Variables Influencing Points Scored Per Game By UVA Basketball

Cliff Lin

March 1, 2020

Introduction and Research Question:

What are the most significant in-game factors that contribute towards points scored and thus win percentage per season in the NCAA?

College basketball is, by nature, a hard sport to predict, so we wanted to face the challenge of correctly predicting points scored per game by a team by using various in-game factors.[1] Moreover, our interest in this space was piqued by several articles including "How Much Do Statistics Matter In Determining Success In Sports?" *Forbes* and "Analytics in College Hoops: A New Type of March Madness", *Michigan State University*. Our dataset encompasses statistics of every NCAA basketball game from 2011 to 2019, but we decided to narrow our focus to UVA. Common metrics in college basketball, and those that we will be testing, relate to three pointers, field goals, rebounds, turnovers, and steals. We will complete exploratory data analysis to decide which in-game factors to focus on, and then we will create and test our hypothesized model to help answer our research question.

Data Summary

The data was collected from the kaggle database and was compiled by Nate Duncan, a reputable NBA sports researcher and analyst. The dataset consists of NCAA Regular Season Basketball Games, and includes every team and game dating back to 2011. Duncan created the dataset, and sourced the the data off the game logs from the website *sports-reference.com*. The dataset can be viewed here: www.kaggle.com/nateduncan/2011current-ncaa-basketball-games/data. While Duncan did not state the purpose of why he created this, it is likely that the dataset is used to analyze the past performance of teams and predicting their chances in the NCAA tournament. Each observation represents a single game and lists the statistics of that game, such as the date it occured, if the team won, how many points were scored, etc. Since our research question narrowed the focus to just UVA, the data was subsetted UVA games only. The "site" variable was re-coded to account for dummy variables, since the values home, away, and neutral are qualitative. They were divided into DumHome (Home - 1, Away - 0) and DumNeutral (Neutral - 1, Away - 0). Furthermore, four dummy variables were added to the dataset in SAS that record whether UVA had more blocks, steals, and rebounds, and fewer turnovers than their opponent.

Below is a data "dictionary" that includes the name of each variables, a brief description, and whether the variable was quantitative or categorical:

Name of Variable	*Description*	*Type (categorical or quantitative)*
pts	Number of points scored in the game by UVA	Quantitative

Site	Determines if the game was played home, away, or at a neutral location	Categorical
opp_pts	Number of points scored by opponent in the game	Quantitative
fg	Number of field goals scored in game	Quantitative
fga	Number of field goals attempted in the game	Quantitative
fg_per	Field goal percentage	Quantitative
3p	Number of three point shots made	Quantitative
3pa	Three point shots attempted	Quantitative
3p_per	Percentage of three point shots made	Quantitative
ft	Free throws made	Quantitative
fta	Free throws attempted	Quantitative
ft_per	Free throw percentage	Quantitative
orb	Offensive rebounds	Quantitative
trb	Total rebounds	Quantitative
ast	Number of Assists	Quantitative
stl	Number of steals	Quantitative
blk	Number of blocks	quantitative
tov	Number of turnovers	Quantitative
pf	Number of personal fouls	Quantitative
opp_fg	Opponent field goals scored	Quantitative

opp_fga	Opponent field goals attempted	Quantitative
opp_fg_ per	Opponent field goal percentage	Quantitative
opp_3p	Opponent three point shots scored	Quantitative
opp_3pa	Opponent three point shots attempted	Quantitative
opp_3p_per	Opponent three point percentage	Quantitative
opp_ft	Opponent free throws scored	Quantitative
opp_fta	Opponent free throws attempted	Quantitative
opp_ft_per	Opponent three throw percentage	Quantitative
opp_orb	Opponent offensive rebound	quantitative
opp_trb	Opponent total rebounds	quantitative
opp_ast	Opponent assists	quantitative
opp_stl	Opponent steals	quantitative
opp_blk	Opponent blocks	quantitative
opp_tov	Opponent turnovers	quantitative
opp_pf	Opponent personal fouls.	quantitative

Advantages and Disadvantages

Advantages:

This dataset possesses a large number of metrics recorded over 8 seasons of college basketball. Every game has the same metrics recorded which allows for easy comparison and standardization. The depth of the statistics recorded will allow for a thorough analysis and multiple linear regression of the data. Furthermore, since the data is sourced directly from in-game logs, it is unbiased and suitable for statistical analysis.

Disadvantages:

A disadvantage that arises when considering our data is that, with only analyzing UVA basketball, it would be difficult to extrapolate our findings to the ACC or the NCAA as a whole. Teams across the NCAA have varying styles of play, so using analysis of UVA, a team with a very unique playing style, to estimate other win percentages across the NCAA would probably not be feasible. Also, since every year brings a different UVA team, without a viable way of separating our analysis by year, we are considering that every UVA team functions similarly despite the fact that players come and go every year. Nevertheless, this disadvantage is mitigated by the tenure of UVA coach Tony Bennett during the entire time span of the dataset, who in fact is known to establish his style of play year after year.

Exploratory Data Analysis

After examining scatter plots for quantitative variables (*Appendix: Appendix: Figures 1-4*) and plots of means for qualitative variables (*Appendix: Figures 5-10*) and their corresponding relationship with points scored, we have identified the nature of each:

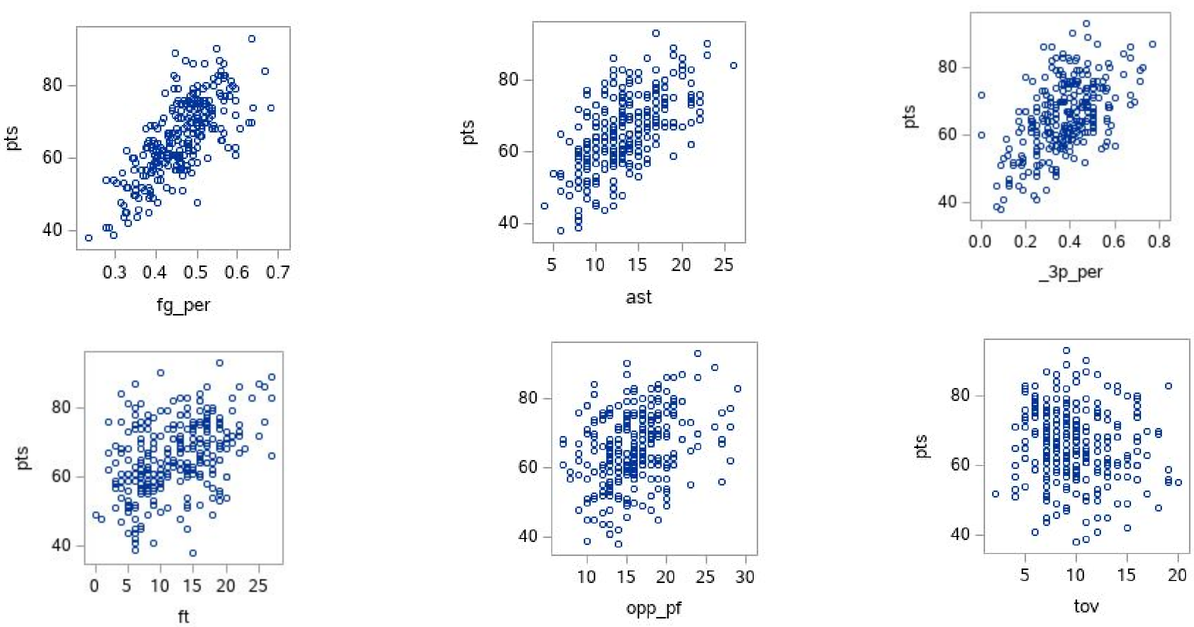
Positive Linear	**Negative Linear**	**Concave Down Curvilinear**	**No Relationship Observed**
Field Goals, Field Goal Percent, Field Goal Attempts, 3 Pointers, 3-point percent, Opponent Points, Free Throw Attempts, Assists, Opponent 3-point Attempts, DumTrb, DumStl, DumBlk, DumTov, DumHome, DumNeutral	Turnovers, Opponent Total Rebounds, Opponent Blocks	Free Throws, Opponent Personal Fouls	3-point attempts, Free Throw Percent, Offensive Rebounds, Total Rebounds, Steals, Blocks, Personal Fouls, Opponent Field Goals, Opponent Field Goals Attempted, Opponent Field Goal Percent, Opponent 3-pointers, Opponent 3-point Percent, Opponent Free Throws, Opponent Free Throw Attempts, Opponent Offensive Rebounds, Opponent Assists, Opponent Steals, Opponent Turnovers

We then selected six quantitative and three qualitative variables that appear to have the strongest relationship with points scored.

Quantitative:

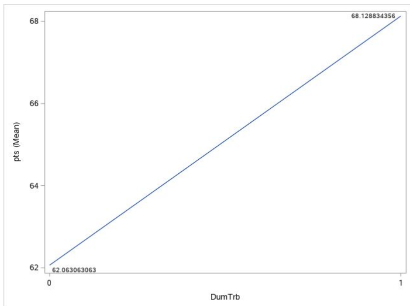
1. Field Goal Percentage - As UVA's field goal percentage increases, the total points scored in that game also appears to increase linearly in a positive relationship.
2. Assists - As the number of assists recorded by UVA increases, the total points scored in that game also appears to increase linearly in a positive relationship.
3. 3-Point percentage - As UVA's three point percentage increases, the total points scored in that game also appears to increase linearly in a positive relationship.
4. Free throws made - As UVA makes more free throws, their total points scored increases in a concave down curvilinear fashion.
5. Opponent Personal Fouls - As UVA's opponent records more personal fouls, their total points scored increases in a concave down curvilinear fashion.
6. Turnovers - As UVA's number of turnovers decrease, the total points scored in that game also appears to increase linearly in a negative relationship.

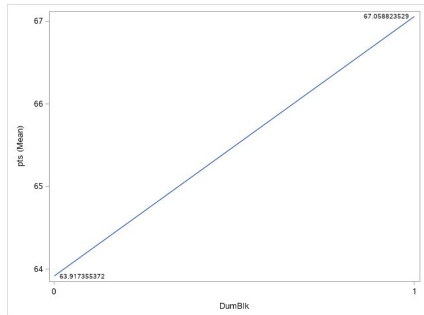
The corresponding scatter plots can be seen below:

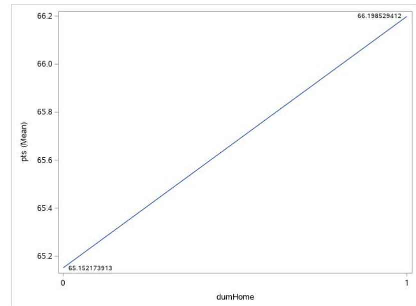


Qualitative:

1. DumTrb - UVA tends to score more points on average where they out-rebound their opponent.
2. DumBlk - UVA tends to score more points on average where they record more blocks than their opponent.
3. DumHome- UVA tends to score more points on average when they are at home rather than away.

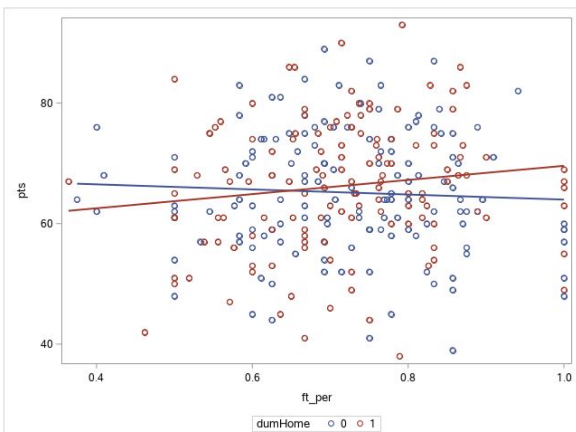


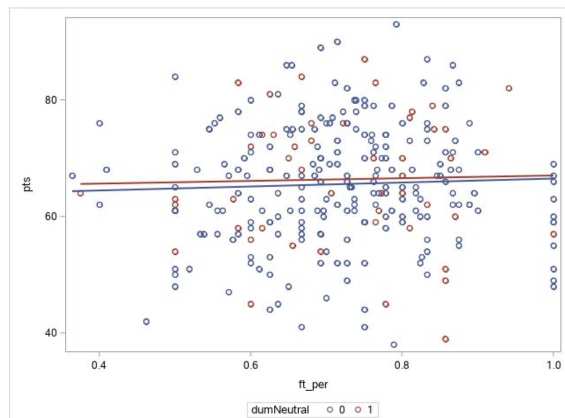




Interaction:

We also explored the interaction between free throw percentage (quantitative) and site (home vs away and neutral vs away) in order to determine whether this interaction has an impact on points scored.





These plots indicate that the ft_per and dumHome have a clear interaction, while ft_per and dumNeutral do not interact. This is indicated by the intersection pictured in the first graph while the second graph has lines that are nearly parallel and do not intersect.

Exploratory Proposed Model

$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ft} + \beta_5 (x_{ft}^2) + \beta_6 x_{ftPer} + \beta_7 x_{oppPf} + \beta_8 (x_{oppPf}^2) + \beta_9 x_{tov}$$
$$+ \beta_{10} x_{DumHome} + \beta_{11} x_{DumTrb} + \beta_{12} x_{DumBlk} + \beta_{13} x_{dumHome} x_{ftPer}$$

Initial Analysis

A rejection region of *p-value > 0.05* will be used throughout the analysis process. We will begin with all the variables of interest determined from EDA, and use a combination of Global F-tests, Nested F-tests, and T-tests to find which are most suitable for multiple linear regression.

Stage 1 - Quantitative and Higher Order Quantitative Terms

$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ft} + \beta_5 (x_{ft}^2) + \beta_6 x_{ftPer} + \beta_7 x_{oppPf} + \beta_8 (x_{oppPf}^2) + \beta_9 x_{tov}$$

We first conducted a test on overall significance using the Global F-test:
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$
H_a: At least one of $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9 \neq 0$
Test statistic: F = 104.99
p-value: < 0.0001
Output: *Appendix: Figure 11*

Since the p-value for the model that included all quantitative predictors is less than the significance level (0.05 > 0.0001), we reject the null hypothesis. Mean Squared Error, MSE, for this model was 24.64 and the adjusted R^2 value was 77.42%, which indicates that the predictors in this model explain about 77.42% of the variation in points scored. Therefore the overall model appears to be statistically significant in predicting UVA's number of points scored per game. In running this test, we were also able to obtain relevant t-test values for each predictor, including the curvilinear terms. As a result, both $oppPf^2$ and ft^2 did not display significance at the 0.05 alpha level. Therefore, these terms can be dropped from the model. Although ftPer also did not display significance individually, it will be retained at this point as we will test its interaction in a later stage.
End Stage 1 Model:

$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ftPer} + \beta_5 x_{tov} + \beta_6 x_{ft} + \beta_7 x_{oppPf}$$

Stage 2 - Adding qualitative variables

$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ftPer} + \beta_5 x_{tov} + \beta_6 x_{ft} + \beta_7 x_{oppPf} + \beta_8 x_{dumHome} + \beta_9 x_{dumTrb} + \beta_{10} x_{dumBlk}$$

In order to determine the significance of adding the qualitative variables to the model, first a Global F-test will be performed to determine the overall significance of the model.

Global F-test:
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$
H_a: at least one of $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10} \neq 0$
Test statistic: F = 95.07
p-value: < 0.0001
Output: *Appendix: Figure 12*

Since the p-value for the model is less than the significance level (0.05 > 0.0001), we reject the null hypothesis. Mean Squared Error, MSE, for this model was 24.55 and the adjusted R^2 value was 77.51%, which indicates that the predictors in this model explain about 77.51% of the variation in points scored. This is slightly higher than the Stage 1 Model.

Next, to test the contribution of just the added qualitative terms, a nested F-test will be used. We will compare the reduced model from the end of Stage 1 and the complete model from the beginning of Stage 2, shown above. If the Mean Squared Error from the complete model is significantly less than that of the reduced model, it will be used as a more accurate predictor of points scored per game.

Nested F-test:
H_0: $\beta_8 = \beta_9 = \beta_{10} = 0$
H_a: at least one of β_7, β_8, $\beta_9 \neq 0$
Test statistic: $F = 1.53$
p-value: 0.2061
Output: *Appendix: Figure 12*

Since the p-value for the alternative model is greater than the significance level (0.05 < 0.2061), we fail to reject the null hypothesis, therefore the alternative, complete model which holds the qualitative terms is not significant in predicting UVA's number of points scored per game. As a result, the dummy variables for blocks and total rebounds are not included in the model. However, we will retain the dummy variable for home vs away, dumHome, as we plan to use it in an interaction term in the next stage.

End Stage 2 Model:
$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ftPer} + \beta_5 x_{tov} + \beta_6 x_{ft} + \beta_7 x_{oppPf} + \beta_8 x_{tov} + \beta_9 x_{dumHome}$$

Stage 3 - Adding Interaction terms
$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ftPer} + \beta_5 x_{tov} + \beta_6 x_{ft} + \beta_7 x_{oppPf} + \beta_8 x_{tov} + \beta_9 x_{dumHome} + \beta_{10} x_{dumHome} x_{ftPer}$$
We first conducted a test on overall significance using the Global F-test:
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = 0$
H_a: At least one of β_1, β_2, β_3, β_4, β_5, β_6, β_7, β_8, β_9, $\beta_{10} \neq 0$
Test statistic: $F = 104.37$
p-value: < 0.0001
Output: *Appendix: Figure 13*

Since the p-value for the model that included all quantitative predictors is less than the significance level (0.05 > 0.0001), we reject the null hypothesis. Mean Squared Error, MSE, for this model was 24.76 and the adjusted R^2 value was 77.31%, which indicates that the predictors in this model explain about 77.31% of the variation in points scored. Therefore the overall model appears to be statistically significant in predicting UVA's number of points scored per game. In running this test, we were also able to obtain relevant t-test values for each predictor, including the interaction term. The interaction between dumHome and Free Throw percent did not display significance at the 0.05 alpha level. Therefore, this term, dumHome, and ft_per can be dropped from the model, as the significance of the last two were shown to not be sufficient above. Furthermore, in looking at the individual t-values, oppPf, which has been retained in the model throughout the process, is not a significant predictor at the 0.05 alpha level and will be dropped before forming the final model.

Final Proposed Model

$$E(y) = \beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ft} + \beta_5 x_{tov}$$

Global F-test:

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

H_a: At least one of $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \neq 0$

Test statistic: 188.93

P-value: < 0.0001

Output: *Appendix: Figure 14*

Since the p-value for the model is less than the significance level (0.05 > 0.0001), we reject the null hypothesis, therefore the overall model is significant in predicting UVA's number of points scored per game.

	Stage 1	Stage 2	Stage 3
End Stage Model	Mean Points Scored = $\beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ft} + \beta_5 x_{ftPer} + \beta_6 x_{tov} + \beta_7 x_{oppPf}$	Mean Points Scored = $\beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ft} + \beta_5 x_{ftPer} + \beta_6 x_{tov} + \beta_7 x_{dumHome} + \beta_8 x_{oppPf}$	Mean Points Scored = $\beta_0 + \beta_1 x_{fgPer} + \beta_2 x_{ast} + \beta_3 x_{3pPer} + \beta_4 x_{ft} + \beta_5 x_{tov}$
Model Utility	Global F-Test, Significant	Global F-Test, Significant	Global F-Test, Significant
MSE	24.55	24.76	24.57
Adj R-Squared	77.51%	77.31%	77.49%

We chose this model since all the variables are significant in predicting the number of points UVA scores per game, as opposed to the earlier models which had several variables and interactions that did not contribute towards the model.

Appendix

SAS Code:

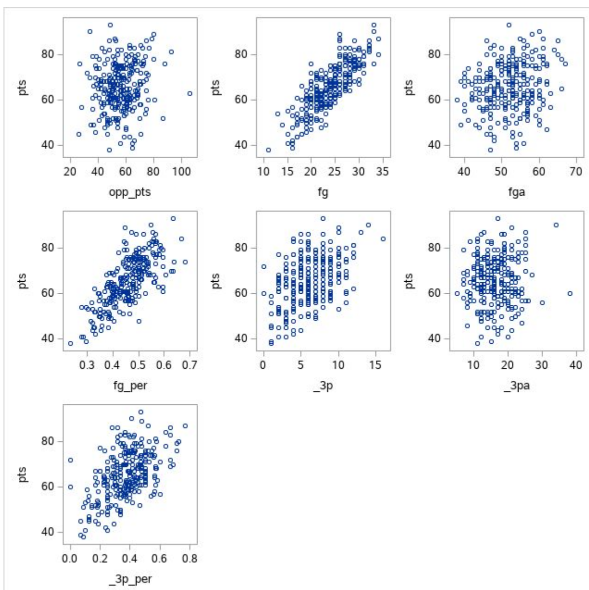
```
filename newdata '/folders/myfolders/DATA/gamesUVA.csv';

 proc import datafile=newdata
        dbms=csv
        out=mydata.gamesUVA
        replace;

     getnames=yes;
run;

*proc contents allows you to ensure your data uploaded as desired;

proc contents data=mydata.gamesUVA;
run;

* setting dummy variables;
data mydata.gamesUVA1 ; *create a new data set saved to your library;
    set mydata.gamesUVA; *set with the original table ;
    DumTrb = 0;
    if trb > opp_trb then DumTrb = 1;

    DumStl = 0;
    if stl > opp_stl then DumStl = 1;

    DumBlk = 0;
    if blk > opp_blk then DumBlk = 1;

    DumTov = 0;
    if tov < opp_tov then DumTov = 1;

run;
```

```
* EDA, multiple scatter plots were made since they would not all fit on the same page;
proc sgscatter data=mydata.gamesUVA1;
plot pts * (opp_pts fg fga fg_per _3p _3pa _3p_per);
run;

proc sgscatter data=mydata.gamesUVA1;
plot pts * (ft fta ft_per orb trb ast stl blk tov);
run;


proc sgscatter data=mydata.gamesUVA1;
plot pts * (pf opp_fg opp_fga opp_fg_per opp_3p opp_3pa opp_3p_per
opp_ft opp_fta);
run;

proc sgscatter data=mydata.gamesUVA1;
plot pts * (opp_ft_per opp_orb opp_trb opp_ast opp_stl opp_blk opp_tov opp_pf);
run;

* for dummy variables;

proc sgplot data = mydata.gamesUVA1;
vline dumTrb/ response = pts stat=mean datalabel;
run;

proc sgplot data = mydata.gamesUVA1;
vline dumStl/ response = pts stat=mean datalabel;
run;
```

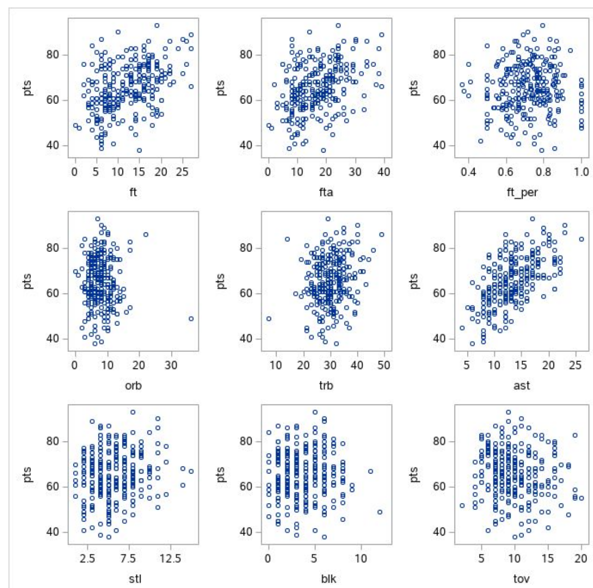
```
proc sgplot data = mydata.gamesUVA1;
vline dumBlk/ response = pts stat=mean datalabel;
run;

proc sgplot data = mydata.gamesUVA1;
vline dumTov/ response = pts stat=mean datalabel;
run;

* creating interactions;
data mydata.gamesUVA2 ; *create a new data set saved to your library;
    set mydata.gamesUVA1; *set with the original table ;
    dumHome = 0;
    if site = "Home" then dumHome = 1;
    dumNeutral = 0;
    if site = "Neutral" then dumNeutral = 1;
    homeFt = dumHome * ft_per;
    neutralFt = dumNeutral * ft_per;

    ftsqr = ft * ft;    * higher order variable;

run;
data mydata.gamesUVA2 ; *create a new data set saved to your library;
    set mydata.gamesUVA2; *set with the original table ;

    oppPfsqr = opp_pf * opp_pf; * higher order variable;
run;

* line plot of means for dummy site variables;
proc sgplot data = mydata.gamesUVA2;
vline dumHome/ response = pts stat=mean datalabel;
run;
```

```
proc sgplot data = mydata.gamesUVA2;
vline dumNeutral/ response = pts stat=mean datalabel;
run;

* plotting interactions;
proc sgplot data = mydata.gamesUVA2;
vline fg / group = dumHome response = pts stat=mean datalabel;
run;

* interaction of home vs. away;
proc sgplot data=mydata.gamesUVA2;
scatter y=pts x=ft_per/group=dumHome;
reg y=pts x=ft_per/group=dumHome;
run;

* interaction of neutral vs. away;
proc sgplot data=mydata.gamesUVA2;
scatter y=pts x=ft_per/group=dumNeutral;
reg y=pts x=ft_per/group=dumNeutral;
run;

* EDA Model;
* first stage, just quantitative variables;
proc reg data=mydata.gamesUVA2 plots=none;
model pts = fg_per ast _3p_per ft ftsqr ft_per opp_pf oppPfsqr tov;
run;
```

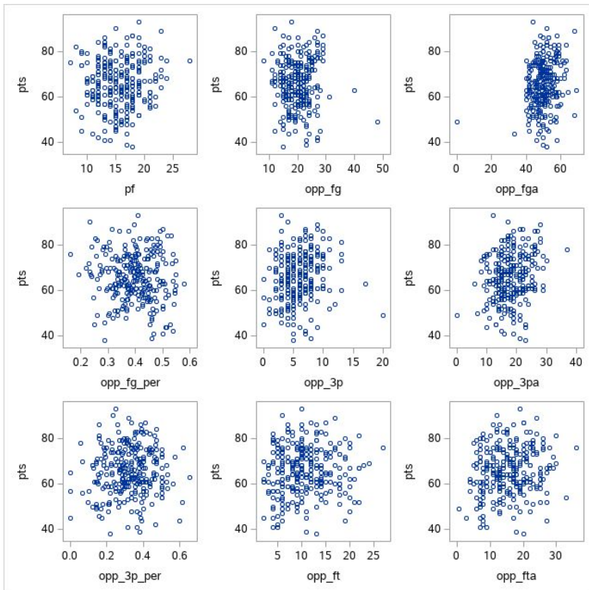
```
* second stage, qualitative variables;
proc reg data=mydata.gamesUVA2 plots=none;
model pts = fg_per ast _3p_per ft ft_per opp_pf tov dumHome dumTrb dumBlk;

run;

* nested f-test;
proc reg data=mydata.gamesUVA2 plots=none;
model pts = fg_per ast _3p_per ft ft_per opp_pf tov dumHome dumTrb dumBlk;
test dumHome, dumTrb, dumBlk;
run;

* third stage, interaction terms;
proc reg data=mydata.gamesUVA2 plots=none;
model pts = fg_per ast _3p_per ft ft_per opp_pf tov dumHome homeFt;
run;

* final model;
proc reg data=mydata.gamesUVA2 plots=none;
model pts = fg_per ast _3p_per ft tov ;
run;
```



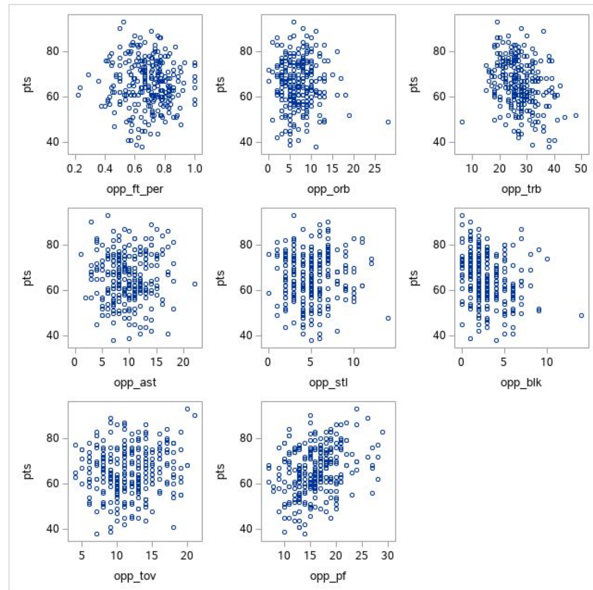
Appendix: Figure 1



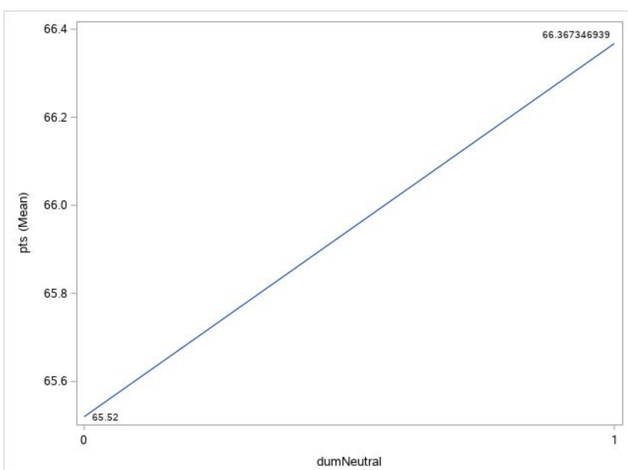
Appendix: Figure 2



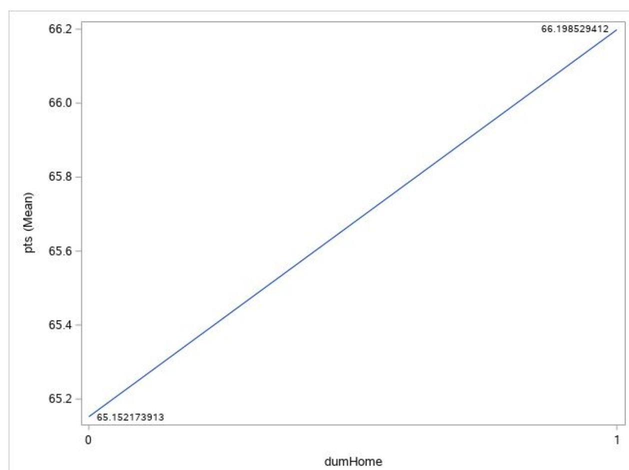
Appendix: Figure 3



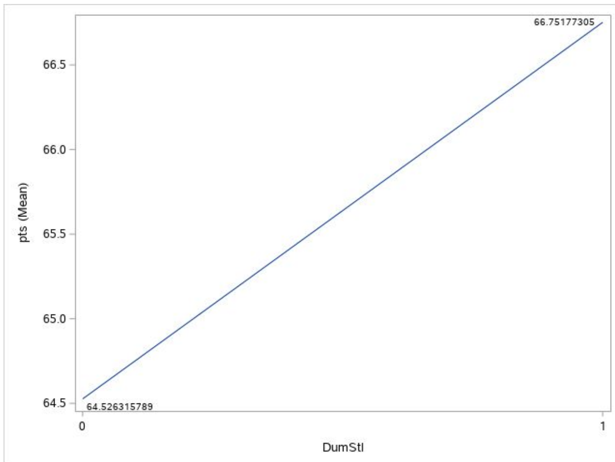
Appendix: Figure 4



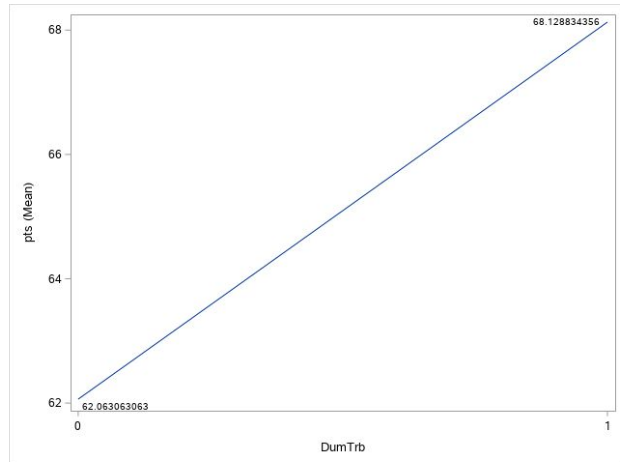
Appendix: Figure 5



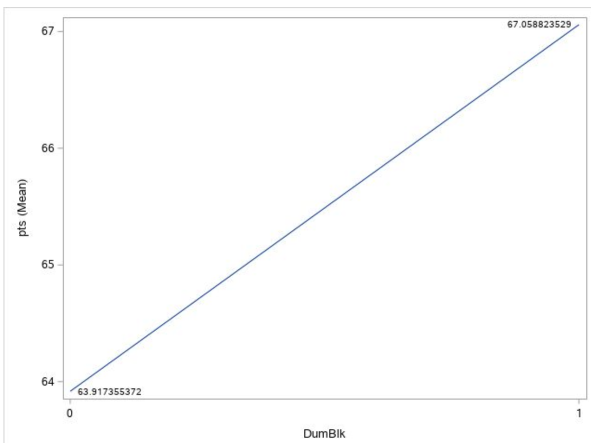
Appendix: Figure 6



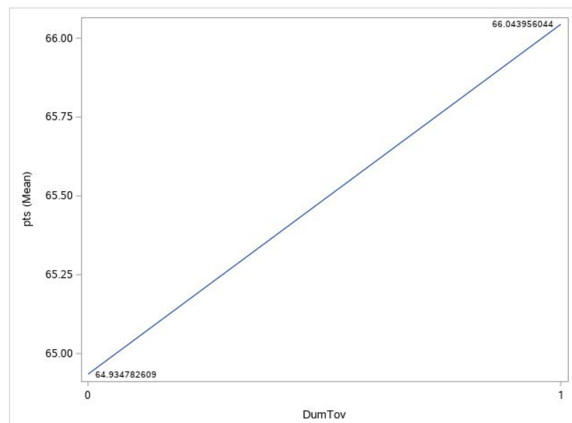
Appendix: Figure 7



Appendix: Figure 8



Appendix: Figure 9



Appendix: Figure 10

The REG Procedure
Model: MODEL1
Dependent Variable: pts

Number of Observations Read	275
Number of Observations Used	274
Number of Observations with Missing Values	1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	23290	2587.74466	104.99	<.0001
Error	264	6506.73598	24.64673		
Corrected Total	273	29796			

Root MSE	4.96455	R-Square	0.7816
Dependent Mean	65.67153	Adj R-Sq	0.7742
Coeff Var	7.55966		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	14.91094	4.64039	3.21	0.0015
fg_per	1	71.10657	5.23528	13.58	<.0001
ast	1	0.62916	0.10032	6.27	<.0001
_3p_per	1	11.00792	2.55997	4.30	<.0001
ft	1	0.24950	0.26604	0.94	0.3492
ftsqr	1	0.01575	0.00987	1.60	0.1117
ft_per	1	1.19454	2.69526	0.44	0.6580
opp_pf	1	0.57034	0.50048	1.14	0.2555
oppPfsqr	1	-0.01434	0.01433	-1.00	0.3180
tov	1	-0.63402	0.09234	-6.87	<.0001

Appendix: Figure 11

The REG Procedure
Model: MODEL1
Dependent Variable: pts

Number of Observations Read	275
Number of Observations Used	274
Number of Observations with Missing Values	1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	23340	2333.96087	95.07	<.0001
Error	263	6456.82926	24.55068		
Corrected Total	273	29796			

Root MSE	4.95486	R-Square	0.7833
Dependent Mean	65.67153	Adj R-Sq	0.7751
Coeff Var	7.54492		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	17.68345	3.34490	5.29	<.0001
fg_per	1	68.49434	5.53873	12.37	<.0001
ast	1	0.65065	0.10080	6.46	<.0001
_3p_per	1	11.32636	2.60449	4.35	<.0001
ft	1	0.64647	0.09904	6.53	<.0001
ft_per	1	0.59037	2.70415	0.22	0.8274
opp_pf	1	0.10810	0.13174	0.82	0.4126
tov	1	-0.64073	0.09236	-6.94	<.0001
dumHome	1	-0.60108	0.63231	-0.95	0.3427
DumTrb	1	1.11362	0.67001	1.66	0.0977
DumBlk	1	-0.61250	0.64892	-0.94	0.3461

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable pts				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	37.65533	1.53	0.2061
Denominator	263	24.55068		

Appendix: Figure 12

The REG Procedure
Model: MODEL1
Dependent Variable: pts

Number of Observations Read	275
Number of Observations Used	274
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	23259	2584.36802	104.37	<.0001
Error	264	6537.12581	24.76184		
Corrected Total	273	29796			

Root MSE	4.97613	R-Square	0.7806
Dependent Mean	65.67153	Adj R-Sq	0.7731
Coeff Var	7.57730		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	17.56469	3.74944	4.68	<.0001
fg_per	1	70.70973	5.23722	13.50	<.0001
ast	1	0.64296	0.10116	6.36	<.0001
_3p_per	1	10.87575	2.58659	4.20	<.0001
ft	1	0.64743	0.09902	6.54	<.0001
ft_per	1	0.10200	3.53315	0.03	0.9770
opp_pf	1	0.10347	0.13232	0.78	0.4349
tov	1	-0.62761	0.09249	-6.79	<.0001
dumHome	1	-2.26303	3.46476	-0.65	0.5142
homeFt	1	2.26902	4.76484	0.48	0.6343

Appendix: Figure 13

The REG Procedure
Model: MODEL1
Dependent Variable: pts

Number of Observations Read	275
Number of Observations Used	274
Number of Observations with Missing Values	1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	23211	4642.25690	188.93	<.0001
Error	268	6585.15345	24.57147		
Corrected Total	273	29796			

Root MSE	4.95696	R-Square	0.7790
Dependent Mean	65.67153	Adj R-Sq	0.7749
Coeff Var	7.54811		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	18.30357	1.96577	9.31	<.0001
fg_per	1	70.64309	5.21224	13.55	<.0001
ast	1	0.62489	0.09930	6.29	<.0001
_3p_per	1	11.30358	2.54922	4.43	<.0001
ft	1	0.70915	0.05326	13.31	<.0001
tov	1	-0.62222	0.09178	-6.78	<.0001

Appendix: Figure 14

Works Cited

Nate Duncan. "NCAA Regular Season Basketball Games." *Kaggle*, 6 Jan. 2019, www.kaggle.com/nateduncan/2011current-ncaa-basketball-games/data.

"Sports Reference: Sports Stats, Fast, Easy, and up-to-Date." *Sports*, www.sports-reference.com/.

"Nate Duncan - MIT Sloan Analytics Conference." *MIT Sloan Sports Analytics Conference*, www.sloansportsconference.com/people/nate-duncan/.