219 |
| **9** | 32.0000 | 30.8868 | 3.8610 | 1.1132 | 2.127 | 0.523 | | |\* | | 0.039 | 0.4944 | 0.7672 |
| **10** | 67.0000 | 66.8169 | 4.1477 | 0.1831 | 1.492 | 0.123 | | | | | 0.005 | 0.1137 | 0.8855 |
| **11** | 56.0000 | 56.0073 | 4.3042 | -0.007287 | 0.950 | -0.0077 | | | | | 0.000 | -0.007100 | 0.9535 |
| **12** | 38.0000 | 34.8208 | 3.0807 | 3.1792 | 3.153 | 1.008 | | |\*\* | | 0.042 | 1.0099 | 0.4885 |
| **13** | 56.0000 | 56.9712 | 3.8385 | -0.9712 | 2.167 | -0.448 | | | | | 0.027 | -0.4211 | 0.7584 |
| **14** | 21.0000 | 23.6366 | 3.4184 | -2.6366 | 2.783 | -0.947 | | \*| | | 0.059 | -0.9395 | 0.6014 |
| **15** | 55.0000 | 53.3081 | 4.1504 | 1.6919 | 1.484 | 1.140 | | |\*\* | | 0.442 | 1.1694 | 0.8866 |
| **16** | 37.0000 | 34.9923 | 3.8568 | 2.0077 | 2.134 | 0.941 | | |\* | | 0.126 | 0.9319 | 0.7656 |
| **17** | 41.0000 | 37.3394 | 3.7319 | 3.6606 | 2.346 | 1.561 | | |\*\*\* | | 0.268 | 1.7893 | 0.7168 |
| **18** | 16.0000 | 17.9768 | 3.8090 | -1.9768 | 2.218 | -0.891 | | \*| | | 0.102 | -0.8762 | 0.7467 |
| **19** | 45.0000 | 41.8228 | 4.0428 | 3.1772 | 1.756 | 1.809 | | |\*\*\* | | 0.754 | 2.2950 | 0.8412 |
| **20** | 17.0000 | 17.1283 | 3.9014 | -0.1283 | 2.051 | -0.0625 | | | | | 0.001 | -0.0579 | 0.7834 |
| **21** | 45.0000 | 43.4063 | 4.1323 | 1.5937 | 1.534 | 1.039 | | |\*\* | | 0.341 | 1.0458 | 0.8789 |
| **22** | 22.0000 | 25.9983 | 3.8751 | -3.9983 | 2.101 | -1.903 | | \*\*\*| | | 0.536 | -2.5371 | 0.7729 |
| **23** | 18.0000 | 18.4606 | 4.3116 | -0.4606 | 0.916 | -0.503 | | \*| | | 0.244 | -0.4742 | 0.9568 |
| **24** | 39.0000 | 38.0179 | 3.4522 | 0.9821 | 2.741 | 0.358 | | | | | 0.009 | 0.3348 | 0.6134 |
| **25** | 51.0000 | 51.9901 | 4.0489 | -0.9901 | 1.742 | -0.568 | | \*| | | 0.076 | -0.5387 | 0.8437 |
| **26** | 29.0000 | 28.5932 | 3.7921 | 0.4068 | 2.247 | 0.181 | | | | | 0.004 | 0.1680 | 0.7401 |
| **27** | 55.0000 | 60.3092 | 3.6147 | -5.3092 | 2.523 | -2.105 | | \*\*\*\*| | | 0.395 | -3.2158 | 0.6725 |
| **28** | 49.0000 | 51.2612 | 3.7872 | -2.2612 | 2.255 | -1.003 | | \*\*| | | 0.123 | -1.0031 | 0.7382 |
| **29** | 38.0000 | 41.3742 | 4.0289 | -3.3742 | 1.788 | -1.887 | | \*\*\*| | | 0.786 | -2.4927 | 0.8354 |
| **30** | 46.0000 | 47.4217 | 4.0530 | -1.4217 | 1.733 | -0.820 | | \*| | | 0.160 | -0.7990 | 0.8455 |

| **Output Statistics** | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Obs** | **Cov Ratio** | **DFFITS** | **DFBETAS** | | | | | | | | | | | |
| **Intercept** | **MP2** | **FG2** | **FGA** | **FGP2** | **P32** | **P3A2** | **P3P2** | **P21** | **P2a1** | **P2p2** | **FT2** |
| **1** | 334.9959 | 0.5204 | 0.0019 | -0.0662 | 0.0778 | -0.0157 | 0.1027 | -0.1645 | 0.1593 | 0.0286 | -0.1492 | 0.1667 | -0.1718 | 0.0078 |
| **2** | 51.9155 | 1.1419 | 0.0252 | -0.3570 | -0.0882 | 0.0174 | -0.1329 | 0.0360 | -0.0270 | 0.1902 | 0.0983 | -0.1703 | 0.1989 | 0.0436 |
| **3** | 99.7425 | 0.7793 | 0.2903 | 0.4102 | 0.0028 | -0.1814 | -0.0049 | 0.2171 | -0.2216 | -0.5359 | 0.0086 | 0.0118 | -0.0793 | -0.1182 |
| **4** | 45.5234 | 1.7562 | 0.2000 | -0.0214 | 0.2479 | -0.1962 | -0.1253 | 0.0507 | -0.0104 | -0.0308 | -0.4551 | 0.2486 | -0.1071 | 0.2517 |
| **5** | 125.6714 | -0.9040 | 0.1047 | -0.2842 | 0.5391 | -0.4652 | -0.2702 | -0.0670 | 0.1037 | 0.1430 | -0.3069 | 0.1340 | -0.0567 | -0.3721 |
| **6** | 3.4106 | 1.3148 | -0.2645 | -0.0724 | 0.5583 | -0.1702 | -0.0997 | -0.1712 | 0.2446 | 0.3841 | -0.3632 | 0.2356 | -0.1043 | 0.3519 |
| **7** | 27.4919 | 0.5231 | -0.2848 | -0.0366 | 0.0888 | 0.1119 | 0.0507 | -0.1670 | 0.1798 | 0.2938 | -0.0251 | 0.0407 | -0.0119 | 0.0476 |
| **8** | 39.2432 | -0.4209 | -0.0203 | 0.0031 | 0.0484 | -0.0587 | -0.1614 | 0.1040 | -0.0964 | 0.1025 | 0.1061 | -0.1441 | 0.1724 | 0.0414 |
| **9** | 59.4258 | 0.8977 | 0.1523 | 0.3302 | -0.1622 | -0.0535 | -0.1519 | 0.2923 | -0.2878 | -0.1849 | 0.2024 | -0.2416 | 0.2134 | 0.0062 |
| **10** | 287.9457 | 0.3162 | -0.0053 | 0.0047 | -0.0319 | 0.0432 | 0.0333 | 0.0557 | -0.0512 | -0.0894 | 0.0202 | -0.0030 | -0.0060 | 0.0625 |
| **11** | 745.5840 | -0.0322 | -0.0050 | -0.0063 | 0.0062 | -0.0012 | 0.0030 | -0.0065 | 0.0078 | 0.0050 | -0.0045 | 0.0037 | -0.0050 | -0.0017 |
| **12** | 1.8315 | 0.9868 | 0.2639 | -0.2925 | 0.2307 | -0.3362 | -0.2271 | -0.0173 | 0.0328 | 0.1169 | -0.1193 | -0.0106 | 0.0336 | -0.0097 |
| **13** | 73.4069 | -0.7459 | -0.0545 | 0.1033 | 0.0552 | 0.0325 | 0.0350 | -0.0644 | 0.0653 | 0.0640 | -0.0533 | 0.0668 | -0.0454 | 0.2602 |
| **14** | 3.7021 | -1.1540 | 0.5287 | -0.1870 | 0.2133 | -0.4651 | -0.2468 | 0.1383 | -0.1519 | -0.3145 | 0.0938 | -0.1155 | 0.0496 | 0.1216 |
| **15** | 2.7176 | 3.2700 | -0.2626 | -0.0091 | -0.1543 | 0.1919 | 0.2017 | -0.0499 | -0.0621 | -0.1996 | 0.7837 | -0.2930 | 0.0731 | -0.2253 |
| **16** | 6.5990 | 1.6843 | -0.5150 | 0.1239 | -0.5696 | 0.6374 | 0.1116 | 0.2389 | -0.2430 | 0.1254 | 0.4830 | -0.3684 | 0.3826 | 0.1379 |
| **17** | 0.0066 | 2.8468 | 0.6292 | 0.6661 | -0.3465 | -0.3615 | -0.4605 | 0.6208 | -0.5345 | -0.2824 | -0.3178 | -0.2284 | 0.3282 | -0.5602 |
| **18** | 8.5782 | -1.5045 | -0.1029 | 0.4396 | -0.2011 | 0.1775 | 0.1839 | -0.0253 | 0.0134 | -0.1031 | 0.1710 | -0.0433 | -0.0417 | 0.1603 |
| **19** | 0.0001 | 5.2826 | 1.2504 | -1.3151 | -0.2698 | -0.6550 | -0.5214 | -0.4636 | 0.1845 | 0.4680 | 0.5931 | -0.4809 | 0.3316 | -0.8193 |
| **20** | 157.9710 | -0.1101 | 0.0125 | -0.0167 | -0.0034 | -0.0033 | 0.0026 | -0.0155 | 0.0147 | 0.0129 | -0.0135 | 0.0082 | -0.0077 | 0.0018 |
| **21** | 6.0792 | 2.8172 | -0.0972 | -0.2174 | 0.8609 | -0.6387 | -0.7426 | 0.2886 | -0.2384 | 0.3183 | 0.1882 | -0.3338 | 0.4747 | 0.2489 |
| **22** | 0.0000 | -4.6803 | -1.3070 | 0.5052 | 0.9638 | 0.0743 | 0.7009 | -0.9524 | 1.1031 | 0.3865 | -1.2595 | 1.1944 | -1.0909 | -1.5364 |
| **23** | 344.3553 | -2.2319 | 0.2416 | 0.2830 | -0.3858 | 0.1755 | 0.0302 | 0.3476 | -0.3424 | -0.3687 | -0.2565 | 0.0996 | 0.0271 | 0.1376 |
| **24** | 58.5536 | 0.4218 | -0.0491 | 0.0048 | -0.0406 | 0.1003 | 0.1558 | -0.1390 | 0.1368 | 0.0201 | -0.1399 | 0.1555 | -0.1566 | 0.0082 |
| **25** | 74.8457 | -1.2517 | 0.1441 | 0.1952 | 0.1319 | -0.0934 | -0.0480 | 0.0257 | -0.0092 | -0.0256 | -0.1269 | 0.0759 | -0.0365 | 0.6830 |
| **26** | 119.7169 | 0.2835 | 0.0259 | -0.0191 | -0.0037 | -0.0279 | -0.0292 | 0.0232 | -0.0213 | -0.0068 | 0.0248 | -0.0359 | 0.0268 | -0.0497 |
| **27** | 0.0000 | -4.6081 | -0.9051 | -1.1034 | -0.8885 | 2.0690 | 2.9514 | -1.7377 | 1.5768 | -0.8644 | -1.0605 | 2.1653 | -2.4895 | 0.7594 |
| **28** | 3.7433 | -1.6845 | -0.4150 | -0.0545 | -0.3605 | 0.4446 | 0.0608 | 0.1436 | -0.1387 | 0.1497 | 0.1165 | -0.1360 | 0.2167 | -0.4084 |
| **29** | 0.0000 | -5.6167 | 0.2340 | 1.9787 | -1.1905 | 0.3640 | -0.7682 | 1.2265 | -1.4711 | -0.3478 | 2.6112 | -1.9942 | 1.6830 | 0.9395 |
| **30** | 21.9154 | -1.8690 | 0.7352 | -0.1929 | -0.3349 | -0.5117 | -0.8728 | 0.9826 | -0.9407 | -0.3417 | 0.5573 | -0.9210 | 0.9081 | -0.6691 |

According to the output statistics and summary statistics I found following are points are going to be either influential points or leverage points.

They are total 30 observations and 22 variables in the dataset.

P=number of xi+1 =22+1=23

N=30

Hii=2p/n=1.53 (leverage point): None

Cook’s D: >1: None

|DFFITS| > 2/sqrt(p/n) =2.28: No.15, NO.17, NO.19, No.21, No.22. No 27 No 29

|DFBETAS|>2/sqrt(n)=0.365:

Following obs has large number of times that exceeds the range:

No17. No 21 No 22 No.27 No.29

COVRATIO: 1±3p/n= (-1.3,3.3): No17 No 22 No 27 No 29

|RSTUDENT|>= 3: No 27

They are total 30 observations and 22 variables in the dataset.

So, over all leverage or influential points is not a big problem only few of them (however all points are belonging to those team that with medium or bad performance, why?)

As I mentioned before, this data does NOT contain any outlier that need to be deleted, the outliers appear in the chart show the prediction equation did not represent the team performance very well since in sports one team with lower statistics can still win the game because it is sports anything can happen!

Mixed Transformation summary:

After executing this mixed transformation, I found that

1. CV and R2 did not change too much
2. Residuals vs Predicted plots still show obvious parent.
3. Some leverage points exist but not big deal

So, I will go back to those 2 old transformations see their plots performance by taking look at their CV and R2 compare to the original model

|  |  |  |  |
| --- | --- | --- | --- |
|  | cv | R2 | MSE |
| Original | 10.5 | 96.21% | 4.29 |
| (1/x) | 6.859 | 98.97% | 2.8 |
| Sqrt(x) | 8 | 98.6% | 3.27 |

Based on the table above, all CV and R2 are accepted so I will consider their plots: are they satisfy those assumptions.

Transformation 1: (all plots are in the Appendix section)



After reviewing the Residual vs Predicted Value plot and Residual VS all regressor plot, I found that there are still obvious patterns there for example the Residual vs predicted value plot it is looks like a close tunnel. So, transformation 1 was decline by rejecting the constant variance assumptions.

Let’s take look at the second transformation: sqrt(x)



For this time, I found that Residual VS Predicted plot is better than last transformation which is all points are separate by the 0 line equally and no obvious pattern found.

Then I checked the Residual VS each regressors plot and found that for most of them, their plots are become better which means no pattern and equally distributed plots. (full plots are list in the Appendix section)

After review 4 different transformation I will go new x=sqrt (x), because not only its statistics are accepted, and its plots are better then other transformation.

Check its influential points, leverage points and outliers.

| **Output Statistics** | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Obs** | **Dependent Variable** | **Predicted Value** | **Std Error Mean Predict** | **Residual** | **Std Error Residual** | **Student Residual** | -2-1 0 1 2 | **Cook's D** | **RStudent** | **Hat Diag H** |
|
| **1** | 60.0000 | 60.2413 | 2.6753 | -0.2413 | 0.845 | -0.286 | | | | | 0.036 | -0.2660 | 0.9093 |
| **2** | 40.0000 | 39.5080 | 2.6314 | 0.4920 | 0.973 | 0.506 | | |\* | | 0.081 | 0.4769 | 0.8797 |
| **3** | 38.0000 | 39.4886 | 2.5189 | -1.4886 | 1.235 | -1.205 | | \*\*| | | 0.263 | -1.2532 | 0.8061 |
| **4** | 33.0000 | 32.2868 | 2.6881 | 0.7132 | 0.803 | 0.888 | | |\* | | 0.384 | 0.8728 | 0.9181 |
| **5** | 50.0000 | 51.0095 | 2.3453 | -1.0095 | 1.540 | -0.656 | | \*| | | 0.043 | -0.6266 | 0.6988 |
| **6** | 53.0000 | 53.1877 | 2.5665 | -0.1877 | 1.133 | -0.166 | | | | | 0.006 | -0.1537 | 0.8369 |
| **7** | 50.0000 | 47.6259 | 1.9663 | 2.3741 | 2.001 | 1.186 | | |\*\* | | 0.059 | 1.2289 | 0.4912 |
| **8** | 30.0000 | 32.1590 | 1.9205 | -2.1590 | 2.045 | -1.056 | | \*\*| | | 0.043 | -1.0659 | 0.4686 |
| **9** | 32.0000 | 28.8231 | 2.3969 | 3.1769 | 1.458 | 2.179 | | |\*\*\*\* | | 0.558 | 3.5569 | 0.7299 |
| **10** | 67.0000 | 66.8186 | 2.5868 | 0.1814 | 1.086 | 0.167 | | | | | 0.007 | 0.1550 | 0.8502 |
| **11** | 56.0000 | 55.8536 | 2.6593 | 0.1464 | 0.894 | 0.164 | | | | | 0.010 | 0.1519 | 0.8985 |
| **12** | 38.0000 | 35.7937 | 2.3263 | 2.2063 | 1.568 | 1.407 | | |\*\* | | 0.189 | 1.5381 | 0.6876 |
| **13** | 56.0000 | 56.6825 | 2.4832 | -0.6825 | 1.305 | -0.523 | | \*| | | 0.043 | -0.4937 | 0.7835 |
| **14** | 21.0000 | 23.6564 | 2.2338 | -2.6564 | 1.697 | -1.565 | | \*\*\*| | | 0.184 | -1.7970 | 0.6340 |
| **15** | 55.0000 | 54.4938 | 2.6596 | 0.5062 | 0.893 | 0.567 | | |\* | | 0.124 | 0.5373 | 0.8987 |
| **16** | 37.0000 | 35.6041 | 2.2365 | 1.3959 | 1.694 | 0.824 | | |\* | | 0.051 | 0.8030 | 0.6355 |
| **17** | 41.0000 | 39.7464 | 2.1340 | 1.2536 | 1.821 | 0.688 | | |\* | | 0.028 | 0.6600 | 0.5786 |
| **18** | 16.0000 | 16.1820 | 2.7045 | -0.1820 | 0.746 | -0.244 | | | | | 0.034 | -0.2268 | 0.9293 |
| **19** | 45.0000 | 45.3008 | 2.6406 | -0.3008 | 0.948 | -0.317 | | | | | 0.034 | -0.2960 | 0.8859 |
| **20** | 17.0000 | 16.2055 | 2.4599 | 0.7945 | 1.349 | 0.589 | | |\* | | 0.050 | 0.5593 | 0.7688 |
| **21** | 45.0000 | 43.9728 | 2.2444 | 1.0272 | 1.683 | 0.610 | | |\* | | 0.029 | 0.5806 | 0.6400 |
| **22** | 22.0000 | 24.4314 | 2.5355 | -2.4314 | 1.201 | -2.025 | | \*\*\*\*| | | 0.795 | -2.9123 | 0.8168 |
| **23** | 18.0000 | 18.8210 | 2.6922 | -0.8210 | 0.789 | -1.040 | | \*\*| | | 0.548 | -1.0475 | 0.9209 |
| **24** | 39.0000 | 38.1077 | 2.1406 | 0.8923 | 1.813 | 0.492 | | | | | 0.015 | 0.4637 | 0.5822 |
| **25** | 51.0000 | 51.4957 | 2.5757 | -0.4957 | 1.112 | -0.446 | | | | | 0.046 | -0.4187 | 0.8429 |
| **26** | 29.0000 | 28.1900 | 2.6615 | 0.8100 | 0.887 | 0.913 | | |\* | | 0.326 | 0.9006 | 0.9000 |
| **27** | 55.0000 | 56.6211 | 2.5721 | -1.6211 | 1.120 | -1.447 | | \*\*| | | 0.480 | -1.6005 | 0.8406 |
| **28** | 49.0000 | 49.8019 | 2.3999 | -0.8019 | 1.453 | -0.552 | | \*| | | 0.036 | -0.5225 | 0.7318 |
| **29** | 38.0000 | 39.0127 | 2.6075 | -1.0127 | 1.035 | -0.978 | | \*| | | 0.264 | -0.9747 | 0.8638 |
| **30** | 46.0000 | 45.8787 | 2.1215 | 0.1213 | 1.836 | 0.0661 | | | | | 0.000 | 0.0612 | 0.5718 |

| **Output Statistics** | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Obs** | **Cov Ratio** | **DFFITS** | **DFBETAS** | | | | | | | | | | | |
| **Intercept** | **MP1** | **FG1** | **FGA1** | **FGP1** | **P31** | **P3A1** | **P3P1** | **P21** | **P2a1** | **P2p1** | **FT1** |
| **1** | 291.8610 | -0.8425 | 0.1834 | -0.2372 | -0.0033 | 0.1086 | 0.2452 | 0.3715 | -0.3604 | -0.3293 | 0.1420 | -0.2658 | -0.2986 | 0.0721 |
| **2** | 122.4374 | 1.2898 | -0.1722 | 0.4367 | 0.1845 | -0.0942 | 0.0113 | -0.1577 | 0.1419 | 0.0212 | -0.2893 | 0.1404 | -0.0082 | -0.1433 |
| **3** | 0.8508 | -2.5556 | -0.3649 | 0.8189 | 0.3144 | 0.2517 | 0.9269 | 0.2188 | -0.1814 | -0.8488 | -0.6729 | -0.1418 | -0.7982 | 0.0655 |
| **4** | 27.0510 | 2.9215 | -0.8201 | -0.2281 | -0.4217 | 0.6114 | 0.2867 | 0.1202 | -0.1080 | 0.1057 | -0.8632 | 0.7088 | 0.4215 | 0.5935 |
| **5** | 26.7795 | -0.9545 | -0.3067 | -0.0523 | -0.3196 | 0.3351 | 0.2756 | 0.1997 | -0.1977 | -0.1508 | 0.0188 | -0.0759 | -0.0928 | -0.2003 |
| **6** | 194.1467 | -0.3481 | 0.0661 | -0.0015 | 0.1714 | -0.1473 | -0.1517 | -0.1472 | 0.1571 | 0.1289 | -0.0524 | 0.0809 | 0.0890 | -0.0166 |
| **7** | 0.3898 | 1.2075 | 0.5809 | -0.0711 | 0.5699 | -0.5052 | -0.0817 | 0.0326 | -0.0324 | -0.2407 | -0.0211 | -0.1551 | -0.3007 | -0.1467 |
| **8** | 1.2085 | -1.0009 | 0.2964 | 0.0239 | 0.3737 | -0.3302 | -0.1081 | 0.0117 | -0.0067 | -0.1230 | -0.1211 | 0.0319 | -0.0832 | -0.0582 |
| **9** | 0.0000 | 5.8477 | -0.1522 | -2.2990 | -0.4501 | 0.8903 | -0.0849 | 0.0118 | 0.0052 | 0.5520 | 0.0196 | 0.2238 | 0.6846 | 1.2974 |
| **10** | 211.0063 | 0.3691 | -0.0123 | -0.0023 | -0.0588 | 0.0620 | 0.0164 | 0.0866 | -0.0879 | -0.0166 | 0.0502 | -0.0305 | 0.0102 | 0.0679 |
| **11** | 312.5383 | 0.4520 | -0.0148 | -0.0350 | 0.0158 | -0.0126 | -0.0121 | -0.0310 | 0.0466 | 0.0310 | -0.0825 | 0.0910 | 0.0399 | 0.0779 |
| **12** | 0.0531 | 2.2817 | 1.0234 | 0.4942 | 0.2484 | -0.5608 | 0.0576 | 0.3955 | -0.4089 | -0.6113 | 0.7531 | -0.7354 | -0.7165 | 0.1072 |
| **13** | 64.0382 | -0.9392 | -0.0880 | -0.0459 | -0.2201 | 0.2080 | 0.1750 | 0.0647 | -0.0635 | -0.0575 | 0.0239 | -0.0424 | -0.0562 | 0.2683 |
| **14** | 0.0047 | -2.3649 | 0.3751 | -0.2693 | -0.2889 | 0.3210 | 0.0984 | 0.2679 | -0.1759 | 0.0956 | 0.9414 | -0.6264 | -0.2291 | 1.0287 |
| **15** | 116.0600 | 1.6005 | 0.0792 | 0.1072 | -0.1820 | 0.0627 | 0.0064 | 0.0514 | -0.0189 | 0.0043 | 0.5489 | -0.3534 | -0.1264 | -0.1338 |
| **16** | 9.0902 | 1.0602 | -0.2177 | 0.1716 | -0.2673 | 0.3060 | 0.3583 | 0.2301 | -0.2424 | -0.2532 | 0.0814 | -0.2128 | -0.2560 | -0.0661 |
| **17** | 16.4079 | 0.7733 | -0.2072 | -0.1125 | -0.0446 | 0.0575 | -0.0644 | -0.1280 | 0.1051 | 0.1308 | -0.2799 | 0.2640 | 0.2174 | -0.0918 |
| **18** | 402.7655 | -0.8223 | -0.2270 | -0.0686 | -0.1561 | 0.1583 | -0.0278 | -0.1450 | 0.1353 | 0.1784 | -0.0318 | 0.0969 | 0.1674 | 0.0051 |
| **19** | 217.6663 | -0.8249 | 0.2764 | -0.3275 | 0.1405 | -0.1183 | -0.1160 | 0.1859 | -0.2041 | -0.0622 | 0.0017 | 0.0130 | 0.0360 | 0.0666 |
| **20** | 46.5644 | 1.0200 | -0.0392 | -0.2233 | -0.0330 | 0.0675 | -0.0628 | 0.0676 | -0.0629 | 0.0710 | 0.0141 | 0.0502 | 0.1190 | -0.0112 |
| **21** | 27.3784 | 0.7742 | -0.1103 | 0.1799 | -0.2829 | 0.1898 | 0.1550 | 0.1212 | -0.1279 | -0.0824 | 0.1597 | -0.1289 | -0.0870 | 0.0525 |
| **22** | 0.0000 | -6.1492 | -0.7081 | -0.4616 | -0.2279 | -0.2658 | -0.7803 | -0.6725 | 0.5088 | 0.8074 | -0.1673 | 0.5669 | 0.8687 | -2.9640 |
| **23** | 9.2009 | -3.5732 | -0.5600 | -0.1541 | 0.0324 | 0.0173 | -0.3765 | -0.3203 | 0.3056 | 0.5120 | -1.2846 | 1.2768 | 0.8918 | 0.4110 |
| **24** | 36.9107 | 0.5473 | 0.0322 | 0.0443 | 0.1706 | -0.2106 | -0.1963 | -0.1918 | 0.1821 | 0.1413 | -0.1683 | 0.1882 | 0.1505 | -0.1345 |
| **25** | 113.7389 | -0.9699 | 0.0513 | -0.0059 | -0.1176 | 0.0653 | 0.0271 | 0.0537 | -0.0637 | -0.0078 | -0.0239 | 0.0427 | 0.0329 | 0.4359 |
| **26** | 18.7621 | 2.7019 | -0.6344 | -0.3097 | 0.0449 | 0.3642 | -0.0208 | -0.2790 | 0.2944 | 0.3593 | -0.4231 | 0.3202 | 0.3912 | -0.1571 |
| **27** | 0.0611 | -3.6750 | -1.0767 | 1.9837 | 0.2102 | -0.9505 | -1.9648 | -2.0599 | 2.0365 | 2.2215 | -0.5387 | 1.6172 | 2.0962 | -0.7241 |
| **28** | 46.4424 | -0.8630 | -0.0631 | 0.1867 | 0.0580 | -0.0977 | 0.0159 | 0.0461 | -0.0637 | -0.0980 | -0.1519 | 0.0676 | -0.0370 | -0.3409 |
| **29** | 8.6568 | -2.4551 | -0.2506 | -0.2344 | -0.9568 | 0.8252 | 0.7494 | 0.3180 | -0.2650 | -0.3079 | 0.8077 | -0.7277 | -0.4878 | 0.4509 |
| **30** | 79.7852 | 0.0707 | 0.0061 | -0.0113 | -0.0000 | -0.0050 | -0.0084 | -0.0123 | 0.0141 | 0.0117 | 0.0023 | 0.0053 | 0.0067 | 0.0109 |

P=number of xi+1 =22+1=23

N=30

Hii=2p/n=1.53 (leverage point): None

Cook’s D: >1: None

|DFFITS| > 2/sqrt(p/n) =2.28: 3,4,9,12,14,22,23,26,27,29

|DFBETAS|>2/sqrt(n)=0.365:

Following obs has large number of times that exceeds the range:

4,14, 9,27,29

COVRATIO: 1±3p/n= (-1.3,3.3): No 3 No7 No8 No9 No 12 No14 No22 No27

|RSTUDENT|>= 3: None

They are total 30 observations and 22 variables in the dataset. After review those points I found that there is less influential or leverage points in this dataset. For Cook’s D, Hii and R-student, no obs violate these 3 criteria, for DFBETAS and DFFITS and COVRSTIO only few obs violate them so overall

After few attempts of different transformation, the leverage points and influential points still exists.

So far, I decided my final model by resolve CV, R2, plots, assumptions, leverage points, influential points, and outliers and linear combination problems. However, after I check the model I selected I found there is one more problem exist which is multicollinearity.

VIF: 0 6.19 77k 31k 15k 250k 198k 3k 143k 128k 7k 13k 19k 5.55 24

6.77 10.31 6.8 17k.

| **Collinearity Diagnostics (intercept adjusted)** | | | | | | | | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Number** | | | **Eigenvalue** | | **Condition Index** | | | **Proportion of Variation** | | | | | | | | | | | | | | | | | |
| **MP2** | | | **FG2** | | | **FGA2** | | | **FGP2** | | | **P32** | | **P3A2** | | **P3P2** | |
| **1** | | | 5.93898 | | 1.00000 | | | 0.00026679 | | | 2.6063E-7 | | | 9.035415E-8 | | | 0.00000111 | | | 6.402948E-8 | | 5.24683E-8 | | 0.00000539 | |
| **2** | | | 4.15963 | | 1.19489 | | | 0.00348 | | | 1.392728E-7 | | | 2.258044E-7 | | | 2.136448E-7 | | | 6.512527E-8 | | 1.058873E-7 | | 1.486485E-8 | |
| **3** | | | 3.12177 | | 1.37929 | | | 0.00085338 | | | 1.381696E-8 | | | 0.00000140 | | | 0.00000118 | | | 3.162891E-8 | | 7.411853E-8 | | 0.00000153 | |
| **4** | | | 2.22285 | | 1.63456 | | | 0.00068360 | | | 4.952764E-9 | | | 8.395587E-7 | | | 8.37623E-7 | | | 8.66391E-9 | | 1.726495E-8 | | 3.39253E-8 | |
| **5** | | | 1.92463 | | 1.75664 | | | 0.00173 | | | 1.943858E-7 | | | 4.270368E-7 | | | 2.224479E-7 | | | 1.314844E-8 | | 6.788291E-9 | | 0.00000228 | |
| **6** | | | 1.26955 | | 2.16288 | | | 0.00148 | | | 1.602273E-8 | | | 0.00000152 | | | 0.00000120 | | | 9.57984E-8 | | 2.951262E-7 | | 0.00001913 | |
| **7** | | | 0.77135 | | 2.77478 | | | 0.05271 | | | 3.024362E-7 | | | 0.00000150 | | | 7.45802E-9 | | | 2.52117E-8 | | 4.561162E-8 | | 6.64033E-10 | |
| **8** | | | 0.73201 | | 2.84837 | | | 0.00492 | | | 1.68026E-8 | | | 0.00000167 | | | 0.00000367 | | | 1.907665E-8 | | 1.004913E-8 | | 0.00000237 | |
| **9** | | | 0.55547 | | 3.26983 | | | 0.05201 | | | 1.705436E-7 | | | 5.673724E-8 | | | 0.00000186 | | | 2.171494E-9 | | 7.38857E-10 | | 0.00000653 | |
| **10** | | | 0.43993 | | 3.67420 | | | 0.00551 | | | 1.974003E-8 | | | 3.276817E-9 | | | 1.442166E-7 | | | 4.914485E-8 | | 1.44632E-8 | | 0.00002439 | |
| **11** | | | 0.40094 | | 3.84871 | | | 0.05604 | | | 3.106602E-7 | | | 4.280417E-7 | | | 0.00000581 | | | 9.197327E-9 | | 5.863264E-8 | | 0.00008298 | |
| **12** | | | 0.30138 | | 4.43913 | | | 0.12112 | | | 4.049106E-8 | | | 0.00000450 | | | 0.00000363 | | | 3.343266E-8 | | 4.422375E-8 | | 0.00018475 | |
| **13** | | | 0.09383 | | 7.95592 | | | 0.01195 | | | 0.00000217 | | | 0.00002233 | | | 0.00000322 | | | 4.834158E-7 | | 4.969898E-8 | | 0.00069130 | |
| **14** | | | 0.06374 | | 9.65238 | | | 0.00227 | | | 0.00000348 | | | 0.00002012 | | | 1.928497E-7 | | | 8.914772E-8 | | 0.00000117 | | 0.00028475 | |
| **15** | | | 0.00317 | | 43.31096 | | | 0.00319 | | | 0.00003259 | | | 0.00117 | | | 0.00142 | | | 0.00018507 | | 0.00040466 | | 0.00096755 | |
| **16** | | | 0.00048047 | | 111.17876 | | | 0.09336 | | | 0.00274 | | | 0.00453 | | | 0.00656 | | | 0.00000679 | | 0.00093865 | | 0.05746 | |
| **17** | | | 0.00011362 | | 228.63177 | | | 0.09146 | | | 0.00294 | | | 0.00732 | | | 0.06098 | | | 0.00518 | | 0.00282 | | 0.03746 | |
| **18** | | | 0.00007790 | | 276.11003 | | | 0.05280 | | | 0.00824 | | | 0.00327 | | | 0.01593 | | | 0.00010250 | | 0.00023442 | | 0.00005428 | |
| **19** | | | 0.00003645 | | 403.63772 | | | 0.01109 | | | 0.08693 | | | 0.09478 | | | 0.41516 | | | 0.01232 | | 0.00455 | | 0.00252 | |
| **20** | | | 0.00002658 | | 472.72632 | | | 0.07937 | | | 0.06941 | | | 0.06388 | | | 0.31445 | | | 0.00038649 | | 0.00000677 | | 0.01354 | |
| **21** | | | 0.00002272 | | 511.22508 | | | 0.17618 | | | 0.04253 | | | 0.02660 | | | 0.14151 | | | 0.00375 | | 0.00012604 | | 0.00315 | |
| **22** | | | 0.00000125 | | 2183.47179 | | | 0.17755 | | | 0.78718 | | | 0.79840 | | | 0.04395 | | | 0.97807 | | 0.99092 | | 0.88354 | |
| **Collinearity Diagnostics (intercept adjusted)** | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| **Number** | | | **Proportion of Variation** | | | | | | | | | | | | | | | | | | | | | | | | |
| **P22** | | | | **P2a2** | | | **P2p2** | | | **FT2** | | | **FTA2** | | | **FTP2** | | | **ORB2** | | **DRB2** | | **TRB2** | |
| **1** | | | 6.61517E-10 | | | | 4.198047E-8 | | | 0.00000272 | | | 3.522568E-8 | | | 1.387566E-7 | | | 0.00000152 | | | 4.927464E-7 | | 3.567253E-7 | | 3.509963E-8 | |
| **2** | | | 3.262326E-7 | | | | 3.140692E-7 | | | 5.682805E-8 | | | 3.807812E-9 | | | 2.560098E-7 | | | 0.00001651 | | | 1.10575E-7 | | 7.723426E-7 | | 2.329686E-7 | |
| **3** | | | 1.809741E-8 | | | | 3.38663E-12 | | | 7.976347E-7 | | | 2.314301E-7 | | | 7.425699E-7 | | | 0.00000503 | | | 0.00000990 | | 0.00000246 | | 0.00000405 | |
| **4** | | | 1.742243E-9 | | | | 1.592125E-8 | | | 0.00000156 | | | 0.00001534 | | | 0.00001666 | | | 0.00000588 | | | 1.405597E-7 | | 3.293378E-9 | | 2.251934E-8 | |
| **5** | | | 1.704973E-7 | | | | 6.643175E-8 | | | 8.231761E-7 | | | 3.111451E-7 | | | 9.840285E-7 | | | 0.00000749 | | | 0.00000170 | | 0.00000179 | | 2.597019E-7 | |
| **6** | | | 8.747802E-8 | | | | 1.134902E-7 | | | 2.784127E-7 | | | 2.018674E-7 | | | 1.23016E-8 | | | 0.00000860 | | | 0.00000127 | | 0.00000230 | | 0.00000187 | |
| **7** | | | 3.268983E-7 | | | | 4.111706E-7 | | | 3.993969E-7 | | | 7.780282E-8 | | | 0.00000162 | | | 0.00014650 | | | 0.00001401 | | 0.00000982 | | 8.572265E-7 | |
| **8** | | | 9.903115E-8 | | | | 3.066914E-8 | | | 0.00001134 | | | 0.00000366 | | | 1.709607E-8 | | | 0.00028177 | | | 0.00001395 | | 0.00001088 | | 0.00000111 | |
| **9** | | | 6.095112E-8 | | | | 1.53479E-11 | | | 0.00000229 | | | 0.00000302 | | | 0.00001321 | | | 0.00013684 | | | 0.00006012 | | 0.00000175 | | 0.00000206 | |
| **10** | | | 3.6004E-10 | | | | 1.62599E-10 | | | 4.10882E-10 | | | 1.426422E-7 | | | 0.00000150 | | | 0.00015005 | | | 4.003902E-8 | | 0.00001196 | | 0.00000537 | |
| **11** | | | 2.824961E-7 | | | | 2.06839E-7 | | | 0.00004317 | | | 0.00000266 | | | 0.00000180 | | | 0.00001453 | | | 0.00000717 | | 9.512546E-7 | | 1.380964E-8 | |
| **12** | | | 1.409622E-7 | | | | 1.691751E-7 | | | 9.305345E-8 | | | 0.00000180 | | | 5.286253E-7 | | | 0.00026934 | | | 0.00009930 | | 0.00001822 | | 2.117071E-8 | |
| **13** | | | 2.390025E-8 | | | | 0.00000179 | | | 0.00010532 | | | 1.973651E-7 | | | 0.00001226 | | | 0.00047183 | | | 0.00016757 | | 0.00000169 | | 0.00000932 | |
| **14** | | | 0.00000225 | | | | 3.781514E-7 | | | 0.00002184 | | | 6.721484E-7 | | | 0.00000167 | | | 0.00034562 | | | 0.00003271 | | 0.00005597 | | 0.00004706 | |
| **15** | | | 0.00028492 | | | | 0.00024429 | | | 0.00615 | | | 0.00012942 | | | 3.945897E-7 | | | 0.00004019 | | | 1.353046E-7 | | 3.077703E-7 | | 0.00000740 | |
| **16** | | | 0.00111 | | | | 0.00029473 | | | 0.06456 | | | 0.00433 | | | 0.00063149 | | | 0.00069395 | | | 0.00058266 | | 0.00004926 | | 0.00020932 | |
| **17** | | | 0.01446 | | | | 0.02069 | | | 0.05327 | | | 0.00235 | | | 0.00903 | | | 0.00918 | | | 0.00001276 | | 0.00005602 | | 0.00002335 | |
| **18** | | | 0.00006834 | | | | 0.00261 | | | 0.02576 | | | 0.25604 | | | 0.52833 | | | 0.53465 | | | 0.03622 | | 0.03620 | | 0.03441 | |
| **19** | | | 4.198083E-7 | | | | 0.00416 | | | 0.08353 | | | 0.04981 | | | 0.02020 | | | 0.02196 | | | 0.12324 | | 0.11742 | | 0.12443 | |
| **20** | | | 0.00492 | | | | 0.01450 | | | 0.20059 | | | 0.60753 | | | 0.19787 | | | 0.20463 | | | 0.00441 | | 0.00486 | | 0.00653 | |
| **21** | | | 0.00031945 | | | | 0.00118 | | | 0.02905 | | | 0.07442 | | | 0.23323 | | | 0.21866 | | | 0.82416 | | 0.82949 | | 0.82475 | |
| **22** | | | 0.97883 | | | | 0.95632 | | | 0.53690 | | | 0.00536 | | | 0.01066 | | | 0.00834 | | | 0.01097 | | 0.01181 | | 0.00956 | |
| **Collinearity Diagnostics (intercept adjusted)** | | | | | | | | | | | | | | | | | | | | |
| **Number** | | | **Proportion of Variation** | | | | | | | | | | | | | | | | | |
| **AST2** | | | **STL2** | | | **BLK2** | | | **TOV2** | | | **PF2** | | | **PTS2** | | |
| **1** | | | 0.00302 | | | 0.00009792 | | | 0.00008051 | | | 0.00014879 | | | 0.00008741 | | | 0.00000122 | | |
| **2** | | | 0.00013615 | | | 0.00051097 | | | 0.00060857 | | | 0.00138 | | | 0.00026049 | | | 5.590792E-8 | | |
| **3** | | | 0.00028514 | | | 0.00019280 | | | 0.00158 | | | 0.00000371 | | | 0.00116 | | | 3.403927E-7 | | |
| **4** | | | 0.00163 | | | 0.00036771 | | | 0.00001460 | | | 0.00000191 | | | 0.00056973 | | | 0.00000111 | | |
| **5** | | | 0.00092759 | | | 0.00452 | | | 0.00595 | | | 0.00570 | | | 0.02084 | | | 5.340404E-7 | | |
| **6** | | | 0.00458 | | | 0.00085047 | | | 0.03834 | | | 0.01259 | | | 0.00001206 | | | 6.008158E-7 | | |
| **7** | | | 0.00040275 | | | 0.00036269 | | | 0.00573 | | | 0.02641 | | | 0.00411 | | | 2.717605E-7 | | |
| **8** | | | 0.01415 | | | 0.00509 | | | 0.00514 | | | 0.00199 | | | 0.02949 | | | 4.511176E-7 | | |
| **9** | | | 0.04753 | | | 0.00253 | | | 0.02033 | | | 0.01717 | | | 0.00025035 | | | 4.005529E-8 | | |
| **10** | | | 0.02829 | | | 0.00397 | | | 0.07798 | | | 0.01148 | | | 0.11317 | | | 1.834833E-7 | | |
| **11** | | | 0.09222 | | | 0.00779 | | | 0.03078 | | | 0.00046084 | | | 0.04239 | | | 4.643644E-9 | | |
| **12** | | | 0.01598 | | | 0.00815 | | | 0.02212 | | | 0.00246 | | | 0.00326 | | | 3.183561E-7 | | |
| **13** | | | 0.54726 | | | 0.00388 | | | 0.00388 | | | 0.04328 | | | 0.04779 | | | 0.00001599 | | |
| **14** | | | 0.10351 | | | 0.22601 | | | 0.09615 | | | 0.34382 | | | 0.04718 | | | 0.00000955 | | |
| **15** | | | 0.00094317 | | | 0.02305 | | | 0.04418 | | | 0.00105 | | | 0.00196 | | | 0.00014470 | | |
| **16** | | | 0.01670 | | | 0.05907 | | | 0.08753 | | | 0.01930 | | | 0.20709 | | | 0.02684 | | |
| **17** | | | 0.00682 | | | 0.02829 | | | 0.00150 | | | 0.07475 | | | 0.00154 | | | 0.00770 | | |
| **18** | | | 0.01037 | | | 0.03348 | | | 0.00327 | | | 0.06179 | | | 0.00746 | | | 0.07347 | | |
| **19** | | | 0.00000981 | | | 0.08877 | | | 0.12861 | | | 0.01036 | | | 0.05179 | | | 0.03633 | | |
| **20** | | | 0.00000126 | | | 0.35851 | | | 0.04873 | | | 0.34251 | | | 0.03467 | | | 0.65264 | | |
| **21** | | | 0.03603 | | | 0.14350 | | | 0.21483 | | | 0.00520 | | | 0.36383 | | | 0.08200 | | |
| **22** | | | 0.06921 | | | 0.00099049 | | | 0.16266 | | | 0.01814 | | | 0.02108 | | | 0.12085 | | |

By checking the VIF, eigenvalues, condition index and Collinearity Diagnostics I found that multicollinearity is a series problem because the few specific VIF value is larger 1000 and in the last line of Collinearity Diagnostics, some specific value is larger than 0.5 which indicate serve multicollinearity problem as well.

By my personal experience and data description the multicollinearity is easily observed because some variables are strongly connected with each other.

1. FG=P3+P2
2. FGP=FG/FGA
3. FGA=P3A+P2A
4. P3P=P3/P3A
5. P2P=P2/P2A
6. TRB=DRB+ORB

This is also explained why there are some linear combination in the original regression model I will use ridge and principal component analysis (both biased) to address multicollinearity problem.



I set the ridge parameter from 0.02 to 2, from the plot above it is clear that when the parameter consistently drops the VIF will drop at the same time. So, I will go with parameter equals 0.02 the minimum value in the plot which indicate the minimum VIF.

The following table describes the detailed information after performing the ridge regression from parameter set to 0.02 to 0.2.

| **Obs** | **\_MODEL\_** | **\_TYPE\_** | **\_DEPVAR\_** | **\_RIDGE\_** | **\_PCOMIT\_** | **\_RMSE\_** | **Intercept** | **MP2** | **FG2** | **FGA2** | **FGP2** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | MODEL1 | PARMS | Games\_Won | . | . | 3.27272 | 87.03 | 1.29613 | 274.88 | -51.73 | 2769.09 |
| **2** | MODEL1 | RIDGEVIF | Games\_Won | 0.00 | . | . | . | 6.19841 | 77874.95 | 31760.12 | 15367.59 |
| **3** | MODEL1 | RIDGE | Games\_Won | 0.00 | . | 3.27272 | 87.03 | 1.29613 | 274.88 | -51.73 | 2769.09 |
| **4** | MODEL1 | RIDGESTB | Games\_Won | 0.00 | . | 3.27272 | 0.00 | 0.02462 | 23.86 | -4.51 | 2.37 |
| **5** | MODEL1 | RIDGEVIF | Games\_Won | 0.02 | . | . | . | 1.75716 | 0.60 | 2.02 | 0.85 |
| **6** | MODEL1 | RIDGE | Games\_Won | 0.02 | . | 6.34519 | -1350.58 | 6.68401 | 0.26 | -2.58 | 182.31 |
| **7** | MODEL1 | RIDGESTB | Games\_Won | 0.02 | . | 6.34519 | 0.00 | 0.12696 | 0.02 | -0.23 | 0.16 |
| **8** | MODEL1 | RIDGEVIF | Games\_Won | 0.04 | . | . | . | 1.59704 | 0.35 | 1.17 | 0.44 |
| **9** | MODEL1 | RIDGE | Games\_Won | 0.04 | . | 6.72222 | -1362.65 | 6.72569 | 0.20 | -2.44 | 200.99 |
| **10** | MODEL1 | RIDGESTB | Games\_Won | 0.04 | . | 6.72222 | 0.00 | 0.12775 | 0.02 | -0.21 | 0.17 |
| **11** | MODEL1 | RIDGEVIF | Games\_Won | 0.06 | . | . | . | 1.46276 | 0.26 | 0.87 | 0.34 |
| **12** | MODEL1 | RIDGE | Games\_Won | 0.06 | . | 7.04758 | -1355.40 | 6.65186 | 0.27 | -2.23 | 203.80 |
| **13** | MODEL1 | RIDGESTB | Games\_Won | 0.06 | . | 7.04758 | 0.00 | 0.12635 | 0.02 | -0.19 | 0.17 |
| **14** | MODEL1 | RIDGEVIF | Games\_Won | 0.08 | . | . | . | 1.34767 | 0.21 | 0.71 | 0.29 |
| **15** | MODEL1 | RIDGE | Games\_Won | 0.08 | . | 7.32963 | -1339.18 | 6.52122 | 0.34 | -2.04 | 202.94 |
| **16** | MODEL1 | RIDGESTB | Games\_Won | 0.08 | . | 7.32963 | 0.00 | 0.12387 | 0.03 | -0.18 | 0.17 |
| **17** | MODEL1 | RIDGEVIF | Games\_Won | 0.10 | . | . | . | 1.24766 | 0.18 | 0.60 | 0.26 |
| **18** | MODEL1 | RIDGE | Games\_Won | 0.10 | . | 7.57519 | -1318.20 | 6.36090 | 0.41 | -1.88 | 200.95 |
| **19** | MODEL1 | RIDGESTB | Games\_Won | 0.10 | . | 7.57519 | 0.00 | 0.12082 | 0.04 | -0.16 | 0.17 |
| **20** | MODEL1 | RIDGEVIF | Games\_Won | 0.12 | . | . | . | 1.15984 | 0.16 | 0.53 | 0.24 |
| **21** | MODEL1 | RIDGE | Games\_Won | 0.12 | . | 7.79115 | -1294.70 | 6.18555 | 0.47 | -1.74 | 198.60 |
| **22** | MODEL1 | RIDGESTB | Games\_Won | 0.12 | . | 7.79115 | 0.00 | 0.11749 | 0.04 | -0.15 | 0.17 |
| **23** | MODEL1 | RIDGEVIF | Games\_Won | 0.14 | . | . | . | 1.08209 | 0.15 | 0.47 | 0.22 |
| **24** | MODEL1 | RIDGE | Games\_Won | 0.14 | . | 7.98331 | -1269.97 | 6.00366 | 0.51 | -1.63 | 196.17 |
| **25** | MODEL1 | RIDGESTB | Games\_Won | 0.14 | . | 7.98331 | 0.00 | 0.11404 | 0.04 | -0.14 | 0.17 |
| **26** | MODEL1 | RIDGEVIF | Games\_Won | 0.16 | . | . | . | 1.01277 | 0.13 | 0.42 | 0.21 |
| **27** | MODEL1 | RIDGE | Games\_Won | 0.16 | . | 8.15624 | -1244.79 | 5.82030 | 0.55 | -1.53 | 193.79 |
| **28** | MODEL1 | RIDGESTB | Games\_Won | 0.16 | . | 8.15624 | 0.00 | 0.11056 | 0.05 | -0.13 | 0.17 |
| **29** | MODEL1 | RIDGEVIF | Games\_Won | 0.18 | . | . | . | 0.95060 | 0.12 | 0.39 | 0.20 |
| **30** | MODEL1 | RIDGE | Games\_Won | 0.18 | . | 8.31348 | -1219.64 | 5.63858 | 0.59 | -1.44 | 191.49 |
| **31** | MODEL1 | RIDGESTB | Games\_Won | 0.18 | . | 8.31348 | 0.00 | 0.10710 | 0.05 | -0.13 | 0.16 |
| **32** | MODEL1 | RIDGEVIF | Games\_Won | 0.20 | . | . | . | 0.89456 | 0.11 | 0.36 | 0.19 |
| **33** | MODEL1 | RIDGE | Games\_Won | 0.20 | . | 8.45775 | -1194.82 | 5.46043 | 0.62 | -1.36 | 189.29 |
| **34** | MODEL1 | RIDGESTB | Games\_Won | 0.20 | . | 8.45775 | 0.00 | 0.10372 | 0.05 | -0.12 | 0.16 |

| **Obs** | **P32** | **P3A2** | **P3P2** | **P22** | **P2a2** | **P2p2** | **FT2** | **FTA2** | **FTP2** | **ORB2** | **DRB2** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | -108.06 | 55.63 | 1333.29 | -160.56 | 38.48 | -1835.15 | 80.29 | -42.11 | -1839.58 | 55.67 | 98.14 |
| **2** | 250885.48 | 198031.39 | 3445.58 | 143620.25 | 128992.02 | 7631.62 | 11845.44 | 9229.45 | 1248.80 | 5848.08 | 13004.02 |
| **3** | -108.06 | 55.63 | 1333.29 | -160.56 | 38.48 | -1835.15 | 80.29 | -42.11 | -1839.58 | 55.67 | 98.14 |
| **4** | -21.23 | 16.66 | 1.52 | -17.29 | 6.78 | -1.82 | 10.77 | -6.51 | -2.13 | 5.59 | 8.42 |
| **5** | 1.12 | 1.90 | 3.38 | 1.32 | 1.06 | 2.26 | 0.44 | 0.50 | 1.87 | 1.90 | 1.20 |
| **6** | -0.43 | -0.32 | -37.57 | -1.47 | -1.83 | 351.09 | 0.37 | -0.34 | 202.24 | 1.02 | 2.96 |
| **7** | -0.08 | -0.10 | -0.04 | -0.16 | -0.32 | 0.35 | 0.05 | -0.05 | 0.23 | 0.10 | 0.25 |
| **8** | 0.43 | 0.69 | 2.51 | 0.58 | 0.49 | 1.20 | 0.33 | 0.40 | 1.58 | 1.58 | 0.98 |
| **9** | -0.14 | -0.09 | -32.42 | -0.94 | -1.46 | 313.90 | 0.36 | -0.21 | 176.52 | 0.66 | 2.62 |
| **10** | -0.03 | -0.03 | -0.04 | -0.10 | -0.26 | 0.31 | 0.05 | -0.03 | 0.20 | 0.07 | 0.22 |
| **11** | 0.27 | 0.41 | 2.01 | 0.39 | 0.34 | 0.88 | 0.31 | 0.36 | 1.38 | 1.36 | 0.84 |
| **12** | -0.01 | 0.01 | -25.26 | -0.71 | -1.29 | 293.42 | 0.36 | -0.14 | 157.72 | 0.42 | 2.41 |
| **13** | -0.00 | 0.00 | -0.03 | -0.08 | -0.23 | 0.29 | 0.05 | -0.02 | 0.18 | 0.04 | 0.21 |
| **14** | 0.20 | 0.30 | 1.67 | 0.31 | 0.28 | 0.71 | 0.29 | 0.33 | 1.23 | 1.19 | 0.75 |
| **15** | 0.07 | 0.06 | -18.08 | -0.58 | -1.18 | 278.85 | 0.36 | -0.10 | 143.69 | 0.25 | 2.27 |
| **16** | 0.01 | 0.02 | -0.02 | -0.06 | -0.21 | 0.28 | 0.05 | -0.01 | 0.17 | 0.03 | 0.19 |
| **17** | 0.17 | 0.24 | 1.42 | 0.27 | 0.24 | 0.61 | 0.28 | 0.31 | 1.12 | 1.07 | 0.68 |
| **18** | 0.12 | 0.09 | -11.39 | -0.49 | -1.10 | 267.40 | 0.36 | -0.06 | 132.87 | 0.12 | 2.16 |
| **19** | 0.02 | 0.03 | -0.01 | -0.05 | -0.19 | 0.27 | 0.05 | -0.01 | 0.15 | 0.01 | 0.19 |
| **20** | 0.14 | 0.20 | 1.24 | 0.24 | 0.21 | 0.53 | 0.26 | 0.30 | 1.03 | 0.97 | 0.62 |
| **21** | 0.16 | 0.12 | -5.30 | -0.42 | -1.04 | 257.90 | 0.35 | -0.04 | 124.30 | 0.02 | 2.07 |
| **22** | 0.03 | 0.03 | -0.01 | -0.05 | -0.18 | 0.26 | 0.05 | -0.01 | 0.14 | 0.00 | 0.18 |
| **23** | 0.13 | 0.18 | 1.09 | 0.21 | 0.19 | 0.47 | 0.25 | 0.28 | 0.95 | 0.88 | 0.58 |
| **24** | 0.19 | 0.13 | 0.23 | -0.37 | -0.99 | 249.79 | 0.35 | -0.02 | 117.33 | -0.06 | 2.00 |
| **25** | 0.04 | 0.04 | 0.00 | -0.04 | -0.17 | 0.25 | 0.05 | -0.00 | 0.14 | -0.01 | 0.17 |
| **26** | 0.12 | 0.16 | 0.97 | 0.20 | 0.17 | 0.43 | 0.25 | 0.27 | 0.88 | 0.81 | 0.54 |
| **27** | 0.21 | 0.14 | 5.22 | -0.34 | -0.95 | 242.71 | 0.35 | -0.00 | 111.57 | -0.13 | 1.94 |
| **28** | 0.04 | 0.04 | 0.01 | -0.04 | -0.17 | 0.24 | 0.05 | -0.00 | 0.13 | -0.01 | 0.17 |
| **29** | 0.11 | 0.15 | 0.87 | 0.18 | 0.16 | 0.39 | 0.24 | 0.26 | 0.83 | 0.75 | 0.50 |
| **30** | 0.23 | 0.15 | 9.74 | -0.31 | -0.92 | 236.43 | 0.35 | 0.01 | 106.72 | -0.18 | 1.89 |
| **31** | 0.05 | 0.05 | 0.01 | -0.03 | -0.16 | 0.23 | 0.05 | 0.00 | 0.12 | -0.02 | 0.16 |
| **32** | 0.10 | 0.14 | 0.79 | 0.17 | 0.15 | 0.36 | 0.23 | 0.25 | 0.78 | 0.70 | 0.47 |
| **33** | 0.25 | 0.16 | 13.85 | -0.28 | -0.89 | 230.80 | 0.35 | 0.02 | 102.59 | -0.23 | 1.84 |
| **34** | 0.05 | 0.05 | 0.02 | -0.03 | -0.16 | 0.23 | 0.05 | 0.00 | 0.12 | -0.02 | 0.16 |

| **Obs** | **TRB2** | **AST2** | **STL2** | **BLK2** | **TOV2** | **PF2** | **PTS2** | **Games\_Won** | **\_IN\_** | **\_P\_** | **\_EDF\_** | **\_RSQ\_** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **1** | -108.63 | 1.55466 | 1.6526 | 1.86776 | -3.1586 | 0.19835 | -76.15 | -1 | 22 | 23 | 7 | 0.98600 |
| **2** | 19275.96 | 5.55960 | 24.0748 | 6.77646 | 10.3141 | 6.80785 | 17777.89 | -1 | . | . | . | . |
| **3** | -108.63 | 1.55466 | 1.6526 | 1.86776 | -3.1586 | 0.19835 | -76.15 | -1 | . | . | . | . |
| **4** | -9.87 | 0.20029 | 0.1991 | 0.19811 | -0.3582 | 0.02119 | -10.85 | -1 | . | . | . | . |
| **5** | 0.98 | 3.46857 | 3.9930 | 1.71064 | 3.1357 | 1.76581 | 0.74 | -1 | . | . | . | . |
| **6** | 2.47 | 1.24952 | 3.2200 | 0.99367 | -3.4937 | -0.55569 | 0.17 | -1 | . | . | . | . |
| **7** | 0.22 | 0.16098 | 0.3879 | 0.10540 | -0.3962 | -0.05938 | 0.02 | -1 | . | . | . | . |
| **8** | 0.72 | 2.69737 | 2.8140 | 1.47815 | 2.2943 | 1.54199 | 0.38 | -1 | . | . | . | . |
| **9** | 2.20 | 1.15480 | 2.7415 | 1.08393 | -2.9150 | -0.76274 | 0.20 | -1 | . | . | . | . |
| **10** | 0.20 | 0.14878 | 0.3302 | 0.11497 | -0.3306 | -0.08150 | 0.03 | -1 | . | . | . | . |
| **11** | 0.58 | 2.19595 | 2.1426 | 1.31985 | 1.8055 | 1.38084 | 0.28 | -1 | . | . | . | . |
| **12** | 1.97 | 1.08073 | 2.3973 | 1.14147 | -2.5141 | -0.89035 | 0.27 | -1 | . | . | . | . |
| **13** | 0.18 | 0.13923 | 0.2888 | 0.12108 | -0.2851 | -0.09514 | 0.04 | -1 | . | . | . | . |
| **14** | 0.49 | 1.84606 | 1.7180 | 1.20041 | 1.4914 | 1.25572 | 0.22 | -1 | . | . | . | . |
| **15** | 1.79 | 1.02210 | 2.1400 | 1.17464 | -2.2197 | -0.97376 | 0.33 | -1 | . | . | . | . |
| **16** | 0.16 | 0.13168 | 0.2578 | 0.12459 | -0.2517 | -0.10405 | 0.05 | -1 | . | . | . | . |
| **17** | 0.42 | 1.58886 | 1.4292 | 1.10473 | 1.2744 | 1.15389 | 0.19 | -1 | . | . | . | . |
| **18** | 1.65 | 0.97517 | 1.9397 | 1.19216 | -1.9942 | -1.02966 | 0.38 | -1 | . | . | . | . |
| **19** | 0.15 | 0.12563 | 0.2336 | 0.12645 | -0.2261 | -0.11002 | 0.05 | -1 | . | . | . | . |
| **20** | 0.38 | 1.39210 | 1.2216 | 1.02509 | 1.1161 | 1.06837 | 0.16 | -1 | . | . | . | . |
| **21** | 1.54 | 0.93711 | 1.7786 | 1.19954 | -1.8158 | -1.06742 | 0.42 | -1 | . | . | . | . |
| **22** | 0.14 | 0.12073 | 0.2142 | 0.12724 | -0.2059 | -0.11406 | 0.06 | -1 | . | . | . | . |
| **23** | 0.34 | 1.23680 | 1.0660 | 0.95702 | 0.9955 | 0.99495 | 0.15 | -1 | . | . | . | . |
| **24** | 1.44 | 0.90585 | 1.6458 | 1.20021 | -1.6711 | -1.09272 | 0.45 | -1 | . | . | . | . |
| **25** | 0.13 | 0.11670 | 0.1982 | 0.12731 | -0.1895 | -0.11676 | 0.06 | -1 | . | . | . | . |
| **26** | 0.31 | 1.11110 | 0.9453 | 0.89777 | 0.9004 | 0.93089 | 0.13 | -1 | . | . | . | . |
| **27** | 1.36 | 0.87985 | 1.5340 | 1.19635 | -1.5514 | -1.10921 | 0.48 | -1 | . | . | . | . |
| **28** | 0.12 | 0.11335 | 0.1848 | 0.12690 | -0.1759 | -0.11852 | 0.07 | -1 | . | . | . | . |
| **29** | 0.29 | 1.00726 | 0.8491 | 0.84546 | 0.8235 | 0.87431 | 0.12 | -1 | . | . | . | . |
| **30** | 1.29 | 0.85797 | 1.4382 | 1.18939 | -1.4506 | -1.11932 | 0.51 | -1 | . | . | . | . |
| **31** | 0.12 | 0.11053 | 0.1732 | 0.12616 | -0.1645 | -0.11960 | 0.07 | -1 | . | . | . | . |
| **32** | 0.27 | 0.92001 | 0.7707 | 0.79879 | 0.7597 | 0.82386 | 0.11 | -1 | . | . | . | . |
| **33** | 1.24 | 0.83935 | 1.3551 | 1.18033 | -1.3646 | -1.12473 | 0.53 | -1 | . | . | . | . |
| **34** | 0.11 | 0.10814 | 0.1632 | 0.12520 | -0.1547 | -0.12018 | 0.08 | -1 | . | . | . | . |

PCA

However, PCA is hard to interpreter because it is involved component instead of variables. PCA is better applied when less variables in the dataset. However, there are too many variables in this dataset, so I will try variable selection see if it will help.

1. Selection of Variable Subset

Even though ridge and PCA help dealing the multicollinearity problem, there are still 22 variables in the dataset. My goal for this part is try to find which variables are important when some criteria are satisfied just like Mallow’s Cp, MSE and R2.

First try forward selection.

Since the variable has less observations and more variables I will try forward selection first.

| **Summary of Forward Selection** | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Entered** | **Label** | | | | **Number Vars In** | | **Partial R-Square** | | | **Model R-Square** | | | **C(p)** | | **F Value** | | **Pr > F** |
| **1** | P2p2 |  | | | | 1 | | 0.6412 | | | 0.6412 | | | 153.378 | | 50.04 | | <.0001 |
| **2** | DRB2 |  | | | | 2 | | 0.1093 | | | 0.7505 | | | 100.750 | | 11.82 | | 0.0019 |
| **3** | PF2 |  | | | | 3 | | 0.0495 | | | 0.8000 | | | 77.9957 | | 6.44 | | 0.0175 |
| **4** | BLK2 |  | | | | 4 | | 0.0344 | | | 0.8344 | | | 62.8002 | | 5.19 | | 0.0315 |
| **Analysis of Variance** | | | | | | | | | | | | | | | | | |
| **Source** | | | | **DF** | | **Sum of Squares** | | | | **Mean Square** | | | **F Value** | | | **Pr > F** | |
| **Model** | | | | 4 | | 4467.85524 | | | | 1116.96381 | | | 31.49 | | | <.0001 | |
| **Error** | | | | 25 | | 886.84476 | | | | 35.47379 | | |  | | |  | |
| **Corrected Total** | | | | 29 | | 5354.70000 | | | |  | | |  | | |  | |
|  | | | |  | |  | | | |  | | |  | | |  | |
| **Root MSE** | | | | | | 5.95599 | | | **R-Square** | | | 0.8344 | | |
| **Dependent Mean** | | | | | | 40.90000 | | | **Adj R-Sq** | | | 0.8079 | | |
| **Coeff Var** | | | | | | 14.56232 | | |  | | |  | | |

This is the result from forward selection, after 4 steps, there are only 4 variables left in the model: P2 DRB PF and BLK. However, the R2 and CV are getting worse when majority of variables are removed so this method is not recommended, besides, only 4 variables in the model after variable selection is insufficient.

Backward selection.

After 7 steps, following variables are removed from the model:

| **Summary of Backward Elimination** | | | | | | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step** | **Variable Removed** | **Label** | | | | **Number Vars In** | | **Partial R-Square** | | | **Model R-Square** | | | | **C(p)** | | **F Value** | **Pr > F** |
| **1** | PF2 |  | | | | 21 | | 0.0001 | | | 0.9859 | | | | 21.0330 | | 0.03 | 0.8610 |
| **2** | MP2 |  | | | | 20 | | 0.0001 | | | 0.9859 | | | | 19.0616 | | 0.03 | 0.8612 |
| **3** | P2a2 |  | | | | 19 | | 0.0003 | | | 0.9856 | | | | 17.2009 | | 0.18 | 0.6834 |
| **4** | FGA2 |  | | | | 18 | | 0.0006 | | | 0.9850 | | | | 15.4864 | | 0.40 | 0.5430 |
| **5** | P3P2 |  | | | | 17 | | 0.0005 | | | 0.9845 | | | | 13.7280 | | 0.35 | 0.5634 |
| **6** | STL2 |  | | | | 16 | | 0.0017 | | | 0.9829 | | | | 12.5587 | | 1.29 | 0.2783 |
| **7** | FTP2 |  | | | | 15 | | 0.0046 | | | 0.9783 | | | | 12.8668 | | 3.51 | 0.0838 |
| **Analysis of Variance** | | | | | | | | | | | | | | | | |
| **Source** | | | | **DF** | | **Sum of Squares** | | | | **Mean Square** | | | **F Value** | | **Pr > F** | |
| **Model** | | | | 15 | | 5238.30959 | | | | 349.22064 | | | 42.01 | | <.0001 | |
| **Error** | | | | 14 | | 116.39041 | | | | 8.31360 | | |  | |  | |
| **Corrected Total** | | | | 29 | | 5354.70000 | | | |  | | |  | |  | |
| **Root MSE** | | | | | | 2.88333 | | | **R-Square** | | | 0.9783 | | |
| **Dependent Mean** | | | | | | 40.90000 | | | **Adj R-Sq** | | | 0.9550 | | |
| **Coeff Var** | | | | | | 7.04971 | | |  | | |  | | |

Following variables are stay in the model:

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | -1698.65365 | 264.40803 | -6.42 | <.0001 | 0 |
| **FG2** |  | 1 | 228.38995 | 43.03601 | 5.31 | 0.0001 | 8987.96440 |
| **FGP2** |  | 1 | 6749.94338 | 1578.41875 | 4.28 | 0.0008 | 1172.22130 |
| **P32** |  | 1 | -61.33179 | 12.60153 | -4.87 | 0.0002 | 3948.26335 |
| **P3A2** |  | 1 | 27.97976 | 6.42188 | 4.36 | 0.0007 | 2382.81732 |
| **P22** |  | 1 | -106.72649 | 15.71517 | -6.79 | <.0001 | 1843.76953 |
| **P2p2** |  | 1 | -4085.09395 | 1156.60590 | -3.53 | 0.0033 | 850.18829 |
| **FT2** |  | 1 | 45.06717 | 10.06450 | 4.48 | 0.0005 | 1173.09501 |
| **FTA2** |  | 1 | -4.24088 | 0.99635 | -4.26 | 0.0008 | 15.26263 |
| **ORB2** |  | 1 | 62.56430 | 24.34756 | 2.57 | 0.0223 | 3847.85689 |
| **DRB2** |  | 1 | 108.93628 | 42.04374 | 2.59 | 0.0213 | 8380.32503 |
| **TRB2** |  | 1 | -122.58517 | 48.69081 | -2.52 | 0.0246 | 12617 |
| **AST2** |  | 1 | 1.90241 | 0.65413 | 2.91 | 0.0115 | 4.57421 |
| **BLK2** |  | 1 | 2.05622 | 0.54117 | 3.80 | 0.0020 | 2.12230 |
| **TOV2** |  | 1 | -2.02025 | 0.53774 | -3.76 | 0.0021 | 2.39498 |
| **PTS2** |  | 1 | -96.32628 | 23.67878 | -4.07 | 0.0012 | 7337.04319 |

After the 7 steps, the table listed above is the ultimate result for backward selection, although its R2 and CV are close to the full model However, there are still large numbers of variables in the model and the multicollinearity still exist, so I will not use this method neither.

Stepwise:

The stepwise result is identical to the forward selection since its R2 and CV are not satisfied so I will not use this method.

All possible regression

Since neither forward, backward and stepwise are accepted so I will try all possible regression using Cp as criteria.

NO. of Var Cp R2 Variables

|  |  |  |  |
| --- | --- | --- | --- |
| **8** | 8.8879 | 0.9402 | P3A2 P22 P2a2 FTP2 TRB2 AST2 STL2 TOV2 |
| **9** | 9.9771 | 0.9428 | P3A2 P22 P2a2 FT2 FTA2 TRB2 AST2 STL2 TOV2 |
| **9** | 10.0178 | 0.9427 | FGP2 P32 P2a2 FTP2 TRB2 AST2 STL2 BLK2 TOV2 |

According to the table above, I find those 3 models suit my needs that Cp = p+1 approximately. Since their R2 are about same I will run all of them to test their VIF and CV.

Model 1:

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 8 | 5034.57994 | 629.32249 | 41.28 | <.0001 |
| **Error** | 21 | 320.12006 | 15.24381 |  |  |
| **Corrected Total** | 29 | 5354.70000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 3.90433 | **R-Square** | 0.9402 |
| **Dependent Mean** | 40.90000 | **Adj R-Sq** | 0.9174 |
| **Coeff Var** | 9.54605 |  |  |

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | -100.72869 | 78.05633 | -1.29 | 0.2109 | 0 |
| **P3A2** |  | 1 | -2.82083 | 0.51939 | -5.43 | <.0001 | 8.50056 |
| **P22** |  | 1 | 4.77204 | 1.17658 | 4.06 | 0.0006 | 5.63648 |
| **P2a2** |  | 1 | -9.47324 | 0.88543 | -10.70 | <.0001 | 8.55784 |
| **FTP2** |  | 1 | 292.44533 | 58.83877 | 4.97 | <.0001 | 1.63022 |
| **TRB2** |  | 1 | 7.44971 | 0.98093 | 7.59 | <.0001 | 2.79285 |
| **AST2** |  | 1 | 1.42879 | 0.57875 | 2.47 | 0.0222 | 1.95283 |
| **STL2** |  | 1 | 4.26336 | 0.91620 | 4.65 | 0.0001 | 4.27829 |
| **TOV2** |  | 1 | -4.89440 | 0.80583 | -6.07 | <.0001 | 2.93324 |



H0: this model is not adequate H1: this model is adequate Alpha=0.05

R2 is 94% CV is 9.54 which is accepted, F-value is 41.28 P-value is smaller than 0.01 so reject the H0, therefore model is adequate.

Both QQ plot and Residual vs Predicted plots are look fine so no assumptions are violated. (Residuals vs each regressors plots are in the appendix)

For the T-test, all P-value for each variable are smaller than 0.05 so all variables are significant in this model. All VIF are smaller than 10

Overall this is quite good model. I will choose one as candidate.

Model 2:

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 9 | 5048.53867 | 560.94874 | 36.64 | <.0001 |
| **Error** | 20 | 306.16133 | 15.30807 |  |  |
| **Corrected Total** | 29 | 5354.70000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 3.91255 | **R-Square** | 0.9428 |
| **Dependent Mean** | 40.90000 | **Adj R-Sq** | 0.9171 |
| **Coeff Var** | 9.56614 |  |  |

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | 185.65006 | 73.96076 | 2.51 | 0.0208 | 0 |
| **P3A2** |  | 1 | -2.99103 | 0.55062 | -5.43 | <.0001 | 9.51352 |
| **P22** |  | 1 | 4.75602 | 1.18763 | 4.00 | 0.0007 | 5.71872 |
| **P2a2** |  | 1 | -9.80990 | 0.95225 | -10.30 | <.0001 | 9.85670 |
| **FT2** |  | 1 | 6.72321 | 1.37167 | 4.90 | <.0001 | **11.83369** |
| **FTA2** |  | 1 | -6.23024 | 1.23816 | -5.03 | <.0001 | **12.80044** |
| **TRB2** |  | 1 | 7.81235 | 1.04397 | 7.48 | <.0001 | 3.15008 |
| **AST2** |  | 1 | 1.29682 | 0.60214 | 2.15 | 0.0436 | 2.10500 |
| **STL2** |  | 1 | 4.41627 | 0.93073 | 4.74 | 0.0001 | 4.39652 |
| **TOV2** |  | 1 | -4.99386 | 0.81341 | -6.14 | <.0001 | 2.97615 |



R2 is 0.9428, CV is 9.56614 they are about same to the first one. However, some VIF number are larger than 10 (free throw and free throw attempt) so I will not choose this model as candidate.

Model 3：

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 9 | 5047.91508 | 560.87945 | 36.56 | <.0001 |
| **Error** | 20 | 306.78492 | 15.33925 |  |  |
| **Corrected Total** | 29 | 5354.70000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 3.91653 | **R-Square** | 0.9427 |
| **Dependent Mean** | 40.90000 | **Adj R-Sq** | 0.9169 |
| **Coeff Var** | 9.57588 |  |  |

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | -445.73510 | 87.86842 | -5.07 | <.0001 | 0 |
| **FGP2** |  | 1 | 482.01061 | 86.85072 | 5.55 | <.0001 | 1.92352 |
| **P32** |  | 1 | -3.87583 | 0.93848 | -4.13 | 0.0005 | **11.86859** |
| **P2a2** |  | 1 | -6.23414 | 0.98156 | -6.35 | <.0001 | **10.45154** |
| **FTP2** |  | 1 | 303.32503 | 60.81692 | 4.99 | <.0001 | 1.73084 |
| **TRB2** |  | 1 | 6.97964 | 1.03690 | 6.73 | <.0001 | 3.10126 |
| **AST2** |  | 1 | 1.69478 | 0.69078 | 2.45 | 0.0234 | 2.76476 |
| **STL2** |  | 1 | 3.95291 | 0.84014 | 4.71 | 0.0001 | 3.57508 |
| **BLK2** |  | 1 | 0.69319 | 0.57734 | 1.20 | 0.2439 | 1.30914 |
| **TOV2** |  | 1 | -4.69441 | 0.77481 | -6.06 | <.0001 | 2.69489 |



R2: 94.27 % CV: 9.57 both are accepted.

However, VIF for P3 and P2A are grater than 10 so reject.

In summary I will choose the first model.

1. **Conclusion**

After all tests, model selections, outlier detection etc. I get my final model

| **Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variance Inflation** |
| **Intercept** | Intercept | 1 | -100.72869 | 78.05633 | -1.29 | 0.2109 | 0 |
| **P3A2** |  | 1 | -2.82083 | 0.51939 | -5.43 | <.0001 | 8.50056 |
| **P22** |  | 1 | 4.77204 | 1.17658 | 4.06 | 0.0006 | 5.63648 |
| **P2a2** |  | 1 | -9.47324 | 0.88543 | -10.70 | <.0001 | 8.55784 |
| **FTP2** |  | 1 | 292.44533 | 58.83877 | 4.97 | <.0001 | 1.63022 |
| **TRB2** |  | 1 | 7.44971 | 0.98093 | 7.59 | <.0001 | 2.79285 |
| **AST2** |  | 1 | 1.42879 | 0.57875 | 2.47 | 0.0222 | 1.95283 |
| **STL2** |  | 1 | 4.26336 | 0.91620 | 4.65 | 0.0001 | 4.27829 |
| **TOV2** |  | 1 | -4.89440 | 0.80583 | -6.07 | <.0001 | 2.93324 |

Since I already check the T-test, F-test, variance assumption, VIFs in the last section about this model is good enough.

The residuals vs predicted plots are given:





Prediction equation:

Y-hat=-100.72869-2.82083sqrt(P3A)+4.77sqrt(P2)-9.47sqrt(P2A)+292.44533sqrt(FTP)+7.449sqrt(TRB)+1.42sqrt(AST)=4.26sqrt(STL)-4.89sqrt(TOV)

To test the appropriate and usefulness of my model I download the 16-17 season data and try to predict the games won using my model

Here is result:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Team | pred | Wons | Diff | RankP |  | Rank |
| Golden State Warriors\* | 85.43407 | 67 | 18.43407 | 1 | [State Warriors\* (1)](https://www.basketball-reference.com/teams/GSW/2017.html) | 1 |
| Houston Rockets\* | 77.11973 | 55 | 22.11973 | 3 | [San Antonio Spurs\* (2)](https://www.basketball-reference.com/teams/SAS/2017.html) | 2 |
| San Antonio Spurs\* | 64.0568 | 61 | 3.0568 | 2 | [Houston Rockets\* (3)](https://www.basketball-reference.com/teams/HOU/2017.html) | 3 |
| Los Angeles Clippers\* | 60.326 | 51 | 9.325996 | 4 | [Los Angeles Clippers\* (4)](https://www.basketball-reference.com/teams/LAC/2017.html) | 4 |
| Utah Jazz\* | 57.71917 | 51 | 6.719171 | 5 | [Utah Jazz\* (5)](https://www.basketball-reference.com/teams/UTA/2017.html) | 5 |
| Denver Nuggets | 54.13031 | 40 | 14.13031 | 9 | [Oklahoma City Thunder\* (6)](https://www.basketball-reference.com/teams/OKC/2017.html) | 6 |
| Memphis Grizzlies\* | 52.94184 | 43 | 9.941841 | 7 | [Memphis Grizzlies\* (7)](https://www.basketball-reference.com/teams/MEM/2017.html) | 7 |
| Minnesota Timberwolves | 50.08433 | 31 | 19.08433 | 13 | [Portland Trail Blazers\* (8)](https://www.basketball-reference.com/teams/POR/2017.html) | 8 |
| Dallas Mavericks | 49.70966 | 33 | 16.70966 | 11 | [Denver Nuggets (9)](https://www.basketball-reference.com/teams/DEN/2017.html) | 9 |
| New Orleans Pelicans | 46.4875 | 34 | 12.4875 | 10 | [New Orleans Pelicans (10)](https://www.basketball-reference.com/teams/NOP/2017.html) | 10 |
| Oklahoma City Thunder\* | 45.35656 | 47 | -1.64344 | 6 | [Dallas Mavericks (11)](https://www.basketball-reference.com/teams/DAL/2017.html) | 11 |
| Portland Trail Blazers\* | 43.48611 | 41 | 2.486112 | 8 | [Sacramento Kings (12)](https://www.basketball-reference.com/teams/SAC/2017.html) | 12 |
| Sacramento Kings | 39.60826 | 32 | 7.608262 | 12 | [Minnesota Timberwolves (13)](https://www.basketball-reference.com/teams/MIN/2017.html) | 13 |
| Phoenix Suns | 30.25568 | 24 | 6.255677 | 15 | [Los Angeles Lakers (14)](https://www.basketball-reference.com/teams/LAL/2017.html) | 14 |
| Los Angeles Lakers | 30.1565 | 26 | 4.156503 | 14 | [Phoenix Suns (15)](https://www.basketball-reference.com/teams/PHO/2017.html) | 15 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| East | -100.729 |  |  |  |  |  |
| Boston Celtics\* | 66.52182 | 53 | 13.52182 | 1 | [oston Celtics\* (1)](https://www.basketball-reference.com/teams/BOS/2017.html) | 1 |
| Toronto Raptors\* | 66.25398 | 51 | 15.25398 | 3 | [Cleveland Cavaliers\* (2)](https://www.basketball-reference.com/teams/CLE/2017.html) | 2 |
| Cleveland Cavaliers\* | 59.12828 | 51 | 8.128283 | 2 | [Toronto Raptors\* (3)](https://www.basketball-reference.com/teams/TOR/2017.html) | 3 |
| Washington Wizards\* | 57.21341 | 49 | 8.213405 | 4 | [Washington Wizards\* (4)](https://www.basketball-reference.com/teams/WAS/2017.html) | 4 |
| Atlanta Hawks\* | 54.96314 | 43 | 11.96314 | 5 | [Atlanta Hawks\* (5)](https://www.basketball-reference.com/teams/ATL/2017.html) | 5 |
| Indiana Pacers\* | 49.83897 | 42 | 7.838969 | 7 | [Milwaukee Bucks\* (6)](https://www.basketball-reference.com/teams/MIL/2017.html) | 6 |
| Milwaukee Bucks\* | 49.78934 | 42 | 7.789344 | 6 | [Indiana Pacers\* (7)](https://www.basketball-reference.com/teams/IND/2017.html) | 7 |
| Chicago Bulls\* | 48.43936 | 41 | 7.439359 | 8 | [Chicago Bulls\* (8)](https://www.basketball-reference.com/teams/CHI/2017.html) | 8 |
| Miami Heat | 42.66909 | 41 | 1.669086 | 9 | [Miami Heat (9)](https://www.basketball-reference.com/teams/MIA/2017.html) | 9 |
| Detroit Pistons | 42.07821 | 37 | 5.078213 | 10 | [Detroit Pistons (10)](https://www.basketball-reference.com/teams/DET/2017.html) | 10 |
| Charlotte Hornets | 39.58669 | 36 | 3.586687 | 11 | [Charlotte Hornets (11)](https://www.basketball-reference.com/teams/CHO/2017.html) | 11 |
| New York Knicks | 39.14582 | 31 | 8.145818 | 12 | [New York Knicks (12)](https://www.basketball-reference.com/teams/NYK/2017.html) | 12 |
| Orlando Magic | 38.01275 | 29 | 9.012746 | 13 | [Orlando Magic (13)](https://www.basketball-reference.com/teams/ORL/2017.html) | 13 |
| Philadelphia 76ers | 33.26213 | 28 | 5.262132 | 14 | [Philadelphia 76ers (14)](https://www.basketball-reference.com/teams/PHI/2017.html) | 14 |
| Brooklyn Nets | 32.3612 | 20 | 12.3612 | 15 | [Brooklyn Nets (15)](https://www.basketball-reference.com/teams/BRK/2017.html) | 15 |
|  |  |  | 8.413675 |  |  |  |

After reviewing my prediction, I found that my model did not perform very well on games won prediction, average difference is 8 games and its predicates better in bad teams rather than good teams generally. However, the its shows almost 100% true ranks of east conference only 2 teams switch their order, in west conference, little bit worse than east conference, order of 4 teams are mess around. Since in the league, the performance of west is better than east and the competition is more intense from west than east.so it is explained when prediction in west is hard than east. Over all since I dropped 2/3 variables and the team may have some roster change during the summer this result is not bad for me.

As an experience NBA fan, I also in interesting in some funny facts or some secret/pro factors such as:

Per 100 possession stats.

Pace

Age

Strength of schedule

Arena attendance

Number of all stars

1. **Appendix.**

[www.nba.com](http://www.nba.com/)

<https://www.basketball-reference.com/leagues/NBA_2017.html#all_team-stats-per_game>

<https://en.wikipedia.org/wiki/2014%E2%80%9315_NBA_season>

Original model partial plots:









Trans 1 (1/x) partial plots:







Partial plot for trans 2 (final model selection)











Trans2 partial plots sqrt (x)









**SAS code:**

libname learn "\\Client\e$\sta6013RegressionAnalysis";

PROC IMPORT OUT= LEARN.Project

DATAFILE= "\\Client\E$\sta6013RegressionAnalysis\nba15.xlsx"

DBMS=EXCEL REPLACE;

RANGE="Sheet1$";

GETNAMES=YES;

MIXED=NO;

SCANTEXT=YES;

USEDATE=YES;

SCANTIME=YES;

RUN;

proc print data=learn.Project;run;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\reg\_origin.rtf';

proc reg data=learn.Project plots(label)=(cooksd dfbetas dffits diagnostics observedbypredicted) ;

model games\_won= MP FG FGA FGP P3 P3A P3P P2 P2A P2P FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS /r p vif partial influence collinoint corrb collin tol ;

run;

ods rtf close;

proc reg data=learn.Project;

model games\_won= MP FG FGA FGP P3 P3A P3P P2 P2A P2P FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS ;

run;

proc sort data=learn.project;by \_all\_ ;run;

proc print data=learn.Project;run;

\*statstics summary;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\means.rtf';

proc means data=learn.project mean std min max p25 p50 p75;

run;

ods rtf close;

\*scatter plots;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\scatter\_box.rtf';

proc sgplot data=learn.project;

reg y=games\_won x=mp;

run;

proc sgplot data=learn.project;

hbox mp;

run;

proc sgplot data=learn.project;

reg y=games\_won x=fg;

run;

proc sgplot data=learn.project;

hbox fg;

run;

proc sgplot data=learn.project;

reg y=games\_won x=fga;

run;

proc sgplot data=learn.project;

hbox fga;

run;

proc sgplot data=learn.project;

reg y=games\_won x=fgp;

run;

proc sgplot data=learn.project;

hbox fgp;

run;

proc sgplot data=learn.project;

reg y=games\_won x=p3;

run;

proc sgplot data=learn.project;

hbox p3;

run;

proc sgplot data=learn.project;

reg y=games\_won x=p3a;

run;

proc sgplot data=learn.project;

hbox p3a;

run;

proc sgplot data=learn.project;

reg y=games\_won x=p3p;

run;

proc sgplot data=learn.project;

hbox p3p;

run;

proc sgplot data=learn.project;

reg y=games\_won x=p2;

run;

proc sgplot data=learn.project;

hbox p2;

run;

proc sgplot data=learn.project;

reg y=games\_won x=p2a;

run;

proc sgplot data=learn.project;

hbox p2a;

run;

proc sgplot data=learn.project;

reg y=games\_won x=p2p;

run;

proc sgplot data=learn.project;

hbox p2p;

run;

proc sgplot data=learn.project;

reg y=games\_won x=ft;

run;

proc sgplot data=learn.project;

hbox ft;

run;

proc sgplot data=learn.project;

reg y=games\_won x=fta;

run;

proc sgplot data=learn.project;

hbox fta;

run;

proc sgplot data=learn.project;

reg y=games\_won x=ftp;

run;

proc sgplot data=learn.project;

hbox ftp;

run;

proc sgplot data=learn.project;

reg y=games\_won x=orb;

run;

proc sgplot data=learn.project;

hbox orb;

run;

proc sgplot data=learn.project;

reg y=games\_won x=drb;

run;

proc sgplot data=learn.project;

hbox drb;

run;

proc sgplot data=learn.project;

reg y=games\_won x=trb;

run;

proc sgplot data=learn.project;

hbox trb;

run;

proc sgplot data=learn.project;

reg y=games\_won x=ast;

run;

proc sgplot data=learn.project;

hbox ast;

run;

proc sgplot data=learn.project;

reg y=games\_won x=stl;

run;

proc sgplot data=learn.project;

hbox stl;

run;

proc sgplot data=learn.project;

reg y=games\_won x=blk;

run;

proc sgplot data=learn.project;

hbox blk;

run;

proc sgplot data=learn.project;

reg y=games\_won x=tov;

run;

proc sgplot data=learn.project;

hbox tov;

run;

proc sgplot data=learn.project;

reg y=games\_won x=pf;

run;

proc sgplot data=learn.project;

hbox pf;

run;

proc sgplot data=learn.project;

reg y=games\_won x=pts;

run;

proc sgplot data=learn.project;

hbox pts;

run;

ods rtf close;

\*loess;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\loess.rtf';

proc loess data=learn.project;

model games\_won=mp/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=fg/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=fga/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=fgp/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=p3/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=p3a/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=p3p/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=p2/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=p2a/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=p2p/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=ft/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=fta/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=ftp/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=orb/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=drb/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=trb/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=ast/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=stl/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=blk/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=tov/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=pf/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

proc loess data=learn.project;

model games\_won=pts/smooth= 0.2 0.4 0.6 0.8 1.0;

run;

ods rtf close;

\*corr;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\corr.rtf';

proc corr data=learn.project plots=matrix (histogram);

var games\_won MP FG FGA FGP P3 P3A P3P

P2 P2a P2p FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS;

run;

ods rtf close;

proc sgscatter data=learn.projectplots=matrix (histogram) ;

matrix games\_won MP FG FGA FGP P3 P3A P3P P2 P2a P2p FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS;

;

run;

\*transformation;

data project\_t1;

set learn.project;

MP1=1/mp; FG1=1/fg ;FGA1=1/fga ;FGP1=1/fgp; P31=1/p3; P3A1=1/p3a; P3P1=1/p3p;

P21=1/p2; P2a1=1/p2a; P2p1=1/p2p; FT1=1/ft; FTA1=1/fta; FTP1=1/ftp; ORB1=1/orb;

DRB1=1/drb; TRB1=1/trb; AST1=1/ast; STL1=1/stl; BLK1=1/blk; TOV1=1/tov; PF1=1/pf; PTS1=1/pts;

games\_won2=(games\_won)\*(games\_won);

run;

data project\_t2;

set learn.project;

MP2=sqrt(mp); FG2=sqrt(fg) ;FGA2=sqrt(fga) ;FGP2=sqrt(fgp); P32=sqrt(p3); P3A2=sqrt(p3a); P3P2=sqrt(p3p);

P22=sqrt(p2); P2a2=sqrt(p2a); P2p2=sqrt(p2p); FT2=sqrt(ft); FTA2=sqrt(fta); FTP2=sqrt(ftp); ORB2=sqrt(orb);

DRB2=sqrt(drb); TRB2=sqrt(trb); AST2=sqrt(ast); STL2=sqrt(stl); BLK2=sqrt(blk); TOV2=sqrt(tov); PF2=sqrt(pf); PTS2=sqrt(pts);

games\_won2=(games\_won)\*(games\_won);

run;

data project\_t3;

set learn.project;

MP1=1/mp; FG1=1/fg ;FGA1=1/fga ;FGP1=1/fgp; P31=1/p3; P3A1=1/p3a; P3P1=1/p3p;

P21=1/p2; P2a1=1/p2a; P2p1=1/p2p; FT1=1/ft; FTA1=1/fta; FTP1=1/ftp; ORB1=1/orb;

DRB1=1/drb; TRB1=1/trb; AST1=1/ast; STL1=1/stl; BLK1=1/blk; TOV1=1/tov; PF1=1/pf; PTS1=1/pts;

games\_won2=(games\_won)\*(games\_won);

MP2=sqrt(mp); FG2=sqrt(fg) ;FGA2=sqrt(fga) ;FGP2=sqrt(fgp); P32=sqrt(p3); P3A2=sqrt(p3a); P3P2=sqrt(p3p);

P22=sqrt(p2); P2a2=sqrt(p2a); P2p2=sqrt(p2p); FT2=sqrt(ft); FTA2=sqrt(fta); FTP2=sqrt(ftp); ORB2=sqrt(orb);

DRB2=sqrt(drb); TRB2=sqrt(trb); AST2=sqrt(ast); STL2=sqrt(stl); BLK2=sqrt(blk); TOV2=sqrt(tov); PF2=sqrt(pf); PTS2=sqrt(pts);

games\_won2=(games\_won)\*(games\_won);

run;

\*transform 1/x;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\tran1.rtf';

proc reg data=project\_t1 plots(label)=(cooksd dfbetas dffits diagnostics observedbypredicted);

model games\_won=MP1 FG1 FGA1 FGP1 P31 P3A1 P3P1 P21 P2a1 P2p1 FT1 FTA1 FTP1 ORB1 DRB1 TRB1 AST1 STL1 BLK1 TOV1 PF1 PTS1/alpha=0.05 r p vif partial influence collinoint corrb collin tol ;

run;

ods rtf close;

\*trans2 sqrt(x);

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\tran2.rtf';

proc reg data=project\_t2 plots(label)=(cooksd dfbetas dffits diagnostics observedbypredicted);

model games\_won=MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2/alpha=0.05 r p vif partial influence collinoint corrb collin tol ;

run;

ods rtf close;

\*boxcox;

proc transreg data=learn.project test plots=(transformation(dependent) scatter

observedbypredicted) COEFFICIENTS ;

model BoxCox(games\_won) = identity (MP FG FGA FGP P3 P3A P3P P2 P2A P2P FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS );

run;

\*trans2 y^2;

data project\_t3;

set learn.project;

games\_won2=(games\_won)\*(games\_won);

run;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\tran3.rtf';

proc reg data=project\_t3 plots(label)=(cooksd dfbetas dffits diagnostics observedbypredicted) ;

model games\_won2=MP FG FGA FGP P3 P3A P3P P2 P2A P2P FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS/alpha=0.05 r p vif partial influence collinoint corrb collin tol ;

run;

ods rtf close;

proc reg data=learn.Project ;

model games\_won= MP FG FGA FGP P3 P3A P3P P2 P2a P2p FT FTA FTP ORB DRB TRB AST STL BLK TOV PF PTS/selection=stepwise rsquare cp sse mse sle=0.05 sls=0.05 vif;

run;

proc reg data=project\_t2 ;

model games\_won2=MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2/vif;

run;

proc reg data=project\_t1;

model games\_won2=MP1 FG1 FGA1 FGP1 P31 P3A1 P3P1 P21 P2a1 P2p1 FT1 FTA1 FTP1 ORB1 DRB1 TRB1 AST1 STL1 BLK1 TOV1 PF1 PTS1/vif;

run;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\tran4.rtf';

proc reg data= project\_t3 plots(label)=(cooksd dfbetas dffits diagnostics observedbypredicted) ;

model games\_won= MP1 FG1 FGA FGP P3 P3A P3P2 P21 P2A P2P1 FT FTA FTP2 ORB1 DRB TRB AST2 STL BLK2 TOV PF PTS2 /r p vif partial influence collinoint corrb collin tol ;

run;

ods rtf close;

proc reg data= project\_t3 plots(label)=(cooksd dfbetas dffits diagnostics observedbypredicted) ;

model games\_won2= MP2 FG2 FGA FGP2 P32 P3A2 P3p2 P21 P2A1 P2P2 FT2 FTA2 FTP2 ORB1 DRB TRB AST1 STL BLK2 TOV PF PTS1 /r p vif partial influence collinoint corrb collin tol ;

run;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\ridge.rtf';

proc reg data=project\_t3 plots(only)=ridge(unpack Ridgeaxis=log)

outest=ridge outvif outstb rsquare ridge=0.0 to 0.2 by .02;

model games\_won=MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2 /influence vif rsquare;

run;

proc print data=ridge;run;

ods rtf close;

proc reg data=project\_t3 ;

model games\_won= MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2 /selection=farward rsquare cp sse mse sle=0.05 sls=0.05 vif ;

run;

proc reg data=project\_t3 ;

model games\_won= MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2 /selection=backward rsquare cp sse mse sle=0.05 sls=0.05 vif ;

run;

proc reg data=project\_t3 ;

model games\_won= MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2 /selection=stepwise rsquare cp sse mse sle=0.05 sls=0.05 vif ;

run;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\best.rtf';

proc reg data=project\_t3;

model games\_won=MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2/selection=Cp ;

run;

ods rtf close;

proc robustreg data=project\_t3 method=m (scale=med wf=huber) ;

model games\_won=MP2 FG2 FGA2 FGP2 P32 P3A2 P3P2 P22 P2a2 P2p2 FT2 FTA2 FTP2 ORB2 DRB2 TRB2 AST2 STL2 BLK2 TOV2 PF2 PTS2;

output out =t1 weight=wgt;

run;

ods rtf style=statistical file='\\Client\E$\sta6013RegressionAnalysis\best3.rtf';

proc reg data=Project\_t3;

model games\_won=P3A2 P22 P2a2 FTP2 TRB2 AST2 STL2 TOV2/vif;

run;

proc reg data=Project\_t3;

model games\_won=P3A2 P22 P2a2 FT2 FTA2 TRB2 AST2 STL2 TOV2/vif;

run;

proc reg data=Project\_t3;

model games\_won=FGP2 P32 P2a2 FTP2 TRB2 AST2 STL2 BLK2 TOV2/vif;

run;

ods rtf close;