2020

Predictive Modeling of the Outcomes of University of Vermont Women's Basketball Games

Jennifer Eberling
University of Vermont

Predictive Modeling of the Outcomes of University of Vermont Women's Basketball Games

Jennifer Eberling

Honors College Thesis

Advisor: Bernard Cole | Department of Mathematics and Statistics

2019 – 2020

Index

Introduction

When I began my exploration of college sports, I was surprised at the vast differences in analytics between programs. The technology open and available to schools in the Big 10 conference dwarfed what the University of Vermont was able to provide. The gap was particularly wide between women's and men's programs. While many databases exist for picking the ideal March Madness bracket, very little was publicly available for women's basketball. All I could find was the most recent season's statistics on the individual schools' pages.

I reached out to the University of Vermont (UVM) Women's Basketball coaching staff with the hopes of contributing to the program through an in-depth analysis. All the typical basketball statistics (book statistics) are recorded and published for the team, but the coaches would look at a few key statistics individually – turnovers, fouls, blocked shots – to summarize games. What I suggested to the coaches was to allow me to record the statistics separately, create a database, and then use R to look for more complex patterns throughout the season.

The coaches agreed, and also proposed that I use the application Breakthrough Stats to record where on the court players of both teams took shots from and whether or not they made the shots. This application allowed me to record the percentages of each player and also where every shot was taken from over the course of a season. There was a learning curve to watching a game, recording statistics, and logging shot locations all at once, but practicing on pre-season and non-conference games made me fast and accurate with my records.

UVM faces a wide variety of levels of opponents in a season, and since a full season can span from mid-October to March, the first few weeks are spent exploring and changing team dynamics, deciding who the starters will be, and what the game strategy will be. Therefore, this analysis will only include data from the conference season, a period of 16 games late in the season, played against eight teams. Each team plays UVM at a home and an away game and plays UVM once in the beginning of the season and once at the end. The order of teams played remains the same for the second eight games played.

Teams played during the conference season are generally at the same level of ability as UVM, although each one certainly has individual strengths, styles, and levels of development. For instance, in the 2019 – 2020 season, the University of Maine's women's basketball roster had five senior and three freshmen, while the University of Vermont's only had one senior and six freshman, a notable difference when a team of seniors means that a team that has had more time to work out, to develop skills, and to improve teamwork.

Still, keeping track of conference play was the best way to find a level playing field for UVM. Since a conference season in only 16 games, I recorded statistics for both the 2018 – 2019 and the 2019 – 2020 season. Between them, UVM lost four starting players (players who are typically your best) either to graduation or transferring, but retained its top player in points, rebounds, and free throws. While the team lost tall players, it gained fast ones. Therefore, I am confident that the overall quality and general strategy of the team remained the same, allowing me to evaluate the seasons together.

Problem Statement

The University of Vermont Women's Basketball team collects lots of data from each of its games. However, the team has previously only looked at one statistic at a time in order to improve performance in specific areas. It is possible that trends in statistics across games or pairs of statistics within games could reveal more effective ways of strategizing. A multivariate linear model was built to predict UVM's performance, hopefully leading to a better understanding of the team and staff's style of play.

Literature Review

Every basketball in the America East Conference records and publishes book statistics for official games, but no one is publicly analyzing them. I could not find any literature exploring the eighteen court zones or even mentioning them. Most of the current research concerns general analysis with the book statistics.

Researchers have been exploring the uses of linear regression in athletics [1-6], but a large part of the studies is still limited to professional games and the NCAA Tournament. A few examples are using linear regression to model the relationships between core stability and jump shots, accuracy in jump shots, and game result predictions. Women are rarely included in these studies, a void that my research will fill, as it exclusively focuses on a female team. Linear regression is versatile in what it can be used for, so taking methods and graphs from other fields is appropriate [8-9].

The methods in those studies act as an example for examining significance in relationships between variables. Studies done by Kyle Steenland and James A. Deddens disregarded opponents' skill levels, something necessary in my own study, but he used a sample size of 8495, which was not feasible in this thesis [8]. Gregory T. Knofczynski and Daniel Mundfrom determined that with a high enough correlation coefficient, the need for a large sample population lessens, which is encouragement to find variables with high correlation coefficients in the variables of this thesis [11]. Aside from linear regression, neutral models, prediction intervals, and empirical Bayes confidence intervals have been proposed as alternative ways to predict outcomes in athletics, which may be complementary to ANOVA tests that will be used in this project if the requirements for tests are met [12-14]. Other theses have been written to predict winning basketball games, but they have not focused on female teams, have not included zone percentages, and often only look at data from professional instead of college games [15-16].

More research has been done on the psychological effects of sports. A player's confidence can greatly affect her decisions and how she plays basketball [17]. This is important, because the models in this thesis are based on statistics from entire games, but the way a team plays in the 4th quarter can be determined by the outcome of the 1st quarter, for better or for worse. Personal lives can also affect the style of a game. Female student athletes' psyches differ from their male counterparts, and that is also important to consider when comparing models across genders [18].

Methods

I. Collection

To collect my data, I attended a selection of home games and watched pre-recorded film on ESPN+ for away games. Data was taken from the 2018 – 2019 and 2019 – 2020 Conference Seasons. I used the iPad application

Breakthrough Stats to record layup percentages, turn over points, put-back misses and makes, offensive and defensive rebounds, assists, steals, blocks, deflects, charges, recovered balls, fouls, jump balls for both teams. The value point system scores (VPS), and efficiency statistics (Effics) was automatically calculated for every player and for each team. The shot locations and results (make or miss) were also recorded with Breakthrough Stats. At the end of the game, the resulting percentage of success was calculated and recorded for each of 18 floor zones, as well as the percentage of each team's shots taken from each zone. See figure 1 for zones.

I recorded these statistics for every individual player from both teams in every game. No time was recorded. The events were recorded mostly in order, but not completely, due to any corrections to the data post-collection. For this reason, order does not affect this analysis of the game. The data was then transferred to a laptop in the form of a CSV file and stored as an excel file. The total statistics for each team were pulled from each of the 32 games and put into separate excel files. One file stores all of the NCAA reported statistics, and one stores the zone percentages for each team. In each of these databases, a new statistic was created – Score difference (S_{diff}).

$$S_{diff} = \text{UVM's final points} - \text{Opponent's final points}$$

Thus, a win by UVM results in a positive Score value and a loss results in a negative Score value. This is the responding variable that my exploratory analysis attempts to model, because wins during the conference season (positive Score values) are what determine whether or not a team continues playing in the post-season, an opportunity that can result in more university funding, more attention from potential recruits, and positive headspace for the following season.

II. Analysis

First, I created S_{diff} variable to reflect the difference in the teams' points for each game. A positive number is how many points UVM won by, and a negative number is how many points they lost by. Then, in order to view all of the variables at once, I merged the opponent's data set and UVM's data set in excel and renaming the data set "combo." I then deleted variables that I did not consistently record at games and ones that were directly calculated with the block victim, opponents' charges taken, who was fouled, when there was a forced rushed shot, jump ball victim, minutes played, value points system number (VPS), and efficiency scores.

Variables were selected on their linearity with S_{diff} and used in proc glmselect in SAS. Glmselect included the Akaike information criterion (AIC) as a criterion for model selection. Using the AIC limits the amount of error in the model that can be attributed to overfitting. It penalizes for adding too many variables into a model by decreasing the R-squared value. I included it in this analysis because of the high number of variables available to use in the analysis.

Results

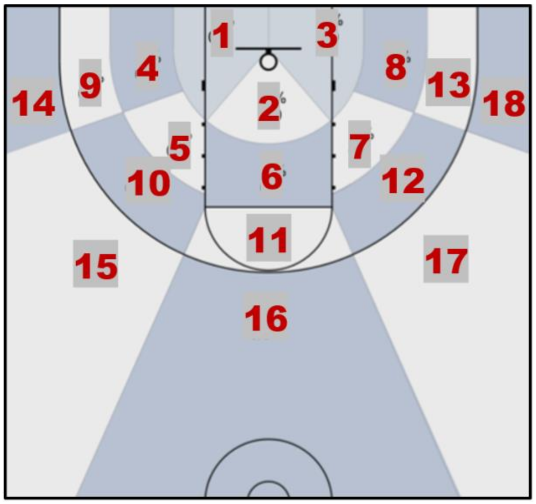


Figure 1

This layout is how the court zones are divided up. UVM's percentage made in zone one is referred to as "one," in zone two as "two," up to zone eighteen as "eighteen." For UVM's opponents, the zones are referred to as "oppone," "opptwo," and so on up to "oppeighteen." Oppenents' statistics are designated in the same manner.

Table 1 and Appendix 3 show the results from

```
     proc glmselect data=full plot = all;
     model score = oppsixteen oppseventeen one fifteen FTA
opp3Pt oppFG oppPts oppTO Foul
/ selection=stepwise(select=AIC) stats=all;
     run;
```

The scatter plots in Figure 2 show that the chosen variables are not linearly associated with one another.

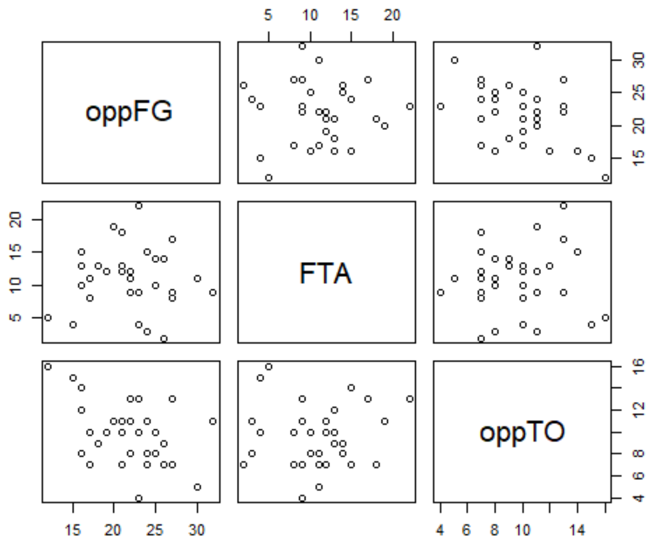


Figure 2

Using the AIC method, the optimal model is

$$S_{diff} = 12.402944 + 1.051618 * FTA - 1.837287 * oppFG + 1.203785 * oppTO$$

Figure 3 shows the above regression model on top of a scatterplot with the score difference predicted by the model on the x-axis and the actual score difference on the y-axis.

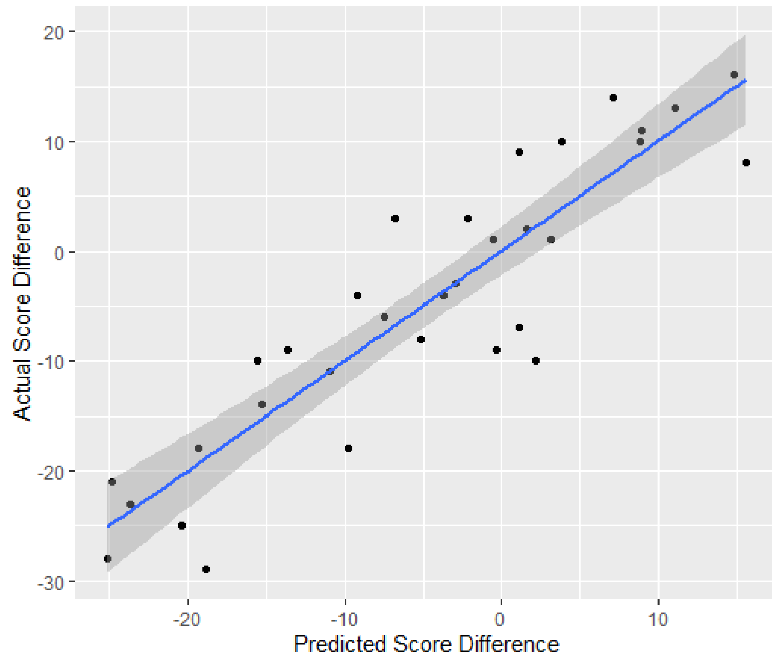


Figure 3

To check on the residual plots:

plot(lm(new$score~new$pred))

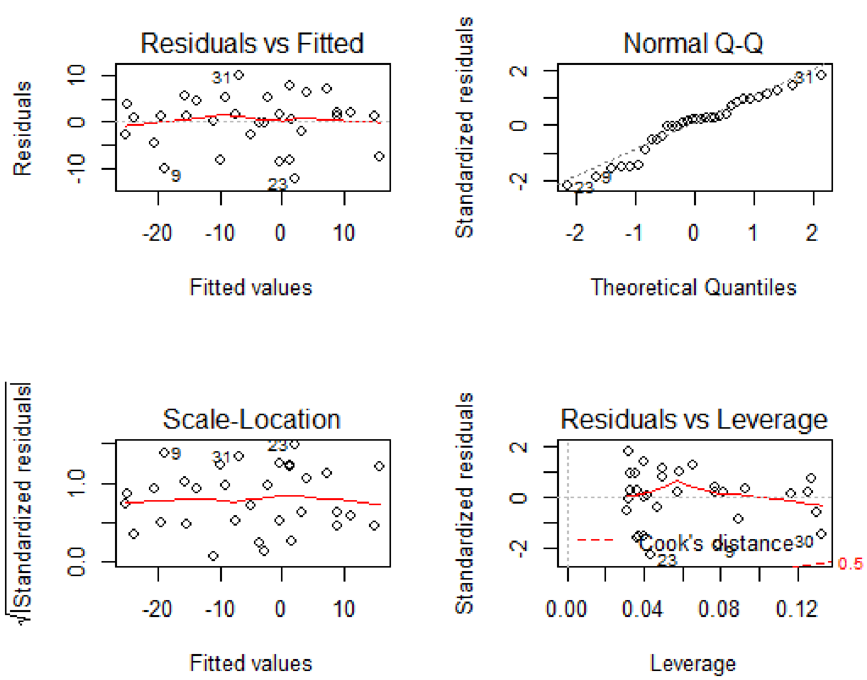


Figure 4

In SAS,

```
proc reg data = full;
model score = FTA oppFG oppTO/clb;
run;
```

results in the confidence intervals for Model 1 in Table 1.

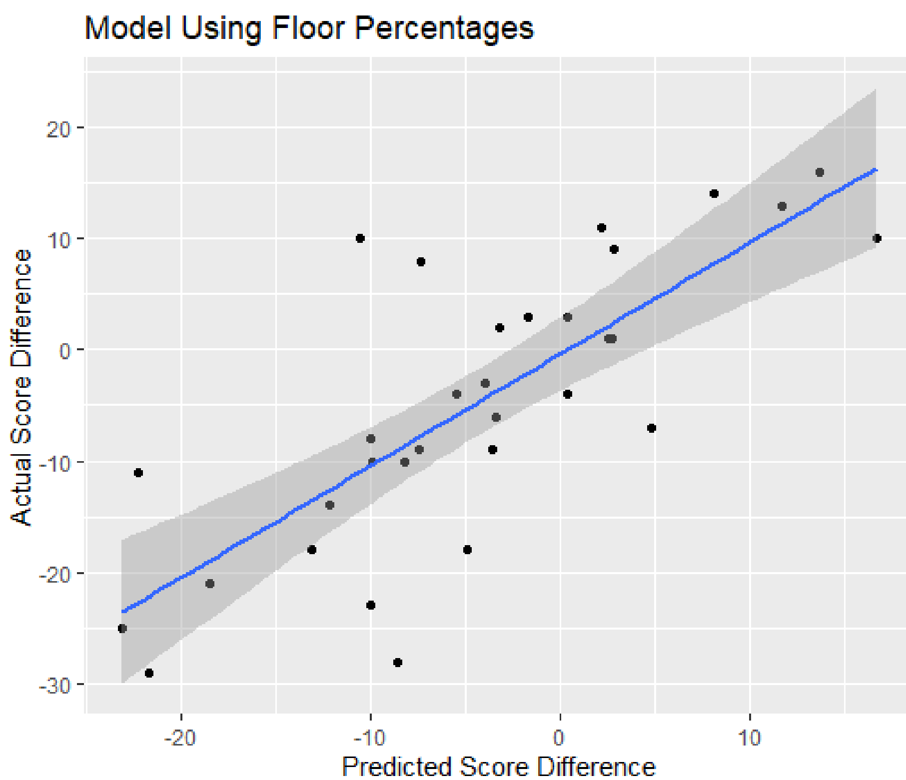


Figure 5

The plot is pictured in Figure 5 with a 95% confidence interval. Only the variables three, nine, thirteen, oppthree, oppfive, and oppsixteen have a statistically significant t-value. The AIC for this model is 176.48153, and the adjusted R-squared value is 0.4969. The results from

```
proc glmselect data = full plot = all;
model score = one two three four five eight nine thirteen
fourteen fifteen sixteen seventeen eighteen oppone opptwo
oppthree oppfour oppfive oppeight oppnine oppten oppfourteen
oppfifteen oppsixteen oppseventeen oppeighteen
/ selection = stepwise(select = AIC) stats=all;
```

```
run;
```

are in Table 1 describing Model 2 and Appendix 3. The full equation is

$$S_{diff} = 10.089135 - 31.871278 * three + 19.859586 * nine + 13.637504 * thirteen + 27.619231 * oppthree - 10.911277 * oppfour - 20.947020 * oppfive - 26.041759 * oppsixteen$$

The variable regression coefficients, confidence intervals, and p-values are shown in Table 1.

Table 1

Model/Variable	Regression Coefficient	95% CI	P-Value
Model 1			
Intercept	12.40	-4.85 to 29.65	0.152
Free Throws Attempted	1.05	0.61 to 1.50	<.001
Opponent's Field Goals	-1.84	-2.36 to -1.32	<.001
Opponent's Turn Overs	1.20	0.37 to 2.04	0.006
Adjusted R-Squared = 0.7950, AIC = 150.23			
Model 2			
Intercept	10.78	-2.14 to 23.70	0.098
Three	-33.77	-57.36 to -10.18	0.007
Nine	18.36	3.33 to 33.38	0.019
Thirteen	13.85	0.51 to 27.19	0.042
Opponent's Three	28.32	-0.91 to 57.56	0.057
Opponent's Four	-10.09	-22.10 to 1.92	0.096
Opponent's Five	-20.58	-31.03 to -10.12	<0.001
Opponent's Sixteen	-17.43	-44.36 to -12.50	0.001
Adjusted R-Squared = 0.4821, AIC = 176.48			

Discussion

The variables identified as the most linear with score difference were oppsixteen oppseventeen FG FTA opp3Pt oppFG oppPts and oppTO. Proc glmselect identified field goals, free throws attempted, opponent's points, and opponent's turn overs as the most impactful variables in that order. The function took into account the strength of association (trying to maximize) as well as the variation (trying to minimize). However, FG – oppPts results is almost exactly the difference in the score, it just does not account for UVM's free throws. I removed field goals from the function then, as it was less linearly correlated with the difference in score statistic. The model that I determined to be the most linear ended up depending on free throws attempted, opponent's field goals, and opponent's turn overs, listed in order of most to least important to the model.

Figure 4 confirmed that the model met all necessary assumptions. The residual versus fitted value plot appeared without pattern, the normal quantile plot showed the residuals followed a normal distribution, homoskedasticity is present, and the leverage values show no alarming outliers.

The model for S_{diff} is extremely linear, with most of its variation occurring where the predicted value was between -10 and 5. Unfortunately, having the variation occur around zero makes the model less useful for ambiguous games, which is when the model would be most helpful. If the variation occurred between predicted values of -20 and -10, then the model would still be accurate in predicting a loss and would thus be very useful. Variation around numbers close to zero prevents the model from giving a definitive statement about which team will win the game. If the coaches know they are likely to lose the next game of a tournament, then they can prepare for whichever team is in the loser's bracket. However, not knowing the outcome means that UVM loses the ability to prepare for specific opponents.

The F value of the model is 41.08 with 31 total degrees of freedom, which has a significance value of <.0001. The predicted values do have a statistically significant association with the actual values. Proc reg tells us that all of the variables are statistically significant, as none of the confidence intervals for the variables' coefficients contain zero. I am 95% confident that the true absolute value of the coefficients is above zero, meaning there is an association.

To see if the model was effective, I applied them to the quarterfinal playoff game against the University of Maine (2020). For the variables, I used the totals for the 2019 – 2020 conference season divided by 16, the number of conference season games. There were no statistics averaged per game available and using data from previous matchups of the teams cannot be used here, because they were used to create the model.

University of Maine Playoff Game

$$S_{diff} = 12.402944 + 1.051618 * 11.1875 - 1.837287 * 25.606 + 1.203785 * 12.0625$$
$$= -8.3569939845 = -8$$

UVM actually lost by 12, so the true $S_{diff} = -12$. That is 4 points lower than the predicted value, but since they are both losses, the model was accurate in predicting whether the game was a win or a loss, which is still helpful to the team.

I also created a model using solely the zone percentages for both UVM and its opponents. The variables I put into glmselect were determined by which had a semblance of linearity. For zone percentages, the zones with only one or two shots in them each game were largely excluded, because the percentages were either zero or one for a majority of games recorded. The variables input were one, two, three, four, five, eight, nine, thirteen, fourteen, fifteen, sixteen, seventeen, eighteen, oppone, opptwo, oppthree, oppfour, oppfive, oppeight, oppnine, oppten, oppfourteen, oppfifteen, oppsixteen, oppsenventeen, and oppeighteen. The model returned was

$$S_{diff} = 10.089135 - 31.871278 * three + 19.859586 * nine + 13.637504 * thirteen + 27.619231 * oppthree - 10.911277 * oppfour - 20.947020 * oppfive - 26.041759 * oppsixteen$$

No game could be run to test the model, as UVM is the only team with data on zones for all of its games, and all of that data was used to create the model.

The model using just zones has a lower R-squared than the model that includes all of the statistics. It shows that when it comes to percentages, UVM's matter more in predicting the outcome of the game, as the first three variables (three, nine, and thirteen) are percentages from UVM. This is contradictory to how two of the three variables in the first model were from the opponent.

This exploratory analysis only included two seasons, but in that time the team lost four seniors and one transfer that saw conference game minutes. The team's top scorer and top defensive player remained, ensuring a similar game dynamic in both seasons. The coaching staff remained the same, and the data collected was from the head coach's first and second season in her position. UVM won 13 games and lost 19 games.

Future research should include a continued collection of UVM Women's Basketball, as the playing style of the team changes throughout seasons. Data on zone percentages should be collected and analyzed alongside the book statistics for other teams in the America East Conference and other NCAA teams. Once data is collected for women's basketball, comparing that to men's basketball could help coaches understand the subtle differences between the styles of play. This could make the transition from men's basketball to women's basketball easier for coaches. For individual teams, this analysis would be helpful in designing plays that optimize their players' talents and reveal which aspects of basketball each player needs to work on.

For a game as fluid as basketball, it might seem like there should be more factors creating a complex model for predicting how much a team will win or lose by, but the style of play varies so much from game to game that only a few variables consistently correlate with the score over games.

References

1. Huang, B. (2014). Linear regression analysis between core stability and jump shot. *Applied Mechanics and Materials, 539*, 911-914 doi:http://dx.doi.org/10.4028/www.scientific.net/AMM.539.911.

2. Yates, Gary, and L. E. Holt. "The development of multiple linear regression equations to predict accuracy in basketball jump shooting." *ISBS-Conference Proceedings Archive*. Vol. 1. No. 1. 1983.

3. Abidin, Nahdiya Zainal, and Mohd Bakri Adam. "Prediction of vertical jump height from anthropometric factors in male and female martial arts athletes." *The Malaysian journal of medical sciences: MJMS* 20.1 (2013): 39.

4. Smith, Tyler, and Neil C. Schwertman. "Can the NCAA basketball tournament seeding be used to predict margin of victory?." *The american statistician* 53.2 (1999): 94-98.

5. Miljković, Dragan, et al. "The use of data mining for basketball matches outcomes prediction." *IEEE 8th International Symposium on Intelligent Systems and Informatics*. IEEE, 2010.

6. Bashuk, Mark. "Using cumulative win probabilities to predict NCAA basketball performance." *Proceedings of the MIT Sloan Sports Analytics Conference*. 2012.

7. Smith, Tyler, and Neil C. Schwertman. "Can the NCAA basketball tournament seeding be used to predict margin of victory?." *The american statistician* 53.2 (1999): 94-98.

8. Bianco, Vincenzo, Oronzio Manca, and Sergio Nardini. "Electricity consumption forecasting inItaly using linear regression models." *Energy* 34.9 (2009): 1413 1421.

9. Cai, T. Tony, and Peter Hall. "Prediction in functional linear regression." *The Annals of Statistics* 34.5 (2006): 2159-2179.

10. Steenland, Kyle, and James A. Deddens. "Effect of travel and rest on performance of professional basketball players." *Sleep*20.5 (1997): 366-369.

11. Knofczynski, Gregory T., and Daniel Mundfrom. "Sample sizes when using multiple linear regression for prediction." *Educational and psychological measurement* 68.3 (2008): 431-442.

12. Maszczyk, Adam, et al. "Application of neural and regression models in sports resultsprediction." *Procedia-Social and Behavioral Sciences* 117 (2014): 482 487.

13. Hu, Feifang, and James V. Zidek. "Forecasting NBA basketball playoff outcomes using the weighted likelihood." *A Festschrift for Herman Rubin*. Institute of Mathematical Statistics, 2004. 385-395.

14. Cox, D. R. "Prediction intervals and empirical Bayes confidence intervals." *Journal of Applied Probability* 12.S1 (1975): 47-55.

15. Yang, Yuanhao. “Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics.” *University of California at Berkeley*, 2015.

16. Shanahan, Kathleen Jean. “A Model for Predicting the Probability of a Win in Basketball.” *University of Iowa*, 1984.

17. Roane, Henry S., et al. “Behavioral Momentum In Sports: A Partial Replication With Women's Basketball.” *Journal of Applied Behavior Analysis*, vol. 37, no. 3, 2004, pp. 385–390., doi:10.1901/jaba.2004.37-385.

18. Sherry, Mark, and Kristi Zeller. “Gender and Motivation: A Study of the Athletic and Academic Motivations of Division I Female College Basketball Players.” *Women's Studies*, vol. 43, no. 1, 2014, pp. 73–92., doi:10.1080/00497878.2014.852425.

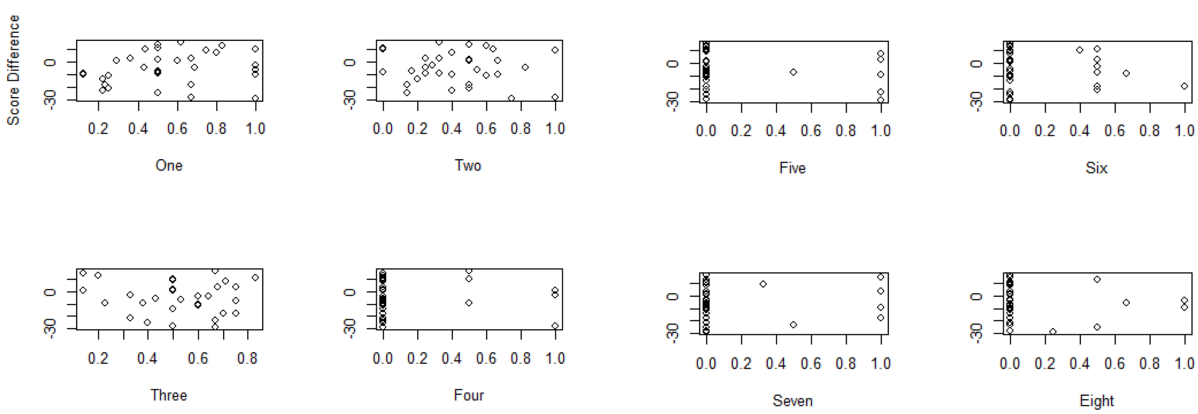
Appendix I

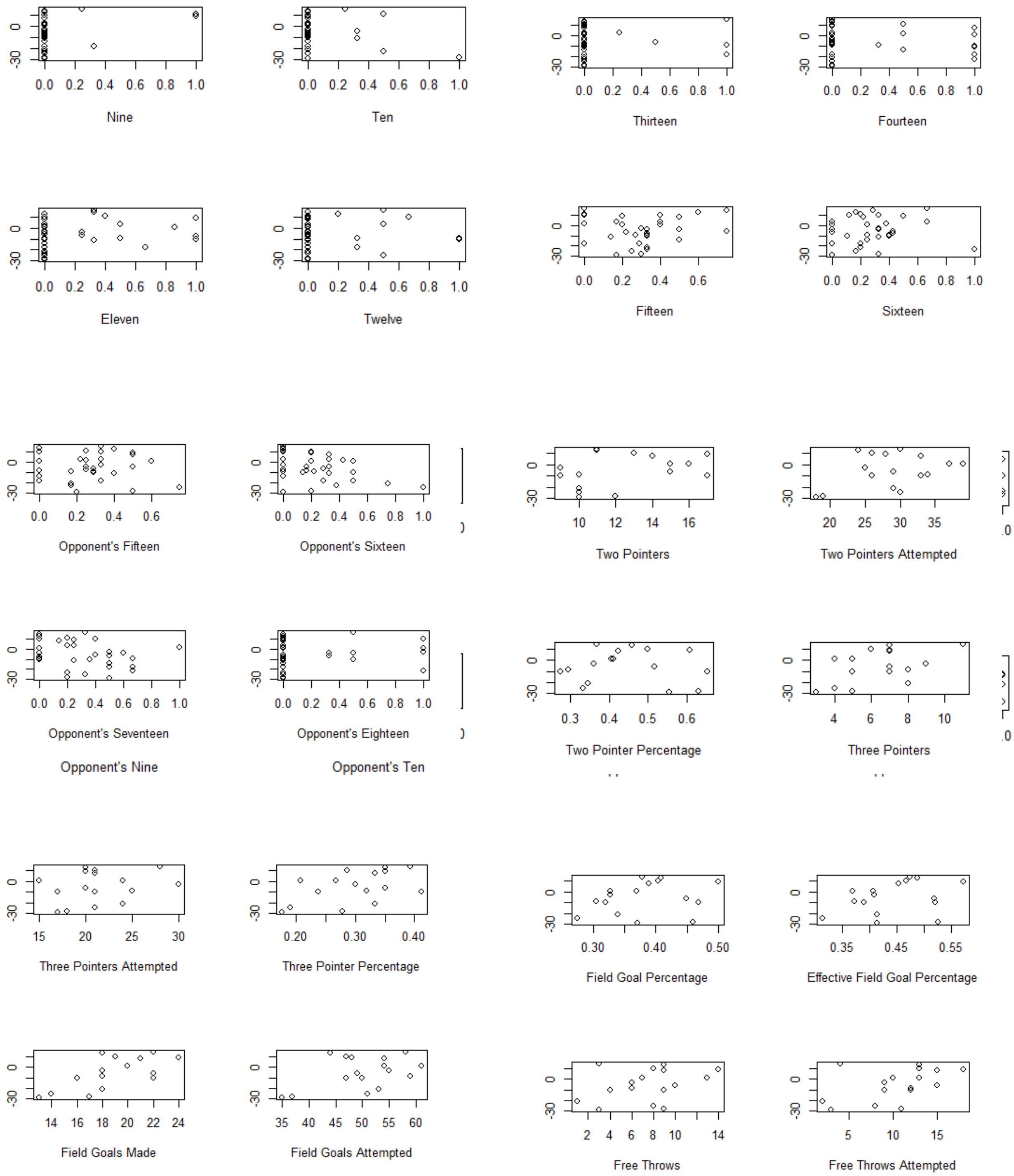
Acronym	Term
UVM	University of Vermont
FG(A)	Field Goals Attempted
FT(A)	Free Throws Attempted
Pts	Points
TO	Turn Overs
Paint	The inner rectangle of the court

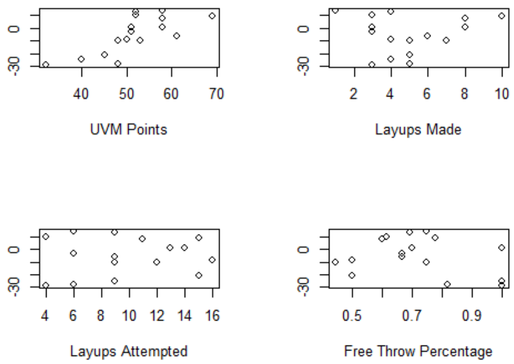
Appendix II

Scatter plots of all variables against $S_{diff.}$

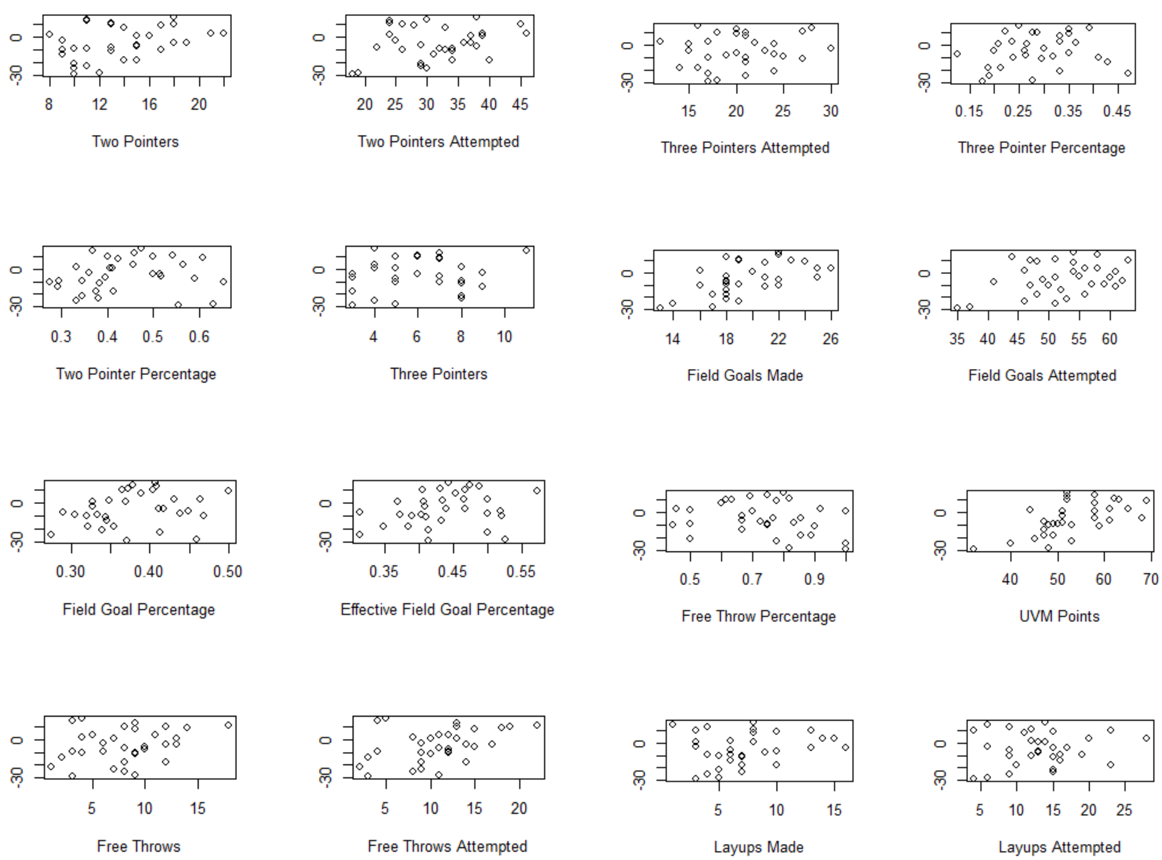
Zones

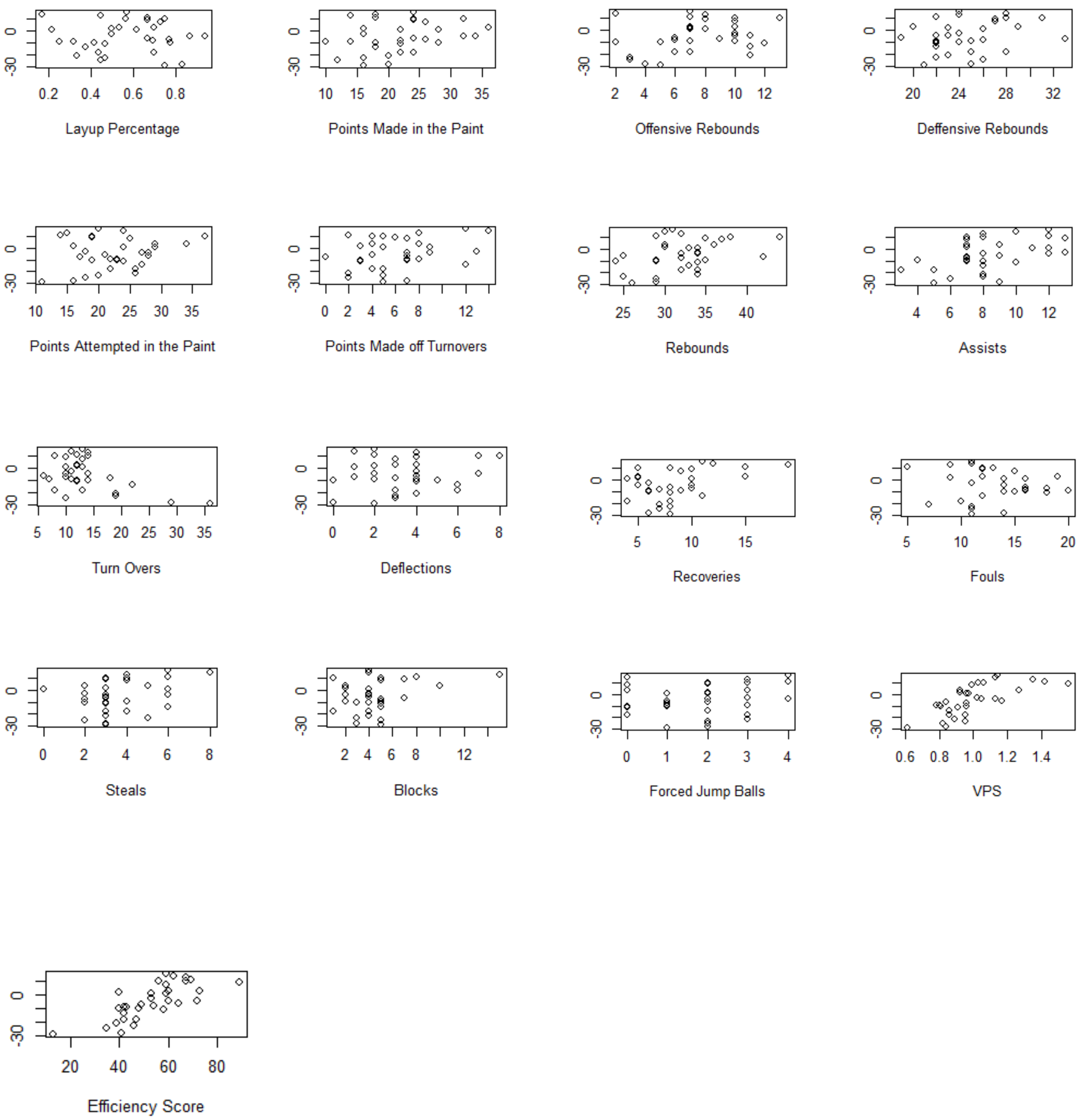



Nine
Ten
Thirteen
Fourteen
Eleven
Twelve
Fifteen
Sixteen
Opponent's Fifteen
Opponent's Sixteen
Two Pointers
Two Pointers Attempted
Opponent's Seventeen
Opponent's Eighteen
Opponent's Nine
Opponent's Ten
Two Pointer Percentage
Three Pointers
Three Pointers Attempted
Three Pointer Percentage
Field Goal Percentage
Effective Field Goal Percentage
Field Goals Made
Field Goals Attempted
Free Throws
Free Throws Attempted

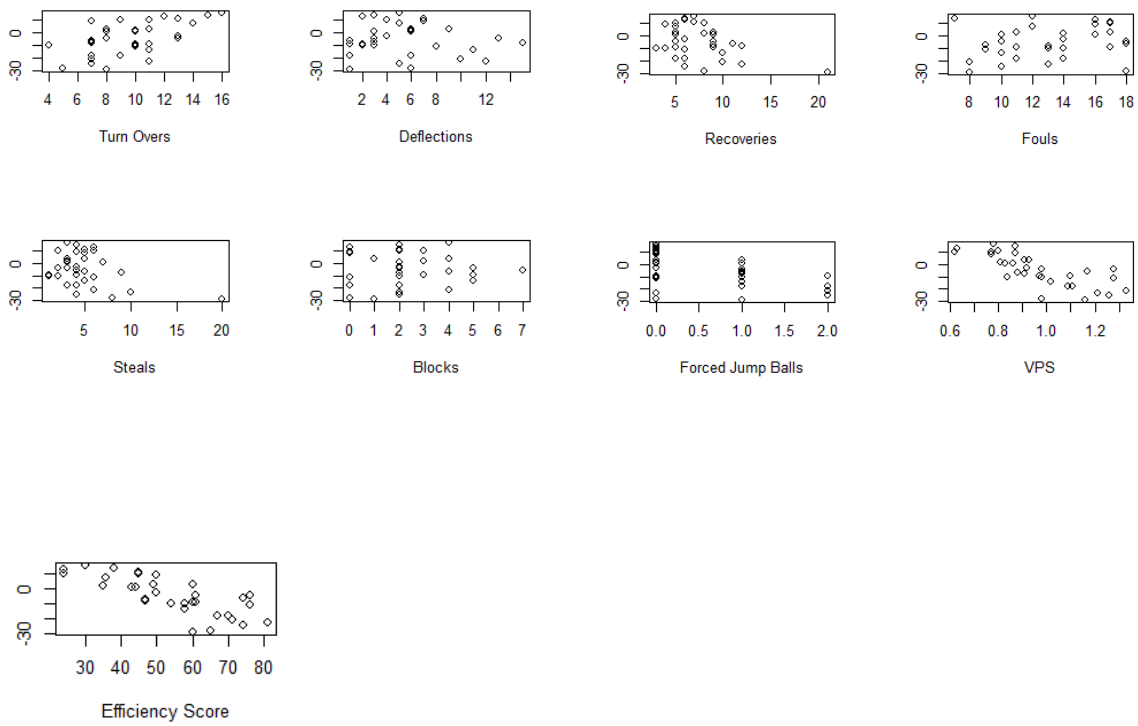


Vermont Statistics




Layup Percentage
Points Made in the Paint
Offensive Rebounds
Deffensive Rebounds
Points Attempted in the Paint
Points Made off Turnovers
Rebounds
Assists
Turn Overs
Deflections
Recoveries
Fouls
Steals
Blocks
Forced Jump Balls
VPS
Efficiency Score

Opponents’ Statistics



Appendix III

```
    proc glmselect data=full plot = all;
    model score = oppsixteen oppseventeen one fifteen FTA
opp3Pt oppFG oppPts oppTO Foul
/ selection=stepwise(select=AIC) stats=all;
    run;
```

Effects:	Intercept FTA oppFG oppTO

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	3	4145.60670	1381.86890	41.08
Error	28	941.89330	33.63905	
Corrected Total	31	5087.50000		

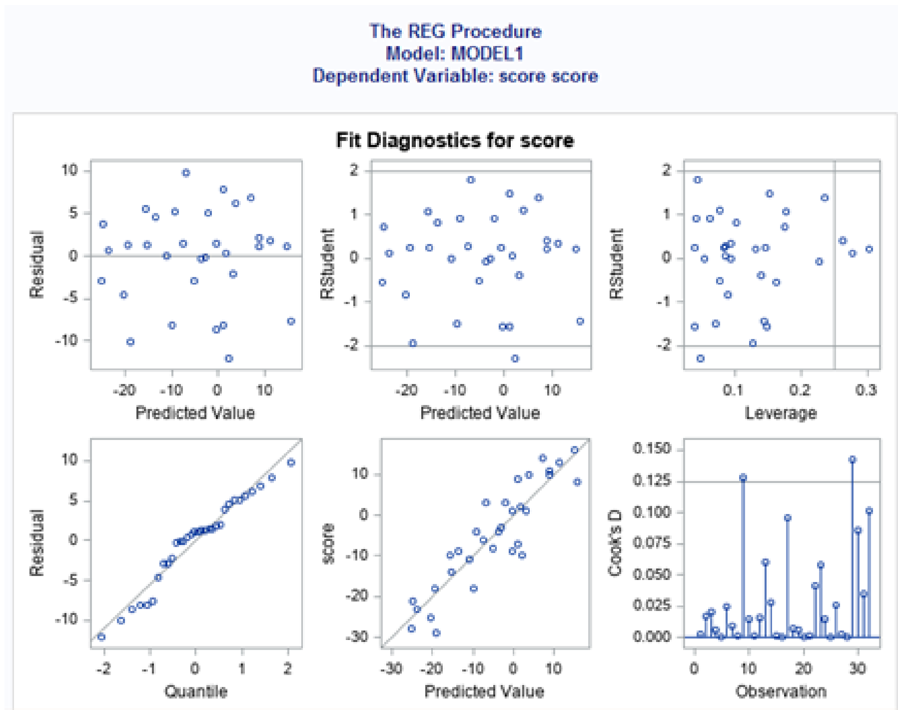
Root MSE	5.79992
Dependent Mean	-4.87500
R-Square	0.8149
Adj R-Sq	0.7950
AIC	150.22900
AICC	152.53669
BIC	120.32309
C(p)	0.96910
PRESS	1178.56254
SBC	122.09194
ASE	29.43417

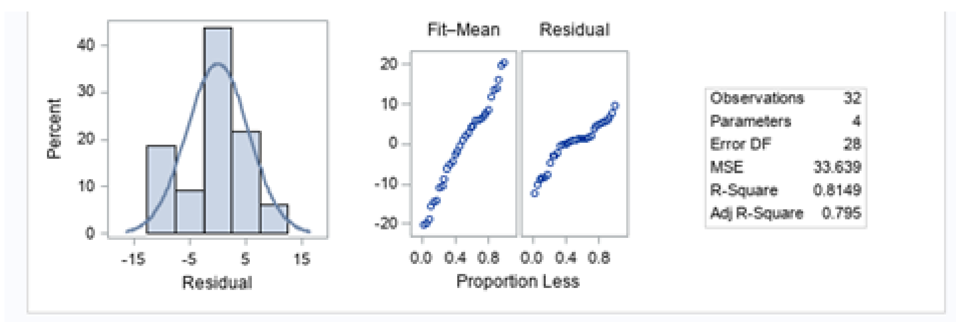
```
proc reg data = full;
model score = FTA oppFG oppTO/clb;
run;
```

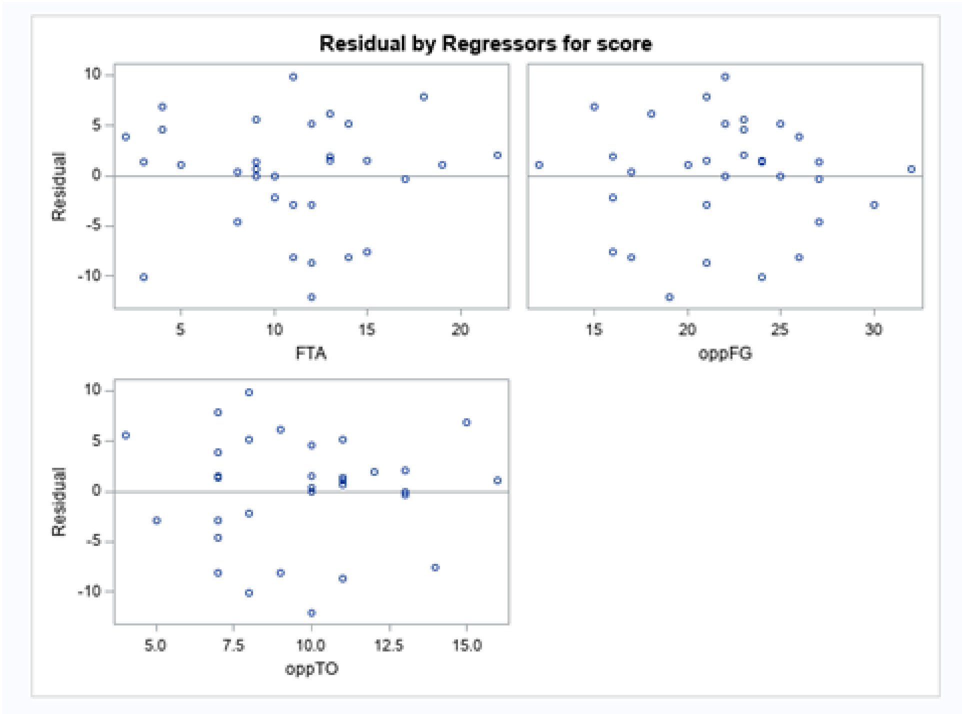
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4145.60670	1381.86890	41.08	<.0001
Error	28	941.89330	33.63905		
Corrected Total	31	5087.50000			

Root MSE	5.79992	R-Square	0.8149
Dependent Mean	-4.87500	Adj R-Sq	0.7950
Coeff Var	-118.97267		

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	12.40294	8.42149	1.47	0.1520	-4.84770	29.65359
FTA	FTA	1	1.05162	0.21725	4.84	<.0001	0.60660	1.49664
oppFG	oppFG	1	-1.83729	0.25415	-7.23	<.0001	-2.35789	-1.31668
oppTO	oppTO	1	1.20379	0.40722	2.96	0.0063	0.36963	2.03794


The REG Procedure
Model: MODEL1
Dependent Variable: score score
Fit Diagnostics for score
Residual
Predicted Value
RStudent
Predicted Value
RStudent
Leverage
Residual
Quantile
score
Predicted Value
Cook's D
Observation


Percent
Residual
Fit–Mean
Residual
Proportion Less
Observations 32
Parameters 4
Error DF 28
MSE 33.639
R-Square 0.8149
Adj R-Square 0.795




```
proc glmselect data = full plot = all;
    model score = one two three four five eight nine thirteen
fourteen fifteen sixteen seventeen eighteen oppone opptwo
oppthree oppfour oppfive oppeight oppnine oppten oppfourteen
oppfifteen oppsixteen oppseventeen oppeighteen
/ selection = stepwise(select = AIC) stats=all;
    run;
```

Effects:	Intercept three nine thirteen oppthree oppfour oppfive oppsixteen

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Value
Model	7	3015.96468	430.85210	5.23
Error	23	1893.71274	82.33534	
Corrected Total	30	4909.67742		

Root MSE	9.07388
Dependent Mean	-4.45161
R-Square	0.6143
Adj R-Sq	0.4969
AIC	176.48153
AICC	185.05296
BIC	176.53762
C(p)	-7.81244
PRESS	3112.14045
SBC	154.95343
ASE	61.08751

```
proc reg data = full;
model score = three nine thirteen oppthree oppfour oppfive oppsixteen
/clb;
run;
```

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	3047.55665	435.36524	5.12	0.0011
Error	24	2039.94335	84.99764		
Corrected Total	31	5087.50000			

Root MSE	9.21942	R-Square	0.5990
Dependent Mean	-4.87500	Adj R-Sq	0.4821
Coeff Var	-189.11623		

Parameter Estimates								
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	95% Confidence Limits	
Intercept	Intercept	1	10.77747	6.25945	1.72	0.0980	-2.14141	23.69634
three	three	1	-33.77019	11.43126	-2.95	0.0069	-57.36315	-10.17723
nine	nine	1	18.35661	7.28100	2.52	0.0187	3.32937	33.38385
thirteen	thirteen	1	13.85119	6.46215	2.14	0.0424	0.51396	27.18841
oppthree	oppthree	1	28.32481	14.16347	2.00	0.0570	-0.90715	57.55677
oppfour	oppfour	1	-10.08883	5.81935	-1.73	0.0958	-22.09937	1.92171
oppfive	oppfive	1	-20.57674	5.06436	-4.06	0.0004	-31.02907	-10.12441
oppsixteen	oppsixteen	1	-28.42837	7.71932	-3.68	0.0012	-44.36026	-12.49648