The Factors of Making the Division I NCAA Basketball Tournament

Jackson Nahom and Jason Phelps

1 Introduction

The goal of this paper is to explore the possible variables that would explain why teams make the NCAA basketball tournament. In collecting large historical data sets of NCAA basketball records, we can conduct statistical analysis on this data to discover what factors influence teams making the NCAA tournament. The source of our data has been collected from a well known-sports statistics site known as sports-references.com. This site is managed by a company known as Sports Reference whose purpose is to provide both basic and sabermetric statistics of historical and current sports and sporting events through their websites[1].
We were able to obtain this data through two means; the first was downloading historical data of schools in CSV. The second method involved web scraping because it would have taken too long to individually download 440 schools season by season statistics. We wrote a python script that retrieved each school's seasonal webpage and downloaded them. We wrote another script that parsed through those webpages to create a master CSV. We then removed any season before 1939 the year that the NCAA tournament began. Although there were 440 schools we only used 399 schools. This was due to some schools never playing division I basketball while the NCAA tournament existed (1939 onward).
To better organize schools in a way that people can visualize, we mapped schools with the respective regions they are from. To this, we used information from the United States Census Bureau. The Census statistically maps regions into statistical groupings to summarize the statistics of the data they gather every 10 years[2].

The variables we were able to work within both these datasets are located in the chart below:

Variable Name	Variable Full Name	Variable Type
School	School Name	Text
Season	Season	Years
conf_abbrev	Conference Abbreviation	Text
wins	Wins	Integer
losses	Losses	Integer
win_loss_pct	Win-Loss Percentage	Percentage/Decimal
wins_conf	Conference Wins	Integer
losses_conf	Conference Losses	Integer
win_loss_pct_conf	Conference Win-Loss Percentage	Percentage?Decimal
srs	Simple Rating System	Decimal
sos	Strength of Schedule	Decimal

pts_per_g	Points Per Game	Decimal
opp_pts_per_g	Opponent Points Per Game	Decimal
rank_pre	Pre-season Rank	Integer or NULL
rank_mid	Mid-Season Rank	Integer or NULL
rank_final	Rank Final	Integer or NULL
seed	Seed	Integer or NULL
coaches	Coach	Text
makeTournament	Made Tournament Value(1or0) 1=yes, 0=no	Binary Value (1,0)
state	State	Text
region	Region	Text
division	Division	Text

In this paper, we will investigate what variables have interaction with collegiate basketball programs making the NCAA tournament. We intend to use statistical methods to do this exploratory statistical investigation. Using the above variables we will see which ones are impactful, unimportant, or ambiguous in making the coveted NCAA basketball tournament; finding what is statistically associated with this analysis.

2. Statistical Methods

2.1 Multiple Linear Regression

Multiple linear regression allows for a response variable y_i to be modeled as a linear function with more than one input variable x_{ki}. This is modeled as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i \tag{1}$$

Where β is the coefficient of the variable, and β_0 is the $y - intercept$ of the regression. K is the number of variables in the regression while $1 \leq i \leq n$ an individual row of input variables as n is the number of rows inputted. Every row of inputs of the input variables has a standard error term ε.

2.2 Binomial Logistic Regression

When the response variable y_i is a binary 1 or 0 multiple linear regression is no longer usable. Instead what is needed is binary logistic regression. This does not assume a linear relationship and large samples are needed, with a minimum of 50 total inputs. While using binomial logistic regression we view 1 as a success and 0 as a failure, not necessarily true depending on your y output. For example, 1 could mean left-handed and 0 could mean right-handed. However, for our analysis, this is true with 1 meaning a school made the NCAA tournament and 0 it did not make the tournament. This success probability is shown by the rate of change $p(x)$:

$$p(x) = 1 - \frac{1}{1+e^{a+bx}} \tag{2}$$

The rate of change of $p(x)$ depends on x and is the largest when $p(x)$ is near 0.5. If we define O(x) to be odds for success when the analysis is run at level x, then

$$O(x) = e^{a+bx} \tag{3}$$

When $b > 0$ the odds increase exponentially in the input variable x; when $b < 0$ the odds decrease exponentially in the input variable x.

$$log[O(x)] = a + bx \tag{4}$$

This will be used to calculate a and b which are unknown and need to be estimated. We accomplish this estimate for a and b by using the maximum likelihood approach. The analysis is performed for levels of $x_1, x_2, ..., x_{k,}$ and y_i is the result. For example, 1 if success and 0 if failure. Now using the Bernoulli Function we can determine the Likelihood Function:

$$P(Y_i = y_i,\ i = 1, 2, ..., k) = \prod_{i=1}^{k} (\frac{e^{a+bx_i}}{1+e^{a+bx_i}})^{y_i} (\frac{1}{1+e^{a+bx_i}})^{1-y_i} \tag{5}$$

Using equation 5 we can run binomial logistic regression.

2.3 Chi-Squared Test

A Chi-Squared is often used to discover differences in categorical data as it pertains to one population. There can be several different usages of the chi-squared test. In our case, we wanted to compare proportions of schools on a regional basis to the total NCAA appearances made by schools in that region. In theory, a competitive NCAA would consist of a tournament with equal representation of schools from each region based on the size of each region, so that there is an equal proportion of schools in the tournament from each region. To do this, we needed to calculate proportions of total competing schools on a regional level for our hypothesis, and use that as a marking for determining if the proportions of tournament appearances match that expected proportion of tournament appearances. Utilizing the chi-squared formula

$$\chi^2 = \Sigma \frac{(Observed - Expected)^2}{Expected} \tag{6}$$

we can calculate a chi-squared value and determine its significance when compared to a chi-square tabulated value from k variable samples. If $\chi^2_{calculated} > \chi^2_{tabulated}$ then it would be statistically significant.

2.4 Two-Way Population Proportion Comparison

In looking at our data set as a whole, one thing was clear. Breaking the data down into a school by school basis and offering a probability of an individual school making it to the tournament would require programming skills and data cleansing beyond our capability. One thing that was evident in the data, is that evidence of regional influence on NCAA appearances chances to teams within that region could be valued. The method involved in determining if the region a school is classified in effects a school in that region's chance of making it to the NCAA tournament is a Two-Way Comparison test. This is a form of discrete data analysis that compares two population proportions to determine a difference denounced by the hypothesis test.

For this test to be conducted, both populations have to have a set population sample size and a recorded observed value to the sample size. With this information, we can determine a pooled estimate of success probability. The equation $p = \frac{x+y}{n+m}$ is used to calculate the pooled success probability. Where x and y are the observed values for the populations and n and m are the respective sample population sizes. Using this pooled success probability we can conduct a z-test with the following formula:

$$z = \frac{pA - pB}{\sqrt{p(1-p)(\frac{1}{n}+\frac{1}{m})}} \tag{7}$$

As with any z-test, once we get our z-calculated value with the formula, we then have to determine whether our null is to be rejected or accepted. We get a $Z\alpha/_2$ from a z-score table and compare it to our calculated z-score. We would accept the null hypothesis if $|z| \leq Z\alpha/_2$ and furthermore reject the null hypothesis if $|z| > Z\alpha/_2$. We can also look at the p-value and depending on our null hypothesis conduct the right calculation based on the hypothesis to determine if our p-value offers significant evidence.

3. Results

3.1 Multi Linear Regression (Historical Dataset)

Table 1

	Regressions					
Variable	1	2	3	4	5	6
Intercept	11.55*** (2.03)	10.41*** (2.06)	-12.76*** (1.79)	1.18 (1.95)	-0.44 (1.67)	11.25*** (3.18)
W-L%	-11.20*** (3.80)	-8.32** (3.85)	29.78*** (3.73)	6.21* (3.22)	10.16*** (3.30)	0.78 (5.83)
CREG	0.5378*** (0.0454)	0.6602*** (0.0384)	0.8419*** (0.0462)	0.5751*** (0.0449)	0.7373*** (0.0380)	--
SRS	0.7831*** (0.0485)	0.7649*** (0.0384)	--	--	--	1.194*** (0.146)
CRTN	0.3790*** (0.0799)	--	--	0.4972*** (0.0810)	--	--
SOS	--	--	--	1.0037*** (0.0672)	0.9694*** (0.0665)	-0.251 (0.198)
Adjusted R-sq	0.7988	0.7878	0.6612	0.7980	0.7792	0.6307
SER	4.66218	4.79748	6.04962	4.67165	4.88401	6.31615
n	399	399	399	399	399	399

* With 90% confidence, ** With 95% confidence, *** With 99% confidence
SER - Standard Error of Regression

3.2 Binary logistic regression (Each season dataset)

Table 2

	Regressions							
Variable	1	2	3	4	5	6	7	8
Intercept	-10.640*** (0.479)	-9.803*** (0.303)	-2.263*** (0,297)	-2.063*** (0.287)	-10.069*** (0.366)	-5.365*** (.324)	-2.266*** (0.259)	-9.303*** (0.264)
Win loss %	8.347*** (0.667)	6.963*** (0.620)	--	--	14.186*** (0.420)	--	--	11.701*** (0.488)
Wins conf	0.1068*** (0.016)	--	--	--	--	--	--	--
Win loss % conference	3.939*** (0.401)	5.694*** (0.330)	--	--	--	4.921*** (0.209)	--	--
Pts per game	-0.1295*** (0.0149)	--	0.16460*** (0.00716)	0.30580*** (0.00508)	0.01438 (0.00989)	0.22338*** (0.00789)	0.31104*** (0.00588)	--
Opp pts per game	0.1314*** (0.0152)	--	-0.17291*** (0.00764)	-0.31797*** (0.00670)	-0.0182* (0.00101)	-0.22292*** (0.00814)	-0.31799*** (0.00620)	--
SRS	0.1840*** (0.00631)	0.0795*** (0.0160)	0.15593*** (0.00471)	--	--	--	--	0.1030*** (0.0111)
SOS	--	0.1158*** (0.0160)	--	0.13712*** (0.00508)	--	--	--	0.0440*** (0.0136)
Adjusted R-sq	0.5181	0.5137	0.4156	0.3864	0.4182	0.3919	0.3310	0.4818
n	15527	16076	17439	17670	17439	15690	17670	18214

* With 90% confidence, ** With 95% confidence, *** With 99% confidence
SER - Standard Error of Regression

3.3 Chi-Squared Analysis

Table 3 School Proportions

	Midwest	**Northeast**	**South**	**West**	**Total**
# of Schools	79	81	176	63	399
Proportion	0.198	0.203	0.441	0.158	1

Table 4 Chi-Squared Table

Expected values in parenthesis

	Midwest	**Northeast**	**South**	**West**	**Total**
# of Tournament appearances	807 (660.528)	560 (677.208)	1363 (1471.176)	606 (527.088)	3336

Hypothesis Test 1

$H_0 : p_{Midwest} = 19.8\%, p_{Northeast} = 20.3\%, p_{South} = 44.1\%\ ,\ p_{West} = 15.8\%$

H_a : At least one is not equal

Chi-Squared: $\chi^2 = \Sigma \frac{(Observed - Expected)^2}{Expected}$

Chi-Squared: $\chi^2 = 72.5343 > \chi^2_{.05,3} = 7.815,\ reject\ H_0\ at\ 95\%\ confidence$

3.4 Two-Way Comparison Tests

Hypothesis Test 2

$H_0 : p_{region\ 1} = p_{region\ 2}\ or\ p_{region\ 1} - p_{region\ 2} = 0$

$H_a : p_{region\ 1} \neq p_{region\ 2}\ or\ p_{region\ 1} - p_{region\ 2} \neq 0$

Hypothesis Test 3

$H_0 : p_{region\ 1} \leq p_{region\ 2}\ or\ p_{region\ 1} - p_{region\ 2} \leq 0$

$H_a : p_{region\ 1} > p_{region\ 2}\ or\ p_{region\ 1} - p_{region\ 2} > 0$

Table 5: Sample Data

	Midwest	**Northeast**	**South**	**West**
# of Tournament appearances (x or y)	807	560	1363	606
Total Proportion Sample Size (n or m)	3987	4154	8204	3244

Table 6: Two-Way Comparison Calculations

Comparison	**Z Calculated**	**P-Value 1**	**Interpretation**	**P-Value 2**	**Interpretation**
Midwest and Northeast	8.15675	0.00002	Reject H_0	0.001	Reject H_0
Midwest and South	4.91141	0.0002	Reject H_0	0.001	Reject H_0
Midwest and West	1.66401	0.097	Fail to Reject H_0	0.0485	Reject H_0
Northeast and South	-4.5385	0.0002	Reject H_0	0.9999	Fail to Reject H_0
Northeast and West	-6.09031	0.00002	Reject H_0	0.9999	Fail to Reject H_0
South and West	-2.64066	0.0082	Reject H_0	0.9959	Fail to Reject H_0

At 95% confidence level

4. Discussion

4.1 Multiple Linear Regression

For our Multiple Linear Regression analysis, we have included six regressions located in Table 1. Although our first regression has the highest adjusted R^2 with 79.88% it is not the most accurate regression. This is because for the majority of the years our data comes from winning a conference tournament guarantees a school to make the NCAA a tournament. For this reason, it is slightly a false narrative because it does help explain what helps make the NCAA tournament because we already know that winning a conference tournament guarantees a spot. For this reason, the regressions that explain the most about making the NCAA tournament are regressions two and five. Regression two includes win-loss percentage, regular-season champions, and SRS statistic with an R^2 of 78.78%; while regression 5 includes win-loss percentage, regular-season champions, and strength of schedule statistic with an adjusted R^2 of 77.92%. With 99%

confidence using the historical database, we can say that win-loss percentage, the regular-season champion, and SRS explain 77.92% of why schools make the tournament.

4.2 Binary Logistic Regression

For our binomial logistic Regression analysis, we looked at the season by season results located in Table 2. Each school either made the tournament that year or did not make it a binary 1or 0 for y variable. For our first regression, we did a kitchen sink regression where we put in as many significant independent variables as existed. For that reason, we take it with a grain of salt. Also in regression 1 points per game tracked in a negative way while opponent points per track positive. This would suggest that the more points a school score per game negatively affect their possibility of making the tournament and vice versa with opponents per game. This logically does not make sense, so for regression 2 we removed both points per game and wins in a conference because win percentage in a conference is a product of the number of wins in a conference. In regression 2 we got adjusted R^2 of 51.37% with all our independent variables at a 99% confidence level; For this reason, we believe that regression 2 is our best explanation of what variables explain what helps determine to make the tournament. Looking into the points per game variables with fewer variables, points per game became positive and opponent's points per game was a negative coefficient. This would suggest that yes scoring more points in games helps chances of making the NCAA tournament and having opponents score fewer points helps as well. However, other variables such as win-loss percentage, conference win-loss percentage, and SRS are more important which may be the cause of the negative-positive switch.

4.2 Multiple Linear Regression vs Binary Logistic Regression

To compare the two datasets we did regressions using the same independent variables. The variables that were the same in both datasets were win-loss percentage, SRS, and SOS. Table 1 regression 6 was significant, however, both SRS and SOS did not meet the requirements to be significant on 90% confidence. The adjusted R^2 of 63.97% even with SRS and SOS not being significant. While in Table 2 regression 8 all three independent variables were significant at a 99% confidence level. However, the adjusted was R^2 of 48.18%. Although theoretically, these are the same variables and one dataset is historical summations while the other is a season by season there is a difference in the R^2 by 15.79%. Although the multiple linear regression is a higher R^2 two of its coefficients are not significant, therefore the binomial logistic regression 8 is the more accurate regression due to the significance level of the coefficients.

4.3 Chi-Squared Testing

Tables 3 and 4 of our results display our results of the computation and order of operations in calculating our Chi-squared test. For our chi-squared test, we wanted to look at each the proportion of teams in each region, and determine how closely they compare to the proportion of total tournament appearances by region. Ideally, the proportion of teams in division 1 college basketball that can make the tournament from any given region, should equal the proportion of that region's appearances with respect to all appearances by all regions. In more specific terms, the proportion of schools from the Northeast region is 20.3% of all eligible schools in all four regions. So, what we are testing is that proportion is equal to the proportion of total NCAA tournament appearances from the Northeast compared to the total number of tournament appearances by all schools since 1939.

After using the chi-squared formula shown in table 4 of the results, we calculated a Chi-Squared value of $\chi^2 = 72.5343$. With a chi-squared value this large, and a $k = 3$, it is statistically irrefutable. The results show that $\chi^2 = 72.5343 > \chi^2_{.05,3} = 7.815$ in which case we reject our null hypothesis that the observed proportions of total tournament appearances in a region is equal to the proportions of schools in that region, and conclude that there is a difference in at least one of the regional basis proportions.

4.4 Two-Way Population Proportion Comparison

The Two-Way Population Proportion Comparison test allows us to conduct testing on regional influence in making the tournament. The tests require a matching comparison of each possible two-way match between regions. The following are all possible cases for matching and comparing population proportions; Midwest and Northeast, Midwest and South, Midwest and West, Northeast and South, Northeast and West, South, and West. In the test, we compare the sample proportion of those teams that make the tournament from their region in a given year and those teams that do not, and we compare all seasons going back to 1939 which was the first year of the tournament.

Using the formulas for two-way population proportion comparison discussed in the statistical methodology section and both tables in the Two-Way Comparison section of the results (Table 5&6) we can input the values from those tables to the formulas and get our results.

The results of these calculations show a couple of different things. The initial test conducted tests the hypothesis that each population proportion is equal meaning that each region sends approximately the same proportion of schools to the tournament over a significant amount of time as each region. The results show that the following comparison shows statistically significant p-values generated by the calculations to indicate that the proportion of appearances from each region differs over a significant amount of time: Midwest and Northeast, Midwest, and South, Northeast, and South, Northeast and West, South and West. All these comparisons came up with statistically significant p-values, in which case we would reject our null hypothesis and conclude that there is enough evidence to suggest that the proportions of the comparing schools are not equal to each other. This would indicate that there is an imbalance in the proportion of schools that make the tournament from each region.

Furthermore, our interest in determining which regions had a greater proportion than the other led us to conduct a second set of tests with an altered hypothesis. The null hypothesis was that the proportions when compared would indicate that observed tournament appearances from one regional population sample would be greater than that of the compared region's population sample. This test showed that the comparisons between Midwest and Northeast, Midwest and South, Midwest, and West all resulted in statistically significant p-values. Both Midwest and Northeast, Midwest, and South comparisons resulted in a p-value < 0.001 which is less than a significance level of $\alpha = 0.05$. And the comparison between regions Midwest and West resulted in a p-value of 0.0485, which is just below the significance level set. However, since all three comparisons resulted in a statistically significant p-value, we reject the null hypothesis and accept the alternative that there is enough evidence to indicate that region A's proportion of teams that make the tournament (first region listed) is greater than region B's proportion of teams that make the tournament (second region listed). The following region proportion comparison comparisons resulted in statistically insignificant p-values; Northeast and South, Northeast and West, South and West. The resulting p-values for the previously stated comparisons are all greater than 0.99, which is significantly larger than our significance level of $\alpha = 0.0$. Since these three comparisons resulted in such statistically insignificant p-values, we fail to reject our null hypothesis and conclude that there is not enough evidence to indicate that region A's proportion of teams that make the tournament (first region listed) is greater than region B's proportion of teams that make the tournament (second region listed).

With this second set of comparisons, we are able to understand that the proportion of tournament appearances made by northeast schools since 1939 is greater than the proportion of other schools. However, based on our earlier testing, it was indicated, if only by such a small margin, that the Northeast and West region's proportions were proved to be equal.

5 Conclusion

5.1 Limitations

No analysis is without its limitations, this exploratory analysis of what contributes to schools making the NCAA tournament is not an exception. The first is that the number of schools invited to the tournament has changed over the years. At the beginning of the tournament's existence, there were only 8 schools expanding in 1951, 1985, 2001, and 2011 to the modern-day 68 teams[3]. This limits the analysis because until 1985 winning a conference tournament did not guarantee a spot in the tournament. This causes anomalies where most of the time winning it means a spot in the tournament and sometimes it does not.

The second limitation is that our two data sources are slightly different. Although the two data sets come from the same source with one being a summation of historic statistics and another being data for each school and season they appear to have some information different.

Another important limitation comes with time. While many data sets on the internet come from trusted reliable sources, if the distributor of that data is also distributing a very high volume of data frequently and updating data tables, there is a possibility for error on their end. This is important to understand for validation purposes. Without a high powered database that can communicate between databases, it is almost impossible to confirm the accuracy of all information for our data set. Data from a third-party data source has the possibility of being low-quality, meaning some inaccurate data, and poorly aggregated, meaning it could be missing some values, which in our data set in some rare cases we found to be true.

5.2 Explanation of Analysis

In our statistical analysis, we were able to make a few different conclusions based on the calculations from different testing procedures. The regression analysis showed us that SRS was the variable that explained making the NCAA tournament the most. This makes sense because SRS is a statistic used to calculate how good a school is at basketball. When we took SRS out of regressions we could see that win-loss percentage, conference win-loss percentage, strength of schedule, regular-season champion, and conference tournament champion. Although for the most part conference tournament champions do not explain making the tournament, rather it tells us that the team makes the tournament by winning the conference tournament. Even though there were other variables that were significant such as points per game and opponent points per game on a unit by unit basis they did not lend much to the overall y variable. Unfortunately with both regressions (Table 1&2), we were not able to find a hidden variable that would help explain making the tournament. Instead, we only proved what is known to be common sense in what helps a school make the tournament. For example, a good SRS rating or not allowing your opponent to score more points. Although Table 1 explains making the tournament more with higher adjusted R^2, we believe that Table 2 is a more accurate regression. This is because Table 2 looks at the season by season basis and has many more data points to compare. Using Regression 2 in Table 2 we can explain 51.37% of making the tournament using win-loss percentage, conference win-loss percentage, SRS, and strength of schedule. This would indicate a moderately positive regression. However, these 4 variables boil down to winning games against good teams which anyone can tell you is needed to play in the NCAA tournament.

While conducting the Chi-Square analysis, we were able to prove our data calculations to be statistically significant. This indicates that our null hypothesis is that the proportion of regional representation of schools in the NCAA tournament is equivalent to the proportion of total NCAA appearances by the respective regions. This null hypothesis was rejected after achieving a statistically significant χ^2 value. This means that the expected proportion of schools said to make the tournament from each region is not equal to the actual proportion of schools that made the tournament.

The comparison tests of region proportions showed that each region has a different proportion of schools that appear in the tournament, with the exception of Midwest and West comparison. With this information, we can infer that the chance of any given school from the Northeast region has a different probability, however significant, of making the tournament compared to any given school from the Southern region. Conducting this test on a school to school basis is subject to extensive testing and can provide further insight.

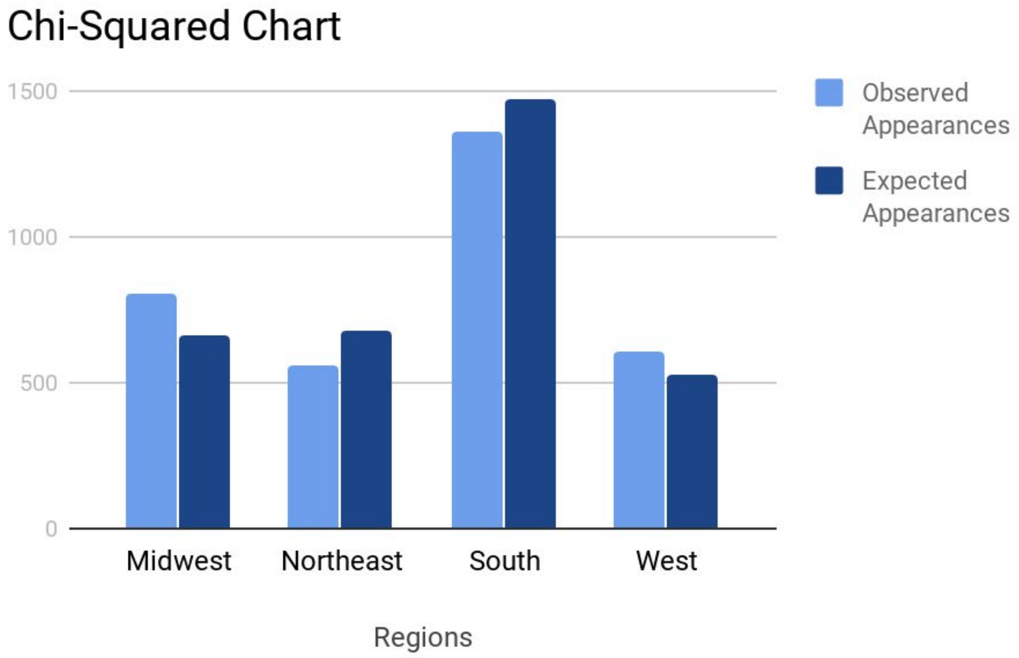
6. Bibliography

1) School Index: College Basketball at Sports. (n.d.). Retrieved May 9, 2020, from https://www.sports-reference.com/cbb/schools/

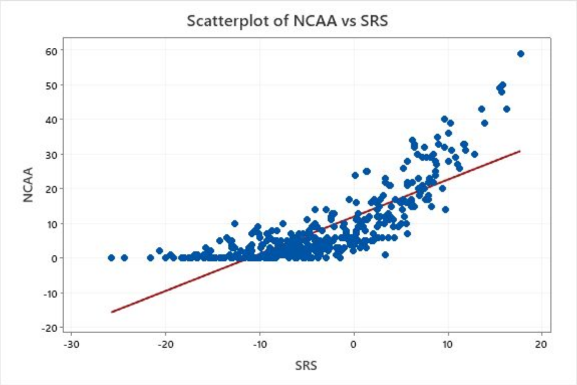
2) U.S. Dept. of Commerce, Economics and Statistics Administration, Bureau of the Census. (1994). Geographic areas reference manual. Washington, D.C. Chapter 6 Retrieved from https://www.census.gov/programs-surveys/geography/guidance/geographic-areas-reference-manual.html

3) Wilco, D. (2020, April 20). What is March Madness: The NCAA tournament explained. Retrieved May 9, 2020, from https://www.ncaa.com/news/basketball-men/bracketiq/2020-04-20/what-march-madness-ncaa-tournament-explaine

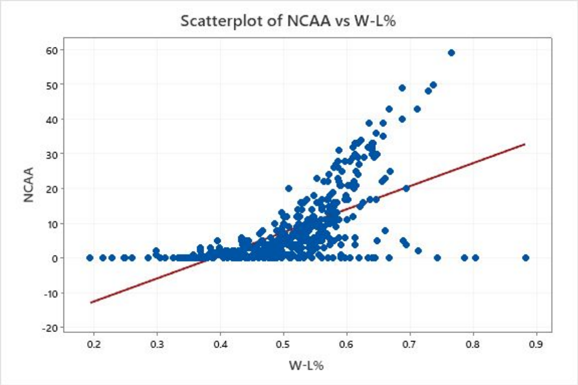
7. Graphs



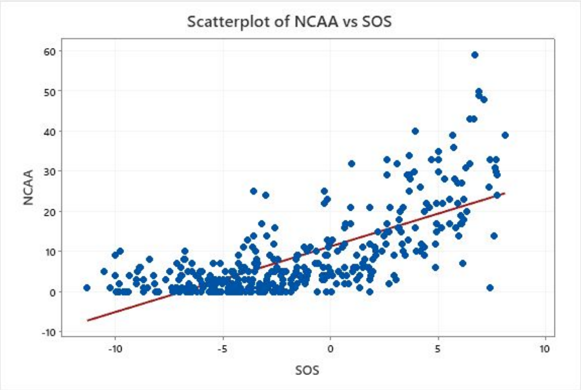
Chi-Squared Chart



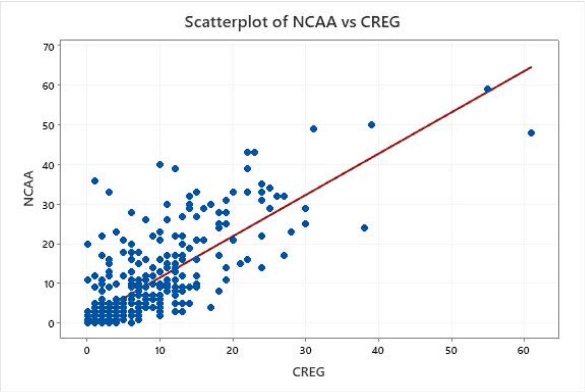
Scatterplot of NCAA vs SRS



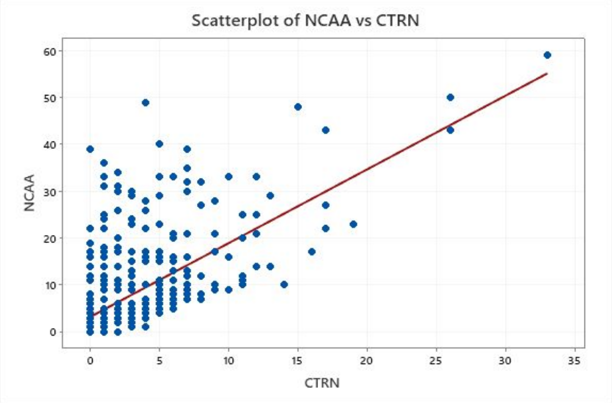
Scatterplot of NCAA vs W-L%

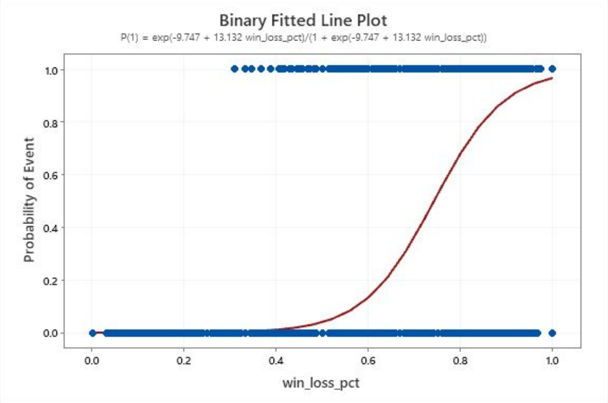


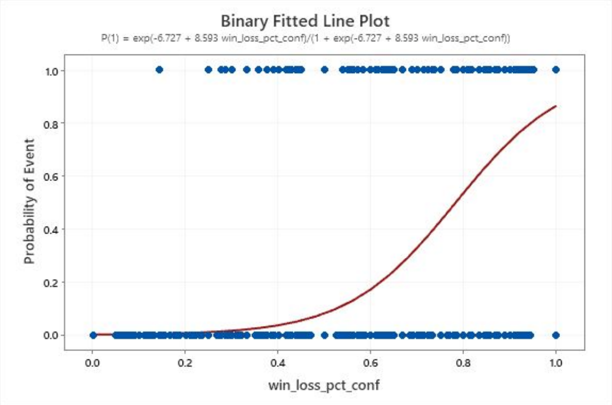
Scatterplot of NCAA vs SOS

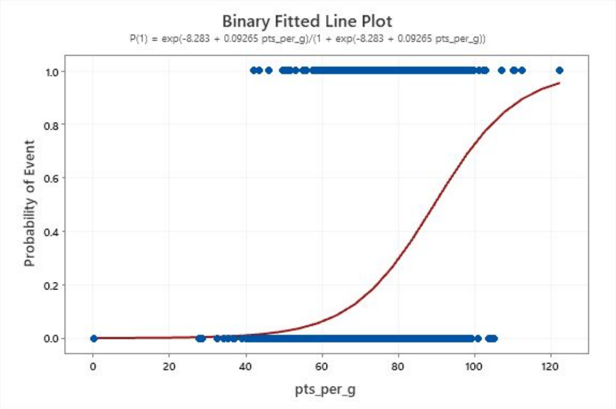


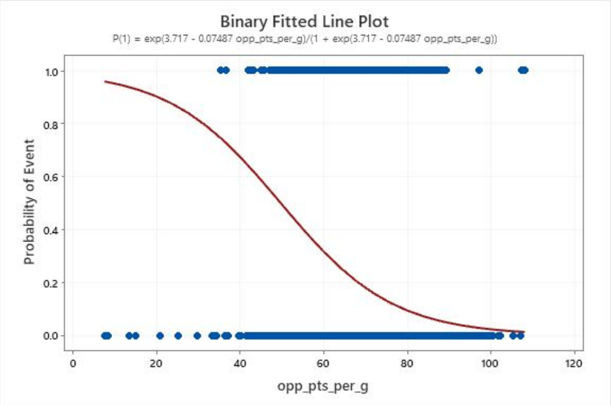
Scatterplot of NCAA vs CREG

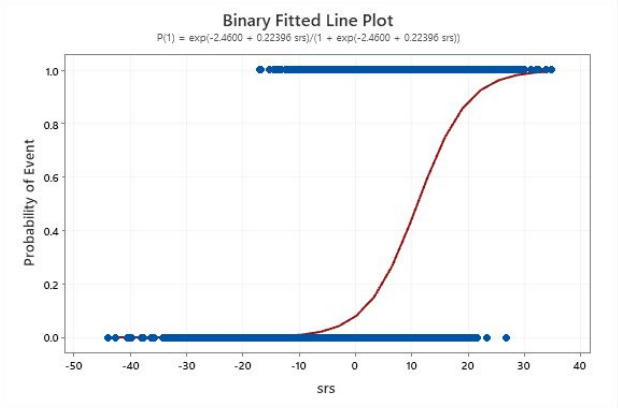

Scatterplot of NCAA vs CTRN
NCAA
0
10
20
30
40
50
60
CTRN
0
5
10
15
20
25
30
35

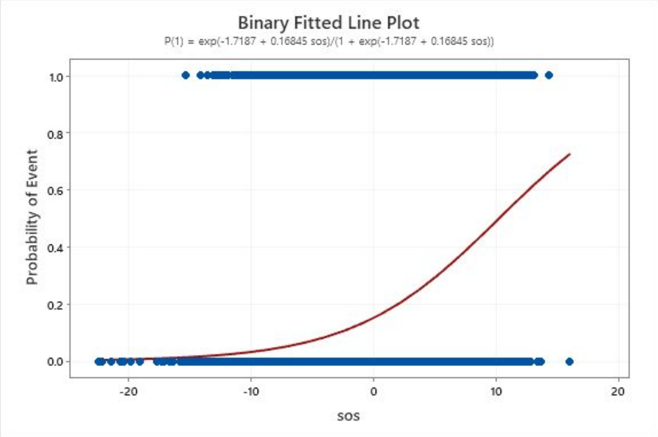

Binary Fitted Line Plot
P(1) = exp(-9.747 + 13.132 win_loss_pct)/(1 + exp(-9.747 + 13.132 win_loss_pct))
Probability of Event
0.0
0.2
0.4
0.6
0.8
1.0
win_loss_pct
0.0
0.2
0.4
0.6
0.8
1.0


Binary Fitted Line Plot
P(1) = exp(-6.727 + 8.593 win_loss_pct_conf)/(1 + exp(-6.727 + 8.593 win_loss_pct_conf))
Probability of Event
0.0
0.2
0.4
0.6
0.8
1.0
win_loss_pct_conf
0.0
0.2
0.4
0.6
0.8
1.0


Binary Fitted Line Plot
P(1) = exp(-8.283 + 0.09265 pts_per_g)/(1 + exp(-8.283 + 0.09265 pts_per_g))
Probability of Event
0.0
0.2
0.4
0.6
0.8
1.0
pts_per_g
0
20
40
60
80
100
120


Binary Fitted Line Plot
P(1) = exp(3.717 - 0.07487 opp_pts_per_g)/(1 + exp(3.717 - 0.07487 opp_pts_per_g))
Probability of Event
0.0
0.2
0.4
0.6
0.8
1.0
opp_pts_per_g
0
20
40
60
80
100
120


Binary Fitted Line Plot
P(1) = exp(-2.4600 + 0.22396 srs)/(1 + exp(-2.4600 + 0.22396 srs))
Probability of Event
0.0
0.2
0.4
0.6
0.8
1.0
srs
-50
-40
-30
-20
-10
0
10
20
30
40


Binary Fitted Line Plot
P(1) = exp(-1.7187 + 0.16845 sos)/(1 + exp(-1.7187 + 0.16845 sos))
Probability of Event
0.0
0.2
0.4
0.6
0.8
1.0
sos
-20
-10
0
10
20