Results from Staying Home on a Fall Saturday

--Home Field Advantage in College-Football--

By

Patrick Yonker

Advisor: Dr. Morris Marx

A Graduate Proseminar
In Partial Fulfillment of the Degree of
Masters of Science in Mathematical Sciences
The University of West Florida
December 2012

Approval Page

This Proseminar of Patrick Yonker approved:

Morris Marx, Ph.D., Proseminar Advisor Date

Josaphat Uvah, Ph.D., Proseminar Committee Chair Date

Accepted for the Department:

Jaromy Kuhl, Ph.D., Chair Date

Table of Contents

Introduction

A fall Saturday means it's time for cool air, college football, and thousands of fans packing stadiums across the country. Two football programs will be fighting for victory. With each victory comes bragging rights. One team calls this stadium home and other will just be visiting for the day. Although bragging rights can be important, there are other important results that come from winning. More wins can lead to a larger fan base, more tickets sold, and a post-season appearance. So more wins ultimately leads to more money coming into a University.

Money is an attractive outcome of winning football games for those running a university. Any fan of any sport will say that playing at home has a built in advantage and increases your chances of winning. This is a reason why schools find room in the budget for new buildings and renovations in excess of 100 million dollars.[espn]

This paper is to determine what factors result from playing at home. As an avid fan of college football that follows teams all over the country, I, like many other fans, have assumptions about home field advantage. The average fan can do a basic search on the internet and find multiple lists that rank stadiums for home field advantage. Many of these rankings are based on the difference between winning percentage at home and winning percentage on the road. This can be very deceiving due to the fact that good teams can have a high road winning percentage also; thus making the difference between home and away small, which would put them at the bottom of the list. Just because a team wins on the road just as much as they win at home does not mean that they do not have an advantage playing at home.

This leads to a detailed analysis on what aspects of a football game differ between the home and away team. It is commonly believed that a home crowd can have an influence on the game. In fact, some research has shown that crowd capacity and denseness, familiarity, travel, and rule factors make a difference in favor for the home team.[Harvard] Instead of analyzing what causes there to be a difference, the focus of this paper is to analyze the result of the game and a few different key components.

In the game of football, turnovers, yards gained/allowed, time of possession, number and yardage of penalties, can all have an effect on the outcome of the game. All of these will be defined in the Background Section of the paper but the idea is to determine if there is a difference between these components for the home and away team. Some questions to be answered are (proposed from the viewpoint of a fan):

1.) Is there a difference in points scored? If so, how confidently can this be said?
2.) Does the home team control the time of possession?
3.) Does the home offense produce more total yards?
4.) Is there a home field advantage difference between conferences?
5.) Does home field have an effect on the number and yardage of penalties?
6.) Will the away team be more susceptible to committing turnovers?
7.) If (1) and (3) are true, is there a correlation between total yards and margin of victory?

By using different statistical analysis methods, we can answer these questions and possibly propose some more for future work. If it can be shown that playing at home causes there to be a difference in these components of the game, then it further validates that there exists home field advantage.

Background/Literature Review

<u>The game of football:</u> As listed in the introduction, the components to be analyzed are number and yardage of penalties, turnovers, time of possession, total offensive yards and margin of victory. These components are defined below:

- Penalty: While on offense or defense, if a player commits a "foul" or "illegal" play, the referee will throw a penalty flag on the field. Penalties result in negative yardage against the guilty team. Depending on the type of penalty, yardage can be 5, 10, or 15 yards.
- Turnover: When a team loses possession of the ball and the other team gains possession. Interceptions and fumbles are the two types of turnovers used in the analysis. Fumbles are when a player drops the ball and the other team recovers the ball. Interceptions are when a defensive player catches a pass attempted by the offense. Turnovers are an instant change of possession and can yield to good field position by the recovering team. Accordingly, causing huge momentum swings, thus having an effect on the outcome of the game.
- Time of Possession: The amount of time of a 60 minute game in which a team control possession of the ball. Time of possession does not count overtime play.
- Total Offensive Yards: The amount of yards that the offense is able to move the ball down the field; including running and passing yards. Does not include special team or defensive yards but does include yards gained in overtime play.
- Margin of Victory: The point differential between the home and away team. For purposes of this analysis, margin of victory can be positive or negative and is relative to the home team. For example, if the home team loses by 10 points, then the margin of victory would be -10.

Background of statistics:

- Normal Distribution: For the following formula, μ is the mean and σ^2 is the variance such that:[zar] $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(\mu-x)^2}{2\sigma^2}}$

  In some of the following analyzes, methods will require the data to be normal and fit the above distribution. In order to verify that the data is normal, the Shapiro-Wilk Test provides a robust test statistic.[engineering] The Shapiro-Wilk Test examines the data and rejects the null hypothesis that the data is normal if the p-value if found to be less than our chosen alpha. This means that if we reject the null hypothesis, we do so with 95% confidence. If the test fails to reject the null hypothesis, we can only assume the data is normal and continue.

- One-Sample T-Test: For this test, the data must be normal. The following test-statistic is used to compare to the value found in the *Student t Table ($t_{\alpha,n}$)*.

$$T_n = \frac{\bar{Y} - \mu}{\frac{S}{\sqrt{n}}}$$

  After calculating T_n, it is compared with the table value using $P(T_n \geq t_{\alpha,n}) = \alpha$. This test is one-sided and is used "to test H_0:$\mu=\mu_o$ versus H_a:$\mu>\mu_o$ at the α level of significance, reject H_0 if t is either $\geq t_{\alpha,n-1}$".[Larson & Marx] The confidence interval for μ using a one-sample t-test is found by using $P\left(\bar{Y} - t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{Y} + t_{\frac{\alpha}{2},n-1} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha$. Here, alpha can be changed to different values, which will change the length of the confidence interval.

- One-Way ANOVA (signal-factor analysis of variance): This analysis of variance test is used to test equal means between groups. The term "One-Way" refers to the fact that the test is testing one factor between multiple groups. Just as in a One-Sample T-Test, the data must be normal. If there are only two groups to be tested, the One-Way ANOVA and the Two-

Sample T-Test will provide the same result. Because of the versatility of the One-Way ANOVA, the Two-Sample T-Test will not be used in this analysis. The One-Way ANOVA tests H_0:μ_1=μ_2=…=μ_k, where k is the number of groups versus H_a: at least one difference among the k number of means. An F-Test statistic is used to determine if the null hypothesis is rejected or not. The following formulas[zar] are used to find the F-value so that it can be compared to the critical table value. In the formulas, N is the total number of data points from all groups, k is the number of groups, and n_j is the number of data points within the j^{th} group.

The degrees of freedom are defined by: $Groups\ DF = k - 1$ and $Error\ DF = N - k$

The value $C = \frac{(\sum\sum X_{i,j})^2}{N}$ is used to find $Groups\ SS = \sum_{i=1}^{k} \frac{(\sum_{j=1}^{n_j} X_{i,j})^2}{n_j} - C$

Once the once value $Error\ SS = \sum_{i=1}^{k}\sum_{j=1}^{n_j} X_{i,j}^2 - \sum_{i=1}^{k} \frac{(\sum_{j=1}^{n_j} X_{i,j})^2}{n_j}$ is found,

$Groups\ MS = \frac{Groups\ SS}{Groups\ DF}$ and $Error\ MS = \frac{Error\ SS}{Error\ DF}$ are calculated.

The F-Test Statistic is $F = \frac{Groups\ MS}{Error\ MS}$ and is compared to $F_{\alpha(1),(k-1),(N-k)}$.

If the p-value is found to be less than α, then H_0:μ_1=μ_2=…=μ_k is rejected.

- Kruskal-Wallis Test: If the data set if found to be non-normal, then the nonparametric Kruskal-Wallis analysis will be used. The Kruskal-Wallis Test uses data ranked from smallest to largest and compares the sample means. This tests the same null hypothesis as the One-Way ANOVA, which is H_0: μ_1=μ_2=…=μ_k, where k is the number of groups. With rejection of H_0, the alternative hypothesis, H_a: not all means are equal, is accepted. If k>2, we cannot draw conclusions as to which means are unequal and which, if any, are equal. As the data is ranked, there will be the set of $n = \sum_{j=1}^{k} n_j$ observations with the rank sum, $R_{.j}$, corresponding to Y_{ij}. Calculating these values allows us to find the Kruskal-Wallis Test statistic, B, to be found. This test statistic can be approximated by a chi-square distribution so that H_0 is rejected if $B > X^2_{1-\alpha,k-1}$. B is calculated using the following formula[Larson & Marx]: $B = \frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{R_{.j}^2}{n_j} - 3(n+1)$.

- Tukey's Test: When using a One-Way ANOVA and the null hypothesis is rejected, it shows that not all means are equal. Tukey's Test uses multiple comparisons to determine which means are equal and which are not. Tukey's Test measures equality of all the pairs of individual means, or H_0:μ_i=μ_j versus H_a: $\mu_i \neq \mu_j$ for all $i \neq j$ [Larson & Marx]. In order to do a Tukey's Test, the data is assumed to be normal. This method for testing equality of means uses confidence intervals for μ_i-μ_j. This intervals are found from the studentized range, $Q_{k,v} = \frac{R}{S}$, where $R = \max_i W_i - \min_i W_i$ and v= n – k. If $\bar{Y}_{.j}$ for j=1,2,…k is the k sample means, $n_j = r$ is the common sample size, and $D = \frac{Q_{\alpha,k,rk-k}}{\sqrt{r}}$, then the confidence interval is defined below:

$$\bar{Y}_{.i} - \bar{Y}_{.j} - D\sqrt{MSE} < \mu_i - \mu_j < \bar{Y}_{.i} - \bar{Y}_{.j} + D\sqrt{MSE}$$

  There is a nice result from this interval. If zero is included in the interval, then the null hypothesis that the two means are equal cannot be rejected. So, if zero is not in the interval, H_0 is rejected and thus we assume the means are not equal [Larson & Marx].

- Poisson Distribution: For a random data set, it can be described by a Poisson Distribution if it follows: $P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$. The sample mean of the k_i's is given by [Larson & Marx]: $\bar{k} = \frac{\sum_{i=1}^{n} k_i}{n}$ At this point, the $P_x(k)$ can be found for each k. Once all of these are found, the test statistic can be calculated by: $d = \sum_{i=1}^{t}\frac{(k_i - np_{io})^2}{np_{io}}$. This test statistic allows us to compare it with a chi-square distribution and determine if the null hypothesis is rejected. The null hypothesis is rejected if d ≥ $X^2_{1-\alpha,t-1}$. By failing to reject the null hypothesis, it means that the data is in fact random and fits a Poisson distribution.

- Linear Modeling: Two sets of data points X and Y can be tested to determine if there is a linear relationship between the two. If a set of points (X_1,Y_1), (X_2,Y_2),….(X_n,Y_n) fits a linear model, then the maximum likelihood estimators for β_0, β_1, and σ^2 are given below:

$$\widehat{\beta_1} = \frac{n\sum_{i=1}^{n} x_i Y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{x}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}$$

These estimators lead us to finding a Student t test statistic $t = \frac{\widehat{\beta_1} - \beta_{1o}}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$. If the p-value is found to be smaller than .05, then the null hypothesis, H_0: $\beta_1 0$ is rejected. This would indicate an effect between the two variables. To test this effect, or correlation, a signal value is found. This value is referred to R-Squared and found with:

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

R-Squared takes a value between 0 and 1 and represents the "proportion of the total variation in the y_is that can be attributed to the linear relationship with x." [Larson & Marx]

Assumptions

A.1) Data used to analyze margin of victory, time of possession, and total yards, comes from games between ranked teams. In order to focus on home field advantage, it is reasonable to take only games where both teams are ranked in the Associated Press Top 25. This assumption attempts to only use games between "even" teams. It has become popular for bigger schools to schedule smaller schools for a home game and payout big money for an assumed beat down. These games provide no interest in trying to analyze home field advantage since these are predetermined mismatches.

A.2) When analyzing if the same home field advantage exists between conferences, it is a fair assumption to only use games from within conference play. By taking a sample of data from 2008-2011, this makes an assumption that each team will play a conference opponent at home and on the road versus the same team. In this assumption, the ranked matchup assumption has been removed so that every Bowl Championship Series Conference (BCS) game for the last 4 years is used. The BCS is comprised of 6 conferences whose winners have the best chance at competing for a national championship.

A.3) For the question relating to number of penalties and penalty yardage, it is reasonable to take a sample of every type of college matchup. The assumption is that the matchup has no effect on penalties. So, when answering this question, every matchup from the first 7 weeks of the 2012 season will be used. Matchups include Division I versus lower divisions, ranked versus ranked, ranked versus unranked, unranked versus unranked, conference games, and non-conference games. This same assumption applies to turnovers. It is assumed that turnovers are random events and therefore are not affected by the matchup.

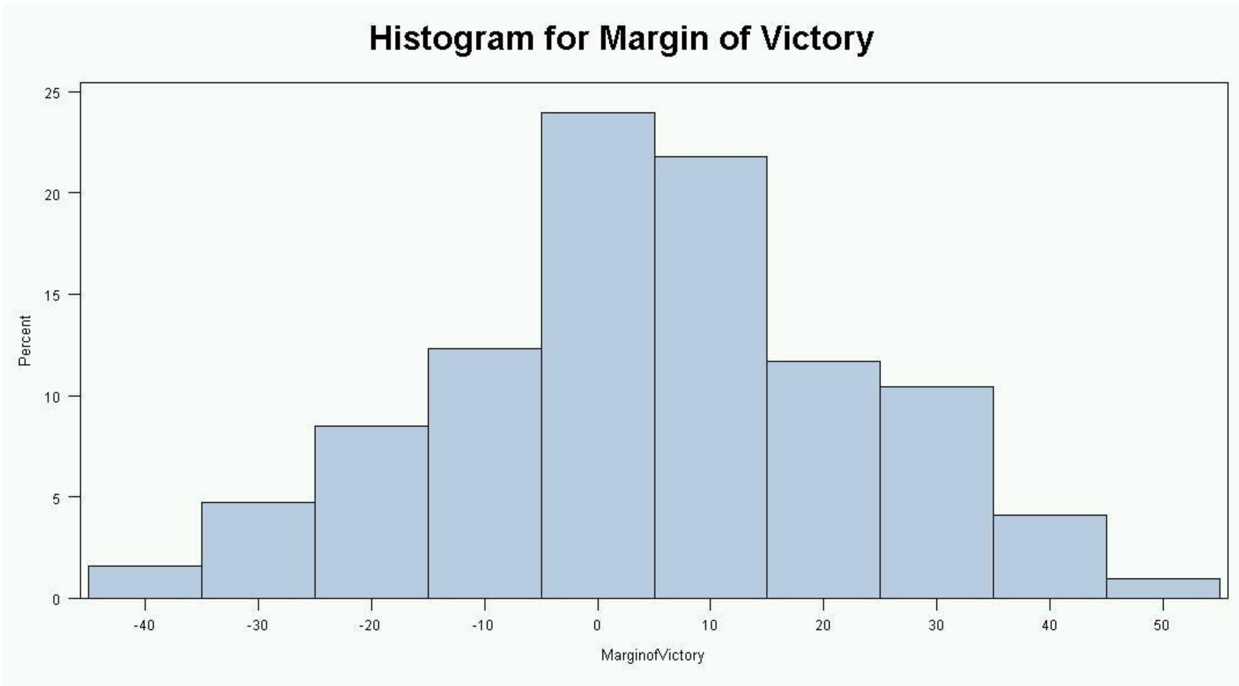
Analysis

- Is there a difference in points scored? If so, how confidently can this be said?

  The Margin of Victory data is taken from all 317 ranked matchups from 2003-2011; which were not played at neutral sites. The first step is to test our assumption that the data is normal.

By Shapiro-Wilks Test: H_o: data is normal versus H_a: data is not normal

The test statistic W=.994 yields a p-value=.2798. Since the p-value is greater than α=.05, it fails to reject H_o, thus the data is assumed to be normal.



N=317, Mean=4.57, Standard Deviation=18.29

Next step is to test if the margin of victory is equal to zero. If the margin of victory is equal to zero, then we can say there is no home field advantage between ranked teams. As shown in the above histogram, the data clearly appears to be favoring the positive side. To test this statistically, the One-Sample T-Test will be used:

H_o: $\mu=0$ or H_a: $\mu>0$

The test statistic of t=4.45 yields a p-value <.0001. Therefore, we can reject the null hypothesis that the mean is equal to zero. Thus, we can say that there may be home field advantage. This brings up the question on how confident can that statement be made. This can be determined by finding a confidence interval for the mean.

Using α=.05, the 95% confidence interval is ($2.5524 \leq \mu \leq 6.5958$).

Using α=.001, the 99.9% confidence interval is (1.1611≤μ≤7.9872).

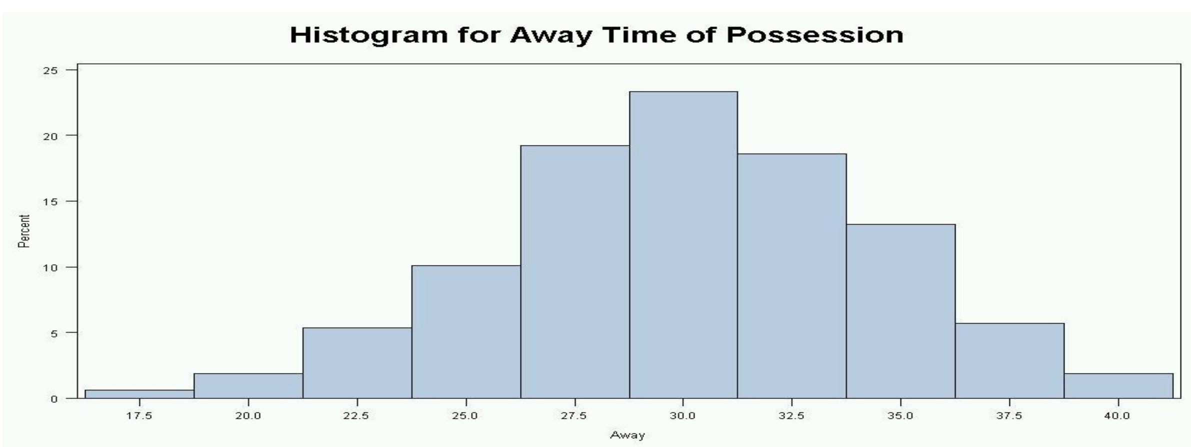
Based on these intervals, it can be said with 99.9% confidence that there exist home field advantage when analyzing margin of victory.

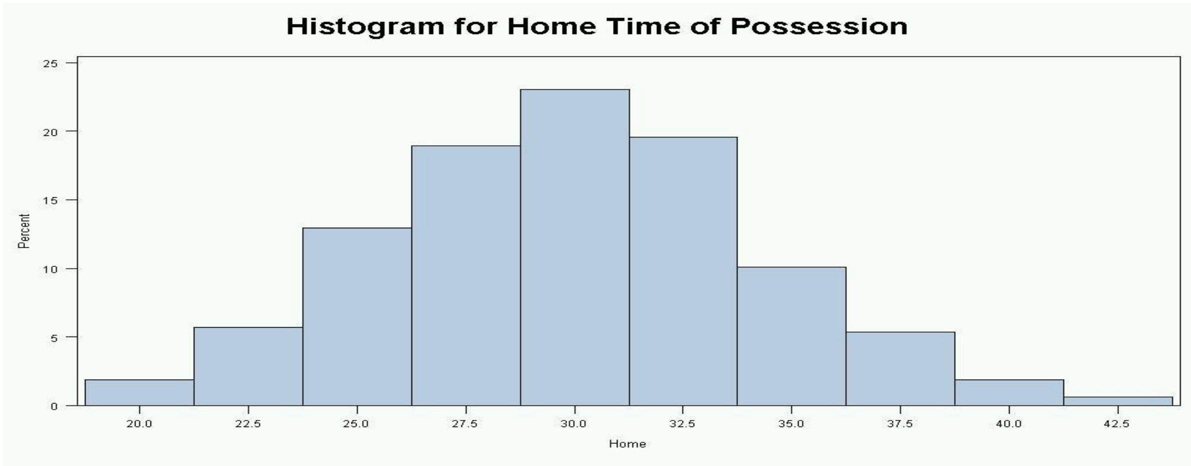
- Does the home team control the time of possession?

The same data set regarding ranked teams from 2003-2011 is used in this analysis. The raw data is given in minutes:seconds format which is converted into decimal. There is a possible total of 60 minutes to possess the ball. The goal is to test whether there is a difference in the mean time of possession.

Test for normality: H_0: data is normal versus H_a: data is not normal

By the Shapiro-Wilks Test, the test statistic is found to be 0.9966 for the away time of possession which yields a p-value=0.7438. Since 0.7438 is greater than α=.05, we cannot reject H_0, thus the data for away time of possession is assumed to be normal. For the home time of possession, the test statistic is .9965, yielding a p-value of .7377. Since .7377 is also greater than α=.05, the data for home time of possession is also assumed to be normal. Being as both sets of data are normal and assumed to be equal variances, a single-factor ANOVA can be done.





$N_{away}=N_{home}=317$

Away mean is 30.08 and Home mean is 29.91.

Away standard deviation is 4.3138 and Home is 4.3146.

H_0: $\mu_{home}=\mu_{away}$ versus H_a: $\mu_{home}\neq\mu_{away}$

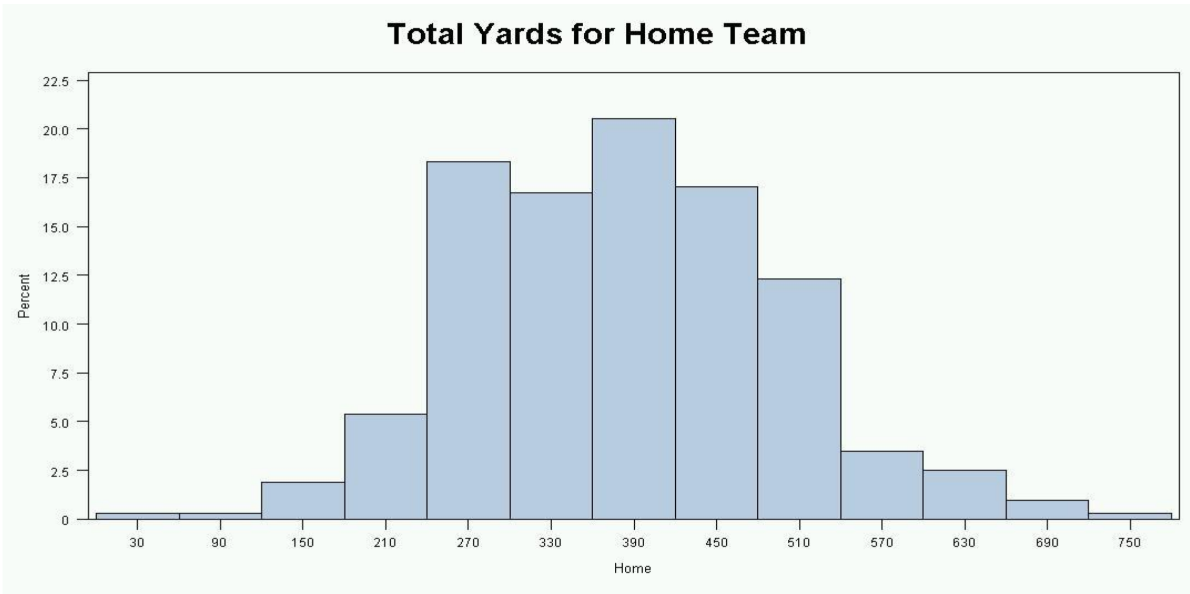
By examining the histograms, it is not clear if the mean of either the home or away team is higher or lower than 30 minutes. Since there is only a possible 60 minutes to possess the ball, equal means would be that both are equal to 30 minutes. This leads us to guess that the means are in fact equal based on the graphs. Using the One-Way ANOVA, it produces an F-test statistic of 0.26 yielding a p-value of 0.6079. Given 0.6079>.05, H_0 cannot be rejected. Therefore, it has been shown that there is no difference in the time of possession for the home and away team.

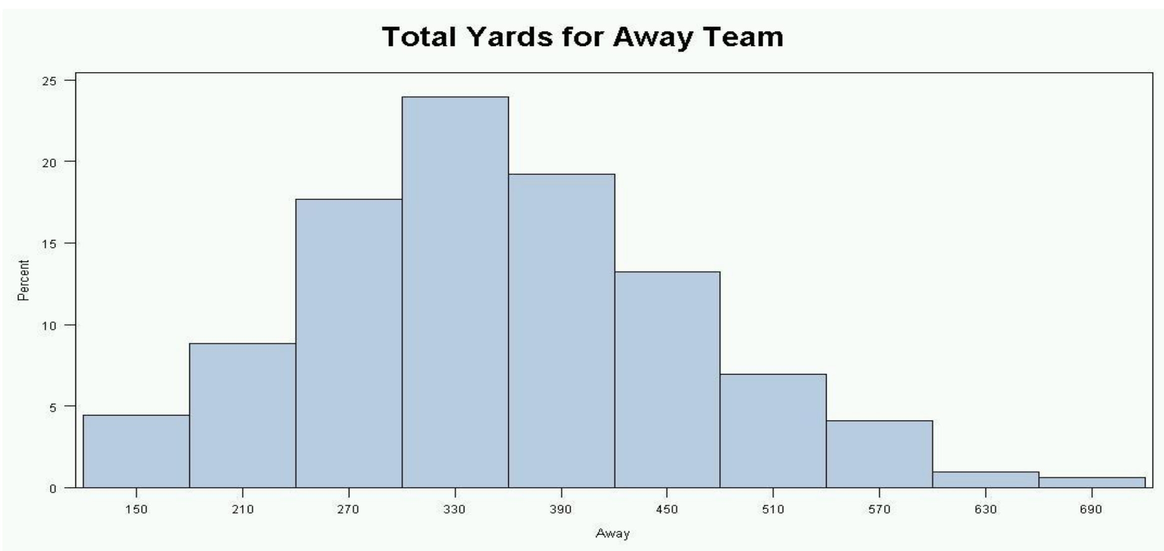
- Does the home team offense gain more total yards?

Using the same data set between ranked teams, the total yards for home and away team needs to be tested for normality.

H_0: data is normal versus H_a: data is not normal

By the Shapiro-Wilks Test, the test statistic for home team total yards is .9952, which yields a p-value of 0.4413. For the home team, we cannot reject that the data is normal. The away team test statistic is .9899 yielding a p-value of .0274. Since .0274<.05, the null hypothesis is rejected; meaning that the data for the away team total yards is non-normal. To use an ANOVA, the data must be normal so this results in using the non-parametric Kruskal-Wallis Test.





$N_{away}=N_{home}=317$

Away mean is 354.67 and Home is 380.92.

Away standard deviation is 106.37 and Home is 112.66.

H_0: $\mu_{home}=\mu_{away}$ versus H_a: $\mu_{home}\neq\mu_{away}$

This test yields a chi-square test statistic of 9.3909. The p-value for this is found to be 0.0022 which means that H_0 is rejected. Since equal means is rejected, it can be assumed that there is a difference between the total yards gained for the home and away team.

- Is there a home field advantage difference between conferences? The question is answered using a different data set. The data set presented in Assumption #2 will be used. This data set consists of the Margin of Victory for the home team. A Shapiro-Wilks normality test is done for each group, where the groups are each of the BCS conferences.

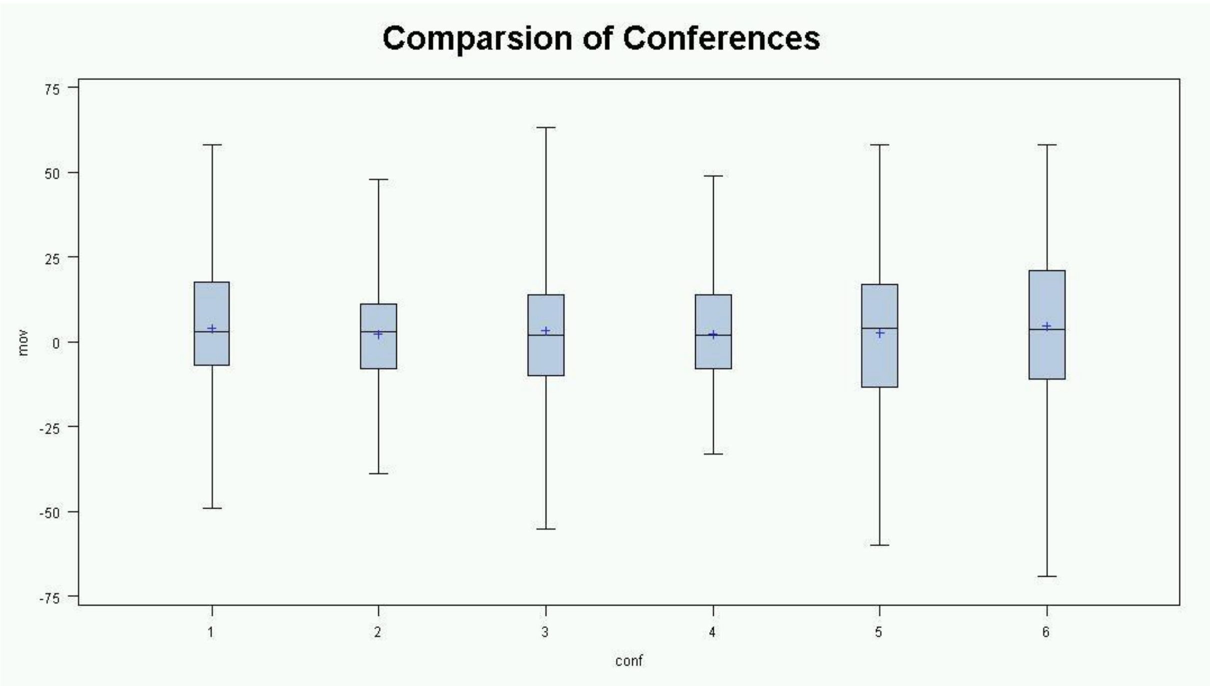
Conference	Test Statistic	P-Value	Reject H_0?
1.) Southeastern (SEC)	.9962	.9237	Do Not Reject
2.) Atlantic Coast (ACC)	.9898	.1926	Do Not Reject
3.) Big 10	.9880	.1315	Do Not Reject
4.) Big East	.9865	.3237	Do Not Reject
5.) Big 12	.9926	.4528	Do Not Reject
6.) Pac 12	.9924	.4185	Do Not Reject

H_o: $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5=\mu_6$ and H_a: not all equal

The One-Way ANOVA test produces F=.38 which has a p-value=.8607. Therefore, the null hypothesis cannot be rejected. This leaves the assumption that there is no difference between the mean Margin of Victory for all six conferences. The histograms for each conference are found in Appendix A. Even though they take different shapes, each histogram is found to be normal with equal means.

Tukey's Test provides a good method to verify this result from the ANOVA. The results and Tukey Intervals are listed in the table below. Since zero is included in the interval for each pairing, every comparison fails to reject the null hypothesis. This further validates the claim from the ANOVA that all conferences have the same home field advantage.

Comparison	μ_i-μ_j	Interval Minimum	Interval Maximum
6-1	0.537	-5.401	6.475
6-3	1.176	-4.828	7.181
6-5	1.851	-4.088	7.789
6-4	2.235	-4.646	9.116
6-2	2.311	-3.603	8.226
1-3	0.640	-5.396	6.675
1-5	1.314	-4.656	7.283
1-4	1.698	-5.210	8.607
1-2	1.775	-4.171	7.720
3-5	0.674	-5.361	6.710
3-4	1.059	-5.907	8.024
3-2	1.135	-4.877	7.147
5-4	0.384	-6.524	7.293
5-2	0.461	-5.485	6.407
4-2	0.076	-6.812	6.964



- Does home field have an effect on the number and yardage of penalties? The average fan will tell you that the home crowd has an effect on the referees calling fewer penalties for the home team. The data set for this consists of that described in Assumption #3. Completing the Shapiro-Wilk's normality test shows that we reject that all of the data are non-normal.

	Test Statistic	**P-Value**	**Reject Ho?**
# of Penalties Home	.9722	<.0001	Reject: non-normal
# of Penalties Away	.9685	<.0001	Reject: non-normal
Penalty Yards Home	.9728	<.0001	Reject: non-normal
Penalty Yards Away	.9642	<.0001	Reject: non-normal

The next step to test is if the number of penalties fits a Poisson distribution. The table for testing Poisson is shown below:

Away: $\bar{k} = 6.2840909$

# Detected	**Frequency**	**Proportion**	**Px(k)**	**d**
0-1	9	.020455	.011725	2.85972
2	27	.061364	.036839	7.18373
3	40	.090909	.077167	1.07678
4	56	.127273	.121231	.132483
5	61	.138636	.152365	.544280
6	53	.120455	.159579	4.22059
7	54	.122727	.143259	1.29474
8	49	.111364	.112531	.005328
9	26	.059091	.078573	2.12545
10	25	.056818	.049376	.493557
11	22	.050000	.028208	7.40755
12	7	.015909	.014772	.038513
13+	11	.025000	.012511	5.4855
Sum	440	1	1	32.8382
$X^2_{.95,11}$= 19.675 Since 32.8382>19.675, reject that this data is random and fits a Poisson Distribution				

Home: $\bar{k} = 6.122727$

# Detected	Frequency	Proportion	Px(k)	d
0-1	16	.036364	.013424	17.2482
2	25	.056818	.041095	2.64694
3	45	.102273	.083872	1.77626
4	48	.109091	.128381	1.27532
5	69	.156818	.157209	.000427
6	50	.113636	.160424	6.00407
7	50	.113636	.140319	2.23251
8	50	.113636	.107392	.159756
9	37	.084091	.073059	.73296
10	17	.038636	.044732	.365487
11	12	.027273	.024898	.099659
12	13	.029545	.012704	9.82362
13+	8	.018183	.010299	2.65473
Sum	440	1	1	45.0199
$X^2_{.95,11}$= 19.675 Since 45.0199>19.675, reject that this data is random and fits a Poisson Distribution				

Based on the rejection of both null hypotheses, it can be concluded that the number of penalties does not fit a Poisson distribution. Even though both are not Poisson, we can still test equal means using the Kruskal-Wallis Test. The test gives a test statistic of .5031 yielding a p-value of .4782. Since .4782>.05, we cannot reject equal means. This shows that the number of penalties for the home and away team can be assumed to be equal. The same test can be done for the yardage. Since different penalties have different yards associated with them, it is acceptable to test equal means between penalty yards. Doing so gives a test statistic of .0828 and p-value of .7735. Since .7735>.05, equal means between penalty yards cannot be rejected. By accepting both null hypotheses, it is clear that home field does not have any effect on penalties.

- Will the away team be more susceptible to committing turnovers? Turnovers can have a major impact on the outcome of a game so showing a difference between home and away would be meaningful. The data set is the same as that used for penalties and described in

Assumption #3. Testing for normality shows that both data sets are non-normal because the away test statistic of .8972 and home test statistic of .8946 produce a p-value<.0001. Since the p-value is less than .05, we can reject that the data is normal. Next, is to test if the number of turnovers fits a Poisson distribution.

Away: $\bar{k} = 1.827273$

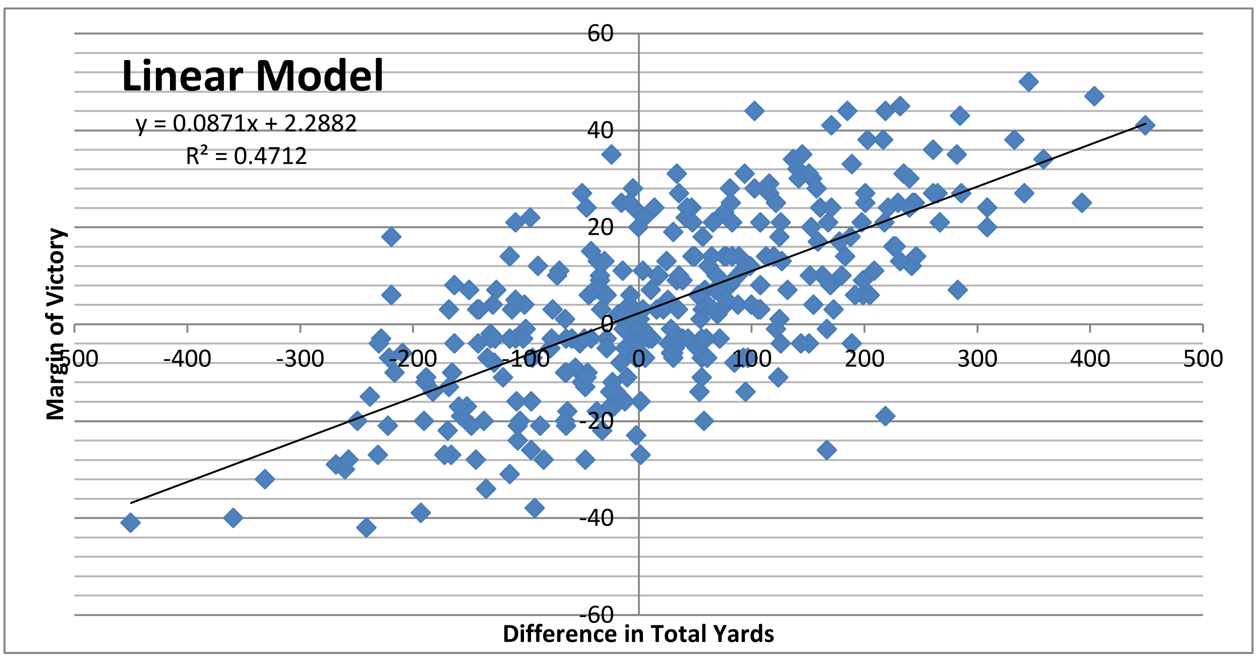
# Detected	Frequency	Proportion	Px(k)	d
0	75	.170455	.160852	.252231
1	125	.284091	.293920	.144627
2	126	.286364	.268536	.52076
3	60	.136364	.163563	1.99015
4	34	.077273	.074718	.038434
5	13	.029545	.027306	.080813
6+	7	.015909	.008316	3.05053
Sum	440	1	1	6.07074
$X^2_{.95,5}$= 11.070 Since 6.07074<11.070, cannot reject that this data fits a Poisson Distribution				

Home: $\bar{k} = 1.618182$

# Detected	Frequency	Proportion	Px(k)	d
0	98	.222727	.198259	1.32869
1	137	.311364	.320819	.122617
2	98	.222727	.259572	2.30115
3	64	.145455	.140011	.093122
4	32	.072727	.056641	2.01017
5	7	.015909	.018331	.140793
6+	4	.009091	.004944	1.53046
Sum	440	1	1	7.52702
$X^2_{.95,5}$= 11.070 Since 7.52702<11.070, cannot reject that this data fits a Poisson Distribution				

Since both test fail to reject that the data fits a Poisson distribution, we can assume the data is truly random and described by the Poisson. Now that both home and away have been shown to fit a Poisson distribution, we can test equal means using the Kruskal-Wallis Test. The test gives a test statistic of 4.8829 and resulting p-value of .0271. We reject equal means between the number home and away turnovers because .0271<.05. This shows that the away team does in fact have more turnovers and this is supported by the Poisson distribution above.

- If (1) and (3) are true, is there a correlation between total yards and margin of victory? From the previous analysis, it has been shown that there exists a difference between away and home teams as far as points scored and total yards gained. As a fan, it would make sense that as more yards are gained offensively, there are more points scored. To test if this relationship exists, a simple linear model will be created. The estimators are calculated which yields a Student t test statistic equal to 3.01. This test statistic produces a p-value=.0028. Since .0028<.05, we reject H_0: $\beta_1=0$ and accept that there is a linear relationship between the difference of total versus the Margin of Victory.



By plotting the points, difference in total yards versus Margin of Victory, it can be reasonably assumed from the graph that there exists the linear relationship. Since the test statistic leads to rejecting the null hypothesis, thus assuming a linear model, we can calculate how much correlation there is between the two variables. The R-Squared value is found to be .4712. This means that 47.12% of the variation of Margin of Victory is explained by the linear regression with the difference of total yards. Or, 52.88% of the variation is explained by other factors.

Conclusion

It has been shown through this analysis that there are some differences between the home and away team. By rejecting the null hypothesis that the home team margin of victory is equal to zero, it is significant to say that the home team does score more points. The home team also gains more total yards which was found by rejecting that the home and away mean total yards were equal. This makes sense from the viewpoint of a fan; being as if you move the ball offensively, it will result in more points. Another null hypothesis that was rejected was the number of turnovers for the home and away teams are equal. By rejecting this, it shows an important aspect of the game is different for the home and away teams.

Although some aspects of the game were found to be different, there are some others that we were unable to reject the null hypothesis. By failing to reject the null hypothesis, it provides the assumption that the home and away measures are equal. This result is much less significant than when the null hypothesis is rejected. This analysis shows that the home and away teams have the same time of possession and commit the same number and yardage of penalties. The null hypothesis that the margin of victory is equal throughout all 6 BCS conferences was also failed to be rejected. This leaves us with the assumption that the same home field advantage exists throughout each of the conferences.

While analyzing the data for turnovers and penalties, it makes sense to test if the data fits a Poisson distribution. For the number of penalties, the home and away sets of data rejected that the data was random and fits a Poisson distribution. On the other hand, the number of turnovers failed to reject that the data fits a Poisson distribution. This shows that turnovers are assumed to be random events and can be modeled with a Poisson distribution. One other point that was shown in this analysis was that there is a linear relationship between the difference of total yards and margin of victory.

Ideas for further study:

- Analyze margin of victory compared with betting lines. It is common knowledge in the betting world that the away team is given 3 points. Why 3 points? Should these 3 points be used through each conference since they were found to have the same home field advantage?
- When betting on a game, there are side bets, or proposition bets that can be made. It would be of interest to do an analysis on total yards and turnovers by comparing with the proposition bets.
- Although margin of victory has been found to be positive toward the home team, I would like to know if it has always been this strong. By doing an analysis with time as a factor, it would show if home field advantage has changed over the years.
- When comparing the margin of victory for each conference, it would be interesting to see how the smaller conferences stack up to the big money ones. This paper made the assumption to only use the 6 BCS conferences but including conference like the Sun Belt and WAC would provide an interesting answer.
- As shown in Appendix C, the histograms for the margin of victory in each conference appear to take a different shape. A more detailed analysis as to why conference histograms take a different shape but have equal means would provide some useful insight.

- Do some individual team analysis and the types of the opponents they face. Is the home field advantage there because of the level of competition or other factors?

References

[1] Bennett, Brian. June 14, 2012 http://espn.go.com/collegefootball/story/_/id/8047787/college-football-facilities-arms-race-proves-recession-proof.

[2] Beyond the bets. "Home Field Advantage rankings: Oklahoma's Memorial Stadium tops Paul Bessire's list". 2012 http://www.beyondthebets.com/home-field-advantage-rankings-oklahomas-memorial-stadium-tops-paul-bessires-list/

[3] Engineering Statistics Handbook, 7.2.1.3. Anderson-Darling and Shapiro-Wilk Tests. http://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm

[4] ESPN. "College Football Scoreboard". 2003-2012. http://scores.espn.go.com/college-football/scoreboard

[5] Hun Myoung Park, Ph.D. "Univariate Analysis and Normality Test Using SAS, Stata, and SPSS". Indian University 2008. http://www.indiana.edu/~statmath/stat/all/normality/normality.pdf

[6] Jamieson, Jeremy P. Journal of Applied Social Psychology, 2010, 4, 7, pp 1819-1848, 2010 Wiley Periodicals. http://www.wjh.harvard.edu/~jamieson/JJ_JASP.pdf

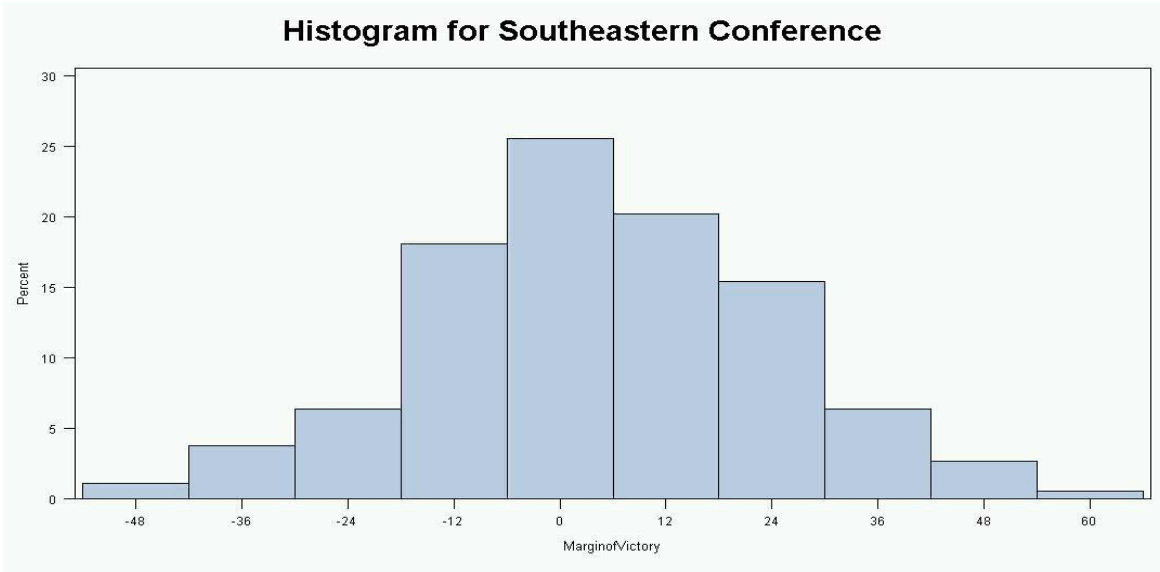
[7] Larsen, Richard and Morris Marx. "An Introduction to Mathematical Statistics and its Applications". 4th edition. 2006, Pearson Prentice Hall.

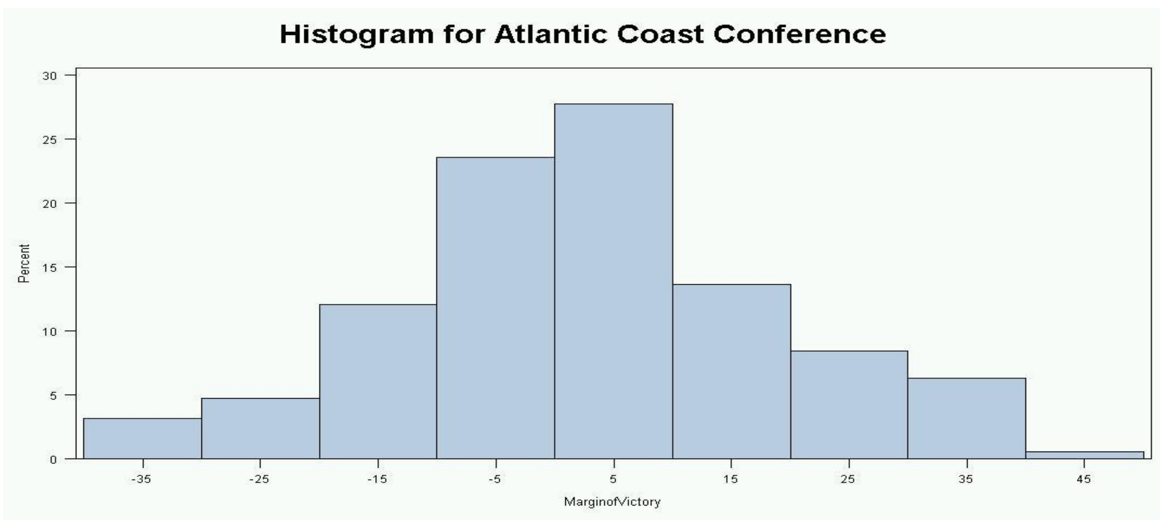
[8] Rivals.com, from YAHOO.COM. "College Football Scoreboard". 2003-2012. http://rivals.yahoo.com/ncaa/football/scoreboard

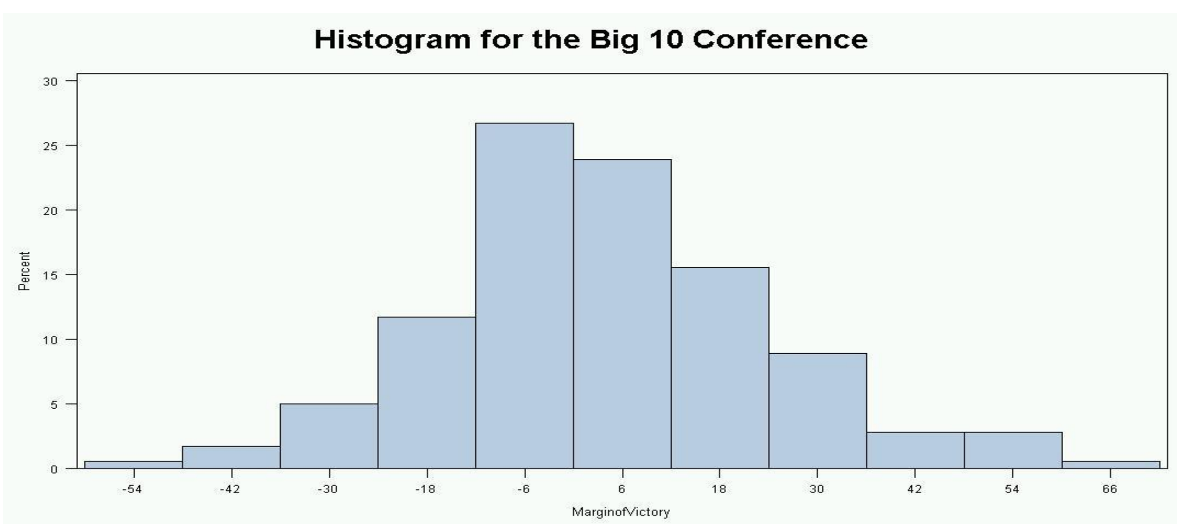
[9] Yendell, Brian. "Linear Models in SAS". 1995. http://www.stat.wisc.edu/~yandell/software/sas/linmod.html

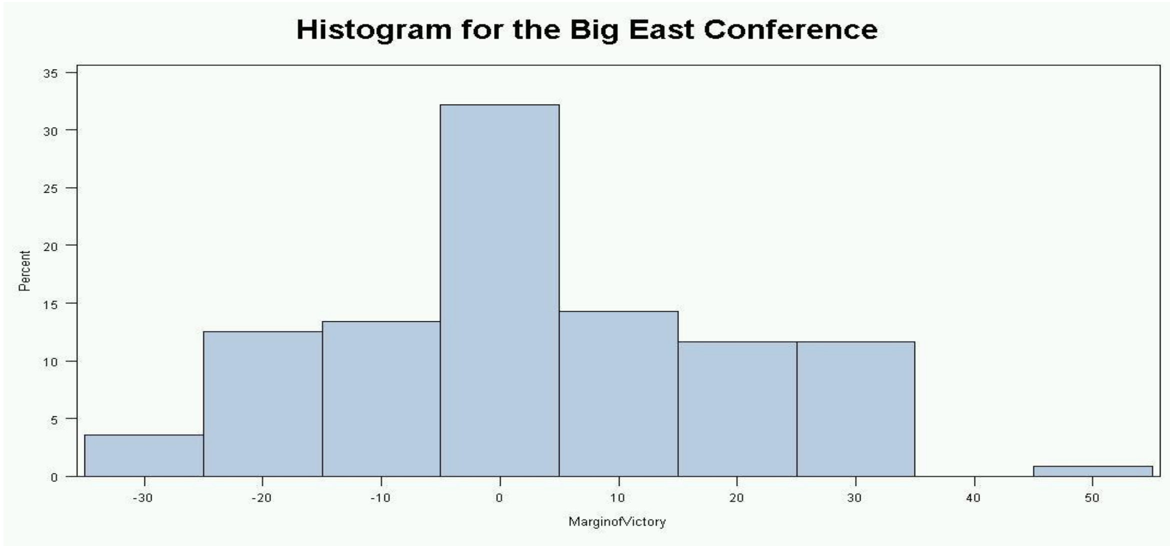
[10] Zar, Jerrold H. "Biostatistical Analysis". 5th Edition. 2010 Pearson Prentice Hall.

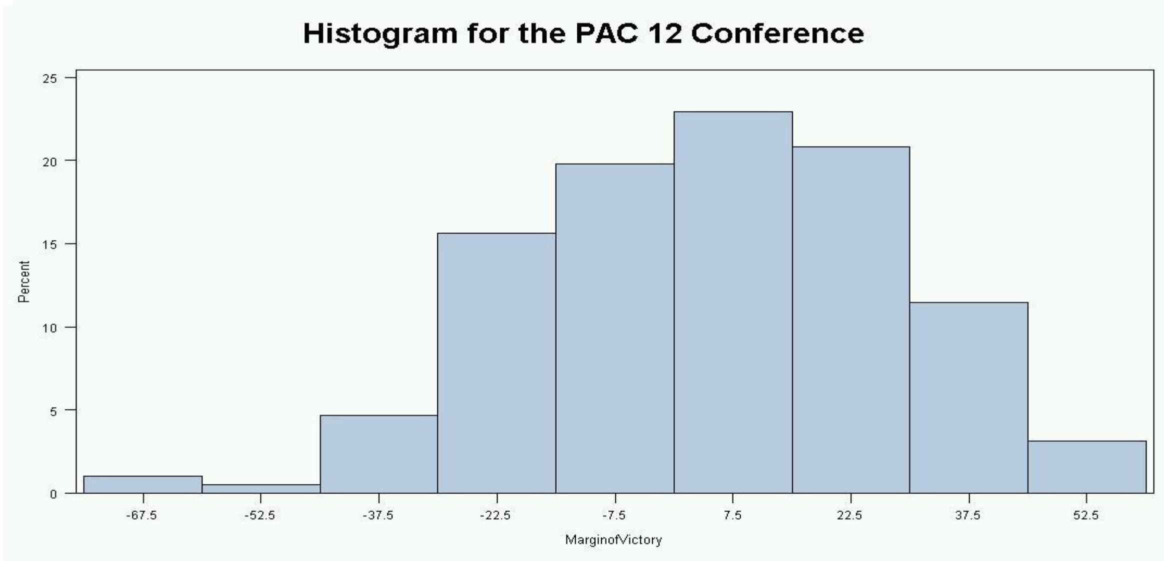
Appendix A

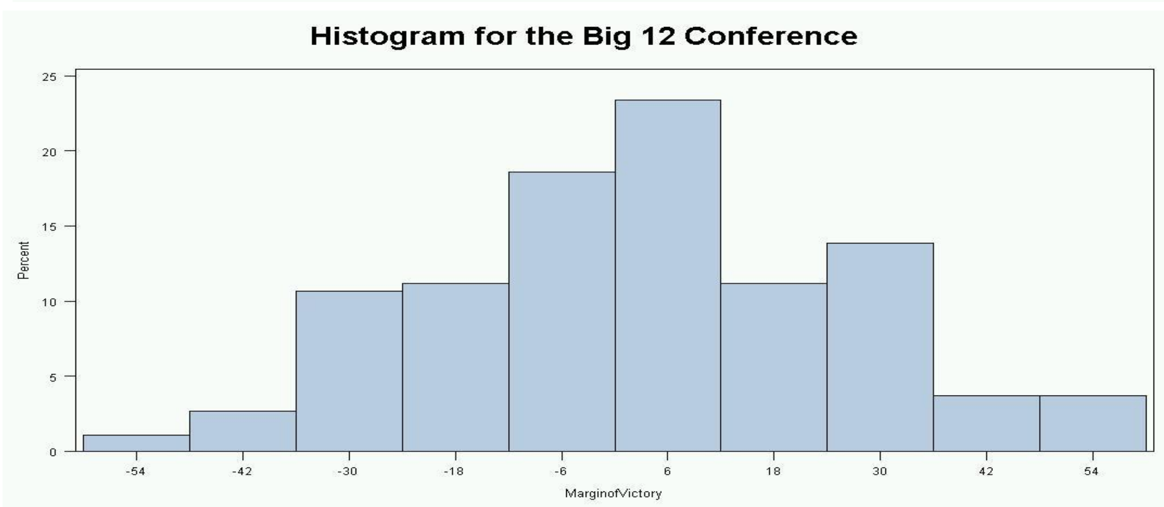

Histogram for Southeastern Conference
Percent
30
25
20
15
10
5
0
-48
-36
-24
-12
0
12
24
36
48
60
MarginofVictory


Histogram for Atlantic Coast Conference
Percent
30
25
20
15
10
5
0
-35
-25
-15
-5
5
15
25
35
45
MarginofVictory


Histogram for the Big 10 Conference
Percent
30
25
20
15
10
5
0
-54
-42
-30
-18
-6
6
18
30
42
54
66
MarginofVictory


Histogram for the Big East Conference
Percent
MarginofVictory


Histogram for the PAC 12 Conference
Percent
MarginofVictory


Histogram for the Big 12 Conference
Percent
MarginofVictory