Determining the Success of NCAA Basketball Teams through Team Characteristics

The Honors Program
Senior Capstone Project
Student's Name: Raymond Witkos
Faculty Sponsor: Alan Olinsky
April, 2010

Table of Contents

Segmentation of the Success of NCAA Basketball Teams by Team Characteristics

Senior Capstone Project for Raymond Witkos

ABSTRACT

Every year much of the nation becomes engulfed in the NCAA basketball postseason tournament more affectionately known as "March Madness." The tournament has received the name because of the ability for any team to win a single game and advance to the next round. The purpose of this study is to determine whether concrete statistical measures can be used to predict the final outcome of the tournament. The data collected in the study include 13 independent variables ranging from the 2003-2004 season up until the current 2009-2010 season. Different tests were run in an attempt to achieve the most accurate predictive model. First, the data were input into Excel and ordinary least squares regressions were run for each year. Then the data were compiled into one file and an ordinary least squares regression was run on that collection of data in Excel. Next, the data were input into Minitab and a stepwise regression was run in order to keep only the significant independent variables. Following that, a regression analysis was run in Minitab. The coefficients from that regression analysis were input into a file with the 2009-2010 data in an attempt to test the model's results against the actual results. All of the models developed, except one for the year 2005-2006, were determined to be significant. There were 6 significant independent variables determined. The final results showed that although the model developed through the study was significant, the ability to accurately predict the outcomes is very difficult.

INTRODUCTION

This paper will present an investigation the determination of team success in the NCAA March Madness tournament through team characteristics. Before beginning the statistical analysis and research, it is important to understand the topic. For that reason, an overview of the March Madness tournament will be included in the final paper. In addition, it is important to understand the statistical categories and methods that were used.

Coaches and players of the sport of basketball alike are constantly trying to gain advantages over their opponents. The competitive nature within humans has been around since we arrived on this earth. The difference is now we are competing for sport, not for survival. Every student, employee and business in this world is trying to gain an advantage over their

competition. In the world of college basketball, this is no different. Over the past decade college basketball has gotten more and more competitive. Revenue for the postseason tournament, March Madness, increases every year. Coaches and players are constantly trying to get a leg up on their opponent. This can be seen through hours of watching film, lifting weights, running, swimming, individual instruction, practicing as a team and many other activities.

Although all of these different angles of gaining an advantage seem to be popular, the use of statistics is not. Very often a college basketball coach uses statistics, but not to the extent they could. For example, a coach will often look up where the team ranks in the league in a certain statistical category. Also, almost every college coach will look at the statistics of their team during halftime of a game, with the assistant coaches keeping an eye on the numbers throughout the entire game. There are many uses a coach may have for statistics, but these numbers are not utilized to their full capacity. A coach rarely designs a practice around what statistical areas need improvement, rather the coach subjectively decides what needs improvement.

The main purpose of this study is to identify the statistical factors that can help segment the success of the teams in the tournament. The research will analyze what statistical factors have a strong correlation to a team's success in March Madness. Many different forms of statistical analysis will be researched and performed. The methods to be considered include linear and logistical regression, discriminant analysis, cluster analysis, data mining, decision trees and factor analysis. The data to be input into the models will include traditional quantitative data and non-traditional qualitative data. The traditional quantitative, data which will be explained later, include field goal percentage, points per game, points allowed per game and many others. The nontraditional qualitative data will include leadership, ability to play on the road, and others which are defined later in the paper.

As well as addressing the main subject, there are issues that this study has attempted to answer. These issues include the following:

- Are there concrete statistical factors that are common between a team's regular season statistics and their postseason success?
- Are there factors that have not been looked at extensively in the past that have strong links to a team's performance in the tournament?
- Are there distinct statistics that should be emphasized in order to help a team increase their chance at success in March Madness?

LITERATURE REVIEW

Before deciding how to define the models to be used for this study, I decided to see what was already done in the subject. By seeing what succeeded and what failed, in addition to what the limitations were of each model, I was able to include the most relevant information. At the same time, it was important that I did not neglect any information that may be able to alter my results.

An article titled "Probability Models for the NCAA Regional Basketball Tournaments," written by Neil C. Schwertman, Thomas A. McCready, and Lesley Howard began the research. The article, written in 1991, used a rather basic model. The purpose of the study was to try and determine the probability of each team winning the National Championship. The models used basic probability equations. The model was based on the assumption that each game is independent of the others and that "intangible effects" are being ignored in these models (Schwertman, and McCready 35-39).

The first thing that needed to be determined in the study was that each team would have to play. The study concentrated on using the strongest team, ranked the 1 seed, winning the tournament. The teams that the 1 seed would have to play were determined by using the probability equations. The probability of any team winning the tournament obviously depends on the ability of that team to win each game.

An important aspect the study mentions is the fact that not every team has an equal chance of winning each game. Obviously the 64th team in the tournament is not the 1st ranked team for

a reason. Every year, each team is given a ranking by professionals based on statistics and their own subjective feelings.

The study suggests three different probability models to be used. The first model computes the probability equal to the opponent's seeding divided by the sum of its own seeding plus that of its opponent.

- P (i, j)= j (i + j)
- 1 Seed = (16/ [16+1]

This model heavily favored the higher seed. The 1 seed is twice as likely to defeat the 2 seed as the 2 seed is likely to defeat the 1 seed (Schwertman, and McCready 35-39). Based on past tournaments, this probability is too high for the one seed and too low for the 2 seed. Although the 1 seed should have a greater probability of winning, twice as likely is too high.

The second and third models are very similar. The second model the study suggests assumes linearity based on the difference of seeding between the two teams. The probability equation for this model looks like this:

$$P(i, j) = .5 + \text{delta}\ (j - i) \quad 0 <= P(i, j) <= 1$$

This model uses a uniform distribution of team strength. As mentioned, the third model is very similar except that a normal distribution of the teams is assumed. Another assumption these models make is the 64 teams that were selected to play in the tournament are in fact the 64 strongest teams in the country. The third model used z scores in an attempt to increase the accuracy of the model (Schwertman, and McCready 35-39).

After developing the models, the predictions of each model were compared to the actual results that occurred during the tournaments of the six previous years. Using the data, the third model was deemed to be most accurate. The chi-squared test was applied to each model in order to attempt to see what model was most relevant and provided the best fit for the data. The third model was revealed to have the lowest chi-squared value and the best model of the three.

This article provided very basic and introductory concepts to the study. The article did provide a good background of information on the topic and where the concepts and studies began. Although this was not the first model ever drawn up for the subject, it was basic and allowed a view of where these studies began. Without having an understanding of the simple concepts, it would be impossible to draft up more complex models.

The article mentioned that there are several weaknesses in the models. The study mentioned that each game is affected by outside influences, affecting the accuracy of the study. Unfortunately just addressing this and assuming independence of these factors and each game alters the accuracy of the model. The article refers to "intangible effects" that may alter the outcome of a game. For the study being conducted, these effects must be incorporated as accurately as the can be.

In 1996 Brett C. Holbrook Neil C. Schwertman. and Kathryn L. Schenck published an article titled "More Probability Models for the NCAA Regional Basketball Tournaments." The article introduced the concept of the probability of a team winning the championship being based on the probability of that team winning one game at a time and the need for each different path to the championship to be analyzed. Holbrook et al. referenced previous studies that had been done and also recognized a need for more accurate results from their models.

Introducing their models, Holbrook et al. again asserted each game to be independent and that the probabilities remain constant throughout the tournament. Seed positions were assumed to be a practical measure to use in the models since they were determined by experts before the tournament began. This article built upon the use of a linear straight line method. The study used a total of 11 different models. The new models presented all used a basic linear equation as show below.

$$E(Y) = \beta o + \beta 1\ (S(i) - S(j))$$

Three of these models being used, labeled models nine through eleven, were from the previous article cited and were nonlinear. The 8 new models developed depended on the type

of regressions, type of intercept, and type of independent variable (Holbrook, Schenk, and Schwertman 34-38).

After developing the models, specifications needed to be made to each. One assumption the models in this study make are that since using linearity, the difference in strength between seeds 1 and 4 is equal to the difference in seeds 13 and 16. This assumption was especially apparent in the first and second models in the study. This is likely untrue, as the 1 seeds will be much stronger than the 4 seeds and there may not be that much difference between the 13 and 16 seeds.

The third and fourth models in the study used a nonlinear function in order to more accurately incorporate team strength. A normal distribution was assumed. By inputting the number given to each team, for example 292 is the best team in the tournament since there were 292 division I programs at the time, the percentile of that team can be determined. From that percentile, the corresponding *z* score was found. Models five through eight use the same concepts as models three and four. The only difference is that logistic regression was used instead of linear regression.

After obtaining the results for each method, the models were analyzed by using a chi-square statistics. This was used as an analysis of the fit of the models. An interesting finding after using the chi-square values was that the models that were best at predicting the probability of any given seed winning a single game were not as good as other models in predicting the winner from that specific region. Models 7 and 8 which used logistics, with and without intercept, and z-scored seeds around the middle of the field at predicting an individual game, but they were the best at predicting the regional winners. Models 3 and 4, which used ordinary least squares, with or without a specific intercept, and z-scored seeds, were the best at predicting an individual game, but were in the middle of the field at predicting the regional champion (Holbrook, Schenk, and Schwertman 34-38).

The study done by Schwertman et al. was a step up from the 1991 study in which Schwertman contributed. This study introduced 8 more models bringing the models introduced to this point up to 11. These additional models were based on a linear equation which was an

improvement from the basic probability equations introduced in the first study. In the end of the study it was discovered that each model had strengths and weaknesses. There was a tradeoff between being able to have a model that was accurate in predicting a single game and a model that could effectively predict the winner of a region.

The study was a step in the right direction. The models in this study were admittedly simplistic. Although more models were introduced, simple equations were still being used. Another weakness to this study was the fact that the only independent data being input into the models were the seeds of the teams. Granted, the seed of the team is based on what experts have determined the team should be ranked at, but many different factors get left out. Although this study had its weaknesses and deficiencies, it incorporated more models and provided a good idea of the challenges involved with creating an effective model.

Kurt T. Dirks wrote an article in 2000 about the effect of trust on team performance. The article, "Trust in Leadership and Team Performance: Evidence from NCAA Basketball," tried to prove two main points. First the study attempted to examine the assumption that a team's trust in the leader had an effect on the team's performance. The second point was to examine the role trust plays in the relationship between past performance and future performance.

Before getting into the models and statistical data, Dirks defined trust and the different measures being used as they pertained to the study. For this study, trust was defined as "an expectation or belief that the team can rely on the leader's actions or words and that the leader has good intentions towards the team (Bass, 1990), causing others to be vulnerable to him or her. As I discuss later, I also take into account the extent to which team members trust each other, because they are also vulnerable to each other, given their independence (Dirks, 2000)." Dirks went on to mention that trust in leadership is important for the success of a team. Reasons for this include the necessity of a team to accept the leader's goals and decisions. If team members do not trust in the direction of their team leader, they are less likely to properly perform their tasks and, consequently, less like to succeed as a team. The three main factors of trust were asserted to be vulnerability to the leader, vulnerability to each team member, and uncertainty. Dirks added that teams provide a good setting in which trust can be measured. There are players on each team that are vulnerable to the coach who controls the playing time,

play selection, and other key decisions. The players are also vulnerable to each other and there is uncertainty among the players over key issues (Dirks 1004-1012).

For the sample, Dirks used data from 30 teams NCAA men's college basketball teams from Division I and III. From the 30 teams, 11 Division I and 19 Division III, 355 representatives from the teams completed surveys. Data was collected from different sources and different methods in an attempt to remove an inflated statistical relationship. The data was collected at the beginning of conference play which allowed the team to have been undergoing official team practice for at least 6 weeks by then. This would allow a certain level of trust to have formed by this time. The study used measures of predictors, used as control variables, of team performance. The coach's control variables were experience and career record and the players control variables were talent, trust, and tenure. The two variables of past performance and preconference performance captured data of the school, coach and players (Dirks 1004-1012).

One thing that was of particular interest in the study was how he would measure each variable. Several of the variables were rather traditional and straightforward. These variables included team performance, measured by the ratio of wins to games played during the conference schedule; prior team performance, measured by the ratio of wins to games played during the conference schedule over the previous four seasons; coach career record, measured by career winning percentage times 1 minus 1 over number of seasons coached; players' tenure, measured by the length of time players have played under the coach; coach's experience, measured in number of games coached by the head coach over his career; preconference performance, measured by winning percentage of the games prior to conference play (Dirks 1004-1012).

With the more traditional measures defined, the more pertinent variables to this study remained. First, trust in the coach was measured by asking each player to complete a survey which asked questions that directly indicated their level of trust in the coach. The questions were posed in simple yes and no responses. The same survey and technique was completed for trust in teammates, just each player responded how they felt about the trust they could place in another player (Dirks 1004-1012).

Next, the team talent variable had to be measured. This variable was measured by determining how many players from each team were named to the all-conference team. All-conference teams consist of a second team, first team, and an MVP. Each all-conference team was evaluated and if a team had a representative on an all-conference team they received a score equal to the school's number of representatives on the all-conference team divided by total number of representatives on the all-conference team. In addition to this, a weight was given to each level of achievement. The MVP rank was given a multiplication of 1.0, the first team had a weight of .66 and the second team had a weight of .33. This allowed for each level of status to be weighted (Dirks 1004-1012).

After gathering the data, Dirks used regression to measure each point previously stated. After analyzing the results, it was shown that trust, as defined in his study and measures, the two hypotheses were confirmed. Trust in leadership does have positive effects on team performance and trust in leadership modes mediates between past team performance and future team performance.

The study done by Dirks was of particular interest to this study. Dirks not only took traditional quantitative statistics such as wins and winning percentage, but he was also able to take variables like team leadership, team performance and team talent and assign values to them. Dirks' study reflected what this study is attempting to do. The regression analysis performed by Dirks was simple yet very accurate. All of his variables pointed toward a significant correlation between trust in leadership and other variables and the variance in team performance. The R-squared value, for example, was .66 suggesting a somewhat strong correlation.

Another interesting article related to this study, titled "Factors Associated with Success among NBA Teams," was written by Anthony J. Onwuegbuzie. The study took data from the 1997-1998 NBA regular season and analyzed it in an attempt to determine the factors that best predicted the ultimate goal of a team, winning percentage. At the time of his study, there was little investigation ever done in the area. Many studies were done on the effects of specific skills, but never tried to determine that skill's effect on team success. The stated goals of his study were to find what factors directly associated with skill level, such as field goal

percentage and number of rebounds best predict a team's winning percentage. A second goal Onwuegbuzie set out to complete was to determine whether offensive or defensive factors were better at predicting a team's success.

Onwuegbuzie obtained 21 different variables from the NBA website for the 1997-1998 season. Winning percentage was treated as the dependent variable with 20 other variables being the independent variables. Some of the variables used in the study included number of points scored per game, field goal percentage, three-point field goal percentage and average number of points scored per game by the opposing team (Onwuegbuzie web).

The study used the Statistical Package for the Social Sciences (SPSS). After the calculations were done, there needed to be an adjustment made for Type I error. In the study, regression analysis was used. First, and all possible subsets (APS) multiple regressions were used. This was done to try and figure out which combination of variables predicted winning the best. In the equation 16 variables were used and two variables, field goal conversion percentage and average three-point conversion percentage of the opposing teams had the most significant correlation. The equation that was used was as follows:

> winning percentage= -159.53 + {(7.90) X field goal conversion percentage} – {(4.24) X average three-point conversion percentage of the opposing teams}

Onwuegbuzie's analysis revealed significant findings. First, for every 1% increase a team has in field goal percentage, there was a 7.9% increase in winning percentage. Second, for every 1% increase in the three-point conversion rate of the opposing teams there was a 4.24% decrease in winning percentage. A 95% confidence interval was used for these results. The field goal percentage variable accounted for 61.4% of the variance in winning percentage and the three-point conversion rate of the opposing teams accounted for 18.9%. Together these two variables accounted for 80.3% of the variance in winning percentage (Onwuegbuzie web).

Onwuegbuzie's article was very important to this study. His study determined some of the variables that are significant to an NBA team's winning percentage. His study also showed

that the offensive variable had more of an effect than the defensive variable. This was not extensively looked into, but the results clearly suggested the conclusion.

Onwuegbuzie's study, however, should have been expanded to different seasons to try and determine if the effect the variables had was altered from year to year. Also, this type of an approach seems reasonable to try and apply to college basketball. The game does change from the NCAA to the NBA level. At the same time, elements of Onwuegbuzie's article can be altered and shaped to help predict the most influential variables on an NCAA basketball team's ability to succeed in March Madness.

Another interesting article that pertained to this study was written by Richard F. Uttenbach and Irvin G. Esters in 1995. The study was titled "Utility of Team Indices for Predicting End of Season Ranking in Two National Polls." Although the article was not attempting to predict a team's success in the NCAA tournament, the methods used and results were relevant.

In their research, Esters and Uttenbach were attempting to use six independent variables including points per game, points allowed, number of field goals, number of free throws, number of three-point goals and number of rebounds to determine if there was an effect on the dependent variables. The dependent variables in this study were selected to be a team's ranking at the end of the year. Two separate polls were used, the USA Today/ CNN scale and the United Press International (UPI) rankings. Because of the wide-spread usage and acceptance of the polls as a rank of success, Esters and Uttenbach accepted them to be input into their model (Esters, and Uttenbach 216-224).

As mentioned, the two main goals of Ester and Uttenbach were to determine if team statistics can be used to predict the final rankings of a team in the medial polls and to identify the variables that provide the strongest predictions. The data was gathered from the 1991 NCAA basketball season. Each of the 64 teams that made it into the March Madness tournament had their statistics analyzed.

In the study, three analyses were conducted. First, statistics were calculated for each independent variable. Second, two regressions model analyses were computed and the

significance of each was tested. Third, beta weights were tested for significance and used to try and determine the most consistent and effective prediction models.

The first hypothesis that Ester and Uttenbach were hoping to confirm was that a significant relationship existed between the six predictor variables and the USA Today/ CNN final polls. A null hypothesis was set up and rejected, meaning a relationship existed. The R-Squared value of the equation was .33 meaning that 33% of the variance in the dependent variable was accounted for through the six predictor variables (Esters, and Uttenbach 216-224).

The second hypothesis was that there was a significant relationship between the six predictor variables and the end of the season UPI ranked score results. Again, a null hypothesis was set up and rejected meaning that a significant relationship did exist. The R-Squared value found in this equation was .49 meaning 49% of the variance was accounted for by the variables included in the study (Esters, and Uttenbach 216-224).

Although each hypothesis achieved similar findings, the variables that were considered to be statistically significant were different between the two. In the USA Today/ CNN poll, only points per game was a strong predictor. This result was not surprising because basketball is based on scoring more points than the other team. This finding suggests although teams may focus on defense to try and slow their opponent, in the end the only thing that matters is having more points on the board than the opponent.

In the UPI scale three variables were found to be statistically significant. The variables were points per game, points allowed, and number of rebounds. Again, points per game was a significant variable.

With this study, there remained a level of subjectivity in the outcome. The outcome was based on final poll positions, either USA Today/ CNN poll or UPI poll, rather than a concrete conclusion like national champion. For this reason, biases by voters could affect the final results. Also, additional testing and increased sample size should be used to test the models that were derived. Because only one year of data was used, the model will naturally fit the observed data well.

Although the study had certain weaknesses, once again there were useful points to be taken into consideration. Although my study aims to use more variables than the study completed by Ester and Uttenbach, one variable that cannot be ignored is points per game. This variable has a significant strength of prediction and must be taken into account no matter what other variables a model might be using.

J.A. Deddens and K. Steenland wrote an article in 1997 titled "The Effect of Travel and Rest on Performance of Professional Basketball Players." Unfortunately, the full article was unable to be located, but an abstract explaining the study and results was.

Deddens and Steenland performed a study of NBA teams over a span of eight seasons. In seasons ranging from the 1987-1988 seasons to the 1994-1995 season, 8,495 separate games were analyzed. The dependent variable in the study was team performance. In their study, they chose to measure performance by points scored per game (Deddens, and Steenland 366-369).

The findings of the study revealed that with more than one day between games, the home team's score increased by 1.1 points and the visitor's score increased by 1.6 points. The peak performance occurred when there were 3 days between games. The effects of the study were suggested to be attributable to jet-lag, however, the more common reason is most likely the fact that the players' bodies have less time to recover (Deddens, and Steenland 366-369).

Although the Steenland and Deddens abstract provided limited information, it did provided valuable suggestions. The NBA schedule calls for professional basketball teams to travel all across the country from night to night. Although the NCAA basketball schedule is about one-third as long as the NBA schedule and much less strenuous with the travel, there could be some use for Steenland and Deddens results. The NCAA March Madness tournament calls for teams to play away games. There are no true home games, all the games are hosted at neutral sites. In addition to this, there are shortened amounts of time between each game in the beginning of the tournament. Teams may only get one day off between games as opposed to the regular season when days between games range anywhere from 1 to 3 days. The study performed by Steenland and Deddens provided useful concepts to be explored.

RESEARCH MODEL DEFINED

The hypothesis of this study is that team success in the NCAA Men's March Madness Tournament can be determined through the analysis of independent variables. This study will attempt to prove that there are independent variables that can be used to determine the probability of a team succeeding in March Madness. The study will also attempt to determine which variables have the strongest predictive power. Finally, the study will attempt to segment the top eight teams in the tournament. This means that characteristics that are common among the top eight teams from each year will be will be analyzed. From this analysis the study will try to find which variables are common among these teams.

As the literature review revealed, there has not been too much research done in this specific area. Out of the research that has been done, most of it concentrated on the independent variables that can be considered traditional. These variables include quantitative variables such as points per game, field goal percentage, three point field goal percentage, free throw percentage, rebounds per game, assists per game, blocks per game, steals per game, turnovers per game, opponents' field goal percentage, opponents' points per game, opponents' three point percentage and others. All of these statistics are important and will still be used in the model in this study. There is no replacement for quantitative statistics in this study.

Where this study will differ and go into much more detail than any other study done in the past is the use of qualitative statistics. This study will attempt to use experience, road play, head coach leadership, campus support, team talent and possible a few other qualitative statistics. Experience will be measured through the number of seniors, juniors, sophomores, and freshmen on the team. Road play will deal strictly with the team's road record. Head coach leadership will be measured by an equation dealing with amount of years experience in coaching and the winning percentage over that time period. Campus support will be measured through home attendance at regular season and conference tournament games. Finally, team talent will be determined by an equation using how many players make it to an all-conference team at the end of the year.

The data will be collected through online websites. Although a database has been located online, the site is temporarily down. For this reason, sites such as ESPN.com and NCAA.org

may have to be used for the data collection. If these websites are used, the data will have to be input manually rather than downloading an Excel file from the website with the database.

After gathering the data this studying will be using various statistical methods in an attempt to analyze the data. The statistical methods to be performed include linear and logistical regression and stepwise regression. Methods may be added as deemed appropriate throughout the study.

METHODOLOGY

Data Used

After the research model was roughly defined, the data needed to be collected. For collecting the data the study attempted to go back the maximum amount of years while still keeping the data current and useful. Data from the 1980s, for example, would not be relevant simply because of changes in the way college basketball is played and coached. The study included the 7 most recent years, beginning with the 2003-2004 Men's Basketball season and ending with the most recent 2009-2010 Men's Basketball season. The data input into the Excel spreadsheets included each season's statistics including the regular season statistics and the conference tournament statistics. Any play after the conference tournament, the NCAA March Madness tournament or any other postseason tournament, was excluded from the data input. The data was excluded for the simple reason that this study was attempting to predict the variables that best predicted a team's success during the tournament. Therefore, any data from the actual March Madness tournament needed to be excluded in order to retain the initial purpose of the study. Most of the data collected was taken from the official NCAA website. There were some entries which needed to be looked up on each team's website as it was unavailable through the NCAA website.

After inputting all of the data into Excel there were 13 independent variables, and 1 dependent variable. The dependent variable, as mentioned, was the final ranking of a team in the NCAA tournament. The independent variables were Average Points per Game, Field Goal Percentage, Three Point Field Goal Percentage, Free Throw Percentage, Rebound Margin per Game, Assists per Game, Blocks per Game, Steals per Game, Turnovers per Game,

Opponents' Points per Game, Opponent's Field Goal Percentage, Experience, and Campus Support. A total of 5785 different independent variable entries were made for the 7 years included.

Although most of the independent variables are self-explanatory and widely used, such as points per game, rebounds, etc., a few of the independent variables need clarification. First, Three Point Field Goal Percentage per game was only included if a team attempted at least 5 three point field goals per game. If the team did not attempt at least 5 three point field goals per game, the NCAA database gave that team an "N/A" for that year. This is a relatively low number of attempts per game and the number of teams that received an "N/A" over the 7 years included was only 40. In the past 7 years there have been 448 teams competing in the NCAA March Madness tournament, so only 8.9% of the teams participating received an "N/A." The "N/A" was then given a 0 for the team for that season. The reason a 0 needed to be input rather than "N/A" was that the statistical software only would allow numbers to be input. A 0 was appropriate as these teams failed to attempt enough three point field goals to count in the study.

The second independent variable needing explanation is Rebound Margin per Game. Typically when a team's statistics are listed "rebounds per game" is a popular statistic used. Although rebounds per game is a valuable measure of how a team "attacks the glass" and is able to efficiently offensively and defensively rebound, it fails to measure the opponent's rebounding against theirs. For this reason, Rebound Margin per Game was used. Rebound Margin per Game measures a team's rebounds per game, subtracts the opponent's rebounds per game and then yields the rebound margin per game. This measure is a better overall picture of how well a team is rebounding the basketball with respect to their opponents.

The third independent variable requiring explanation is one that is not typically included in past models attempting to measure a team's success, whether in the regular season or in the postseason. This independent variable was "Experience." The variable was measure by taking the roster of each team and assigning each player a value based on their year. A Senior was assigned a value of 4, a Junior received a 3, a Sophomore received a 2, and a Freshman

received a 1. A team with more Seniors and Juniors would receive a higher Experience ranking than a team that consisted mostly of Freshmen and Sophomores.

The final independent variable that needed explanation is "Campus Support." This variable is another area that is not typically included in these types of studies. This study attempted to measure Campus Support by the average attendance at home games throughout the course of a season. Again, it was important that only home games were included in order to exclude the attendance figures that were taken into account when a team played on the road or at a neutral site.

Explanation of Final Rankings

For the NCAA March Madness Tournament it is difficult to determine which team ends up coming in last, 64th place. For all of the rounds except for the last two rounds it is difficult to figure out which team placed where in each round. For this reason, the rankings were done in levels. For example, any team that lost in the first round received a 33 as their ranking. This meant that 32 teams, half of the field, would receive a 33 as a ranking. The next round gave 16 teams a ranking of 17. This system of ranking teams in levels follows all the way up until the last four teams, the Final Four. The winner of March Madness received a 1 and the loser of the championship game a 2 for a ranking. Next, the team that lost to the eventual champion in the semifinals received a 3 ranking and the team that lost to the eventual runner up received a 4 ranking. This system of ranking the 3rd and 4th place team is often used in tournaments that do not have an explicit game to determine the 3rd and 4th place teams. From there on out, the levels were used as a ranking device again, as the next round down had 4 teams losing, all receiving rankings of 5.

Excluded Data

Originally there were other quantitative measures that were going to be input as independent variables in order to attempt to gain a more accurate and well-rounded model. Unfortunately, some of these measures were not able to be gathered. First, a measure of each team's coach was going to be attained and ranked using a formula cited earlier in this paper. Unfortunately, the NCAA website did not provide statistical data about the coaches and in addition, school websites were not consistent with the facts they provided about a coach.

The next measure that was originally considered, but later omitted was ranking a team's talent based on the number of players from each team that made an all-conference team. This measure was decidedly left out for the a couple reasons. First, an all-conference team selection is highly subjective to the coaches and media that vote. Second, just because a team does not have a player on the first or second all-conference team does not mean they do not have a good team. Perhaps that team has an abundance of talent that is spread out among their team rather than kept within one or two all-conference players.

The third measure that was left out was a team's road performance. Unfortunately, this measure had to be left out because of the lack of records. On the NCAA website there were no data on this. The only way to research the road record of a team was to manually go through each team's website, view their historical schedules, determine which games were away games, and input the record of those games. After beginning to do this, the number of teams that fully included their team's record back to the 2003-2004 season began to diminish. Many of the big schools had full records, but the mid-major schools often did not include records back that far. For the reason of inconsistency, this measure was omitted.

Within the data input into Excel, there was only one set of data that was unable to be found in the database. The 2007-2008 Turnovers per Game data was not provided for this year. In order to input the data into a formula that would be accepted by both Excel and Minitab, there had to be some value in those cells. The agreed handling of this was to take an average of all the other years Turnovers per Game and input that number into the Turnovers per Game for each team during the 2007-2008 season. This number of Turnovers per Game given to each team during the 2007-2008 season was 13.9.

The final set of data that was omitted was Experience measure and the Campus Support measure for the 2009-2010 season. At the time of the study the NCAA website did not provide either of these as the season was still in progress.

Testing Methods

The first method used for testing was regression with the Excel program. The Excel program used a least squares regression to analyze the data. The independent variables were run

through the program and tested to see what relationship each had on the dependent variable. The output gave a significance for the overall model, *Significance F*, and the significance for each independent variable, the P-value. For this study an alpha of .05 was assigned. This meant that any independent variable with a P-value below .05 was considered significant, with the lower the P-value, the better. Any independent variable close to the cut-off point of alpha, .05, would also be considered significant.

Starting with the 2003-2004 season and ending with the current 2009-2010 season, each regular season's data was run through a regression in Excel. After this was done, it was decided that it would be most beneficial to compile all of the data into one file and run the regression for the previous 6 years as one regression. This would allow for a more complete set of data and a better overall picture of the significant independent variables.

In addition to running the ordinary least squares regression in Excel, it was decided that a stepwise regression would be beneficial to the study as well. The reason a stepwise regression was included was because stepwise regression only includes the significant independent variables in the model. The ordinary least squares regression run in Excel includes all independent variables regardless of their significance. When only significant independent variables are included, it is more clear what influence those variables have on the model. For the stepwise regression, a model was not run for every year. The model run for the stepwise regression was run for the 2003-2004 season through the 2008-2009 season. Because stepwise regression is not available in Excel, Minitab was used to run the analysis.

After a stepwise regression was run in Minitab, a regression analysis was run in Minitab only including the significant variables. This allowed the regression to only include those variables that significantly influenced the overall model. After this stepwise regression was attained the final step of the study could be completed.

The last analysis of data included using the coefficients as determined in the stepwise and regression analysis in Minitab. These coefficients were then used with the current 2009-2010 regular season and conference tournament data. The result was a result for each of the 64

teams that made the March Madness tournament. The team with the lowest numerical number was the team that had the best ranking.

STATISTICAL RESULTS

Individual Years

As mentioned, the majority of emphasis was placed on the results that came from the historical data as a collective unit, the 6 years combined. However, it was important to look at each individual year. First, the model ordinary least squares regression model for the 2003-2004 season f-Stat had a significance of 0.0011. This low value means that the overall model was significant. After observing that the overall model was significant it was important to look at the P-value of each independent variable. Although the overall model was significant, only 2 variables were significant. Turnovers per Game had a P-value of .0048 and Opponents' Field Goal Percentage had a P-value of .0673 (borderline significant). Next, when looking at each significant independent variable, it is important to look at the coefficient of each. Turnovers per Game had a coefficient of 3.6104 and Opponents' Field Goal Percentage had a coefficient of 1.8386. This means that for each additional turnover a team commits their rank increases by 3.6104 and for each percent a team allows their opponent to shoot 1% higher, their rank increases by 1.8356.

Next, the 2004-2005 was viewed. Again the model was determined to be significant with a *Significance F* of .0414. This value was below the .05 cut-off and determined to be significant. The only independent variable that was considered to be significant for the 2004-2005 season was Free Throw Percentage, which had a P-value of .0303. The coefficient associated with Free Throw Percentage was -1.0769. The interpretation of this coefficient is for every additional 1% point a team averaged shooting from the free throw line, their rank decreased by -1.0769.

The 2005-2006 season was looked at next. The *Significance F* associated with the model was .2534. This high value meant that the overall model was not significant. Although the overall model was not significant, there was one independent variable that was considered borderline significant. Opponent's Points per Game received a P-value of .0699. The coefficient

associated with the independent variable was 1.2269. The interpretation for this variable is for each additional point a team allows their opponent to score per game, their ending rank increases by 1.2269.

Next, the 2006-2007 season was observed. This model had a *Significance F* of .0063. This low value again meant that the overall model was significant. Within this model there is only one variable that is significant. The Campus Support variable was significant with a P-value of .0248 and a coefficient of -.0006. The interpretation of this communicates that as a team increase the number of people that attend their home games by 1 person, their final ranking in the March Madness tournament decreases by .0006.

The next season to be taken into consideration was the 2007-2008 season. The *Significance F* had a value of .000014. Again, this revealed that the overall model for the 2007-2008 season was significant. Within the model there were 3 independent variables that were significant. Points per Game was considered significant with a P-value of .0074, Opponents' Points per Game with a P-value of .0002, and Campus Support with a P-value of .0001. Points per Game had a coefficient of -1.4885 meaning for every additional point a team scored per game, their rank decreased by 1.4885. Opponents' Points per Game had a coefficient of 1.8515 meaning for every additional point per game the team's opponents scored, their final rank increased by 1.8515 points. Finally Campus Support had a coefficient of -.0009 meaning for every additional person that attended a team's home games, a team's rank decreased by .0009.

Next, the 2008-2009 season was considered. The overall model had a *Significance F* of .0005. This low value again meant that the overall model for that season was significant. Along with the model there were 2 independent variables that were significant. First, Points per Game was significant with a P-value of .0354. The coefficient for this variable was -1.3285 meaning for each additional point a team averaged throughout the year their final ranking was decreased by 1.3285. The second variable that was significant was Campus Support with a P-value of .0013. The coefficient for this variable was -.0009 meaning for each additional person included in the average home attendance the final ranking of the team decreased by .0009.

Finally the current 2009-2010 season was analyzed. Although at the time of the study the March Madness tournament was still underway and the Experience and Campus Support variables were unable to be included, the model was still carried out. This model again received a *Significance F* which was under the cutoff point of .05. The value given for the overall model was .0004, suggesting the model was significant. In this model there was one significant variable. The Blocks per Game variable had a P-value of .0101. This variable had a coefficient of -6.2020, meaning for each additional block a team averaged throughout the regular season and conference play their final ranking in the March Madness tournament decreased by 6.2020.

Individual Season Conclusions

For the individual years all but one model was significant. The only year that lacked a significant overall model was the 2005-2006 season. In the 7 separate models there were 7 variables that were considered significant or borderline significant at least once. Turnovers per Game, Opponent's Field Goal Percentage, Free Throw Percentage, Opponents' Points per Game, Campus Support, Points per Game, and Blocks per Game were all significant or borderline significant at least one time.

These models provided valuable insight into which variables were most significant during which years. The most important models, however, were the ones that included all the most recent 6 seasons (not including the current 2009-2010 season). With more data entries, the more complete model would be able to provide a bigger picture to what the most important variables are.

Collective Historical Data

After analyzing the output of each individual year, all of the data entries from the 2003-2004 season up until the 2008-2009 seasons were combined in a single Excel file. After combining the data, an ordinary least squares regression was again run in Excel. This provided a more accurate picture of the past 6 seasons of data combining the statistics of each season.

With the combined years' data the *Significance F* assigned to the overall model was 1.9126 E-18. Through this extremely low value the overall model was determined to be significant. Within the model there were 5 independent variables determined to be significant. The

variables included Points per Game, Three Point Field Goal Percentage, Opponents' Field Goal Percentage, Opponents' Points per Game and Campus Support. The Points per Game variable had a P-value of 1.3460 E-05 and a coefficient of -.7673. This means that for every additional point a team averaged per game, their final ranking at the end of the year decreased by .7673. The Opponents' Field Goal Percentage variable had a P-value of 3.2311 E-06 and a coefficient of .5824. The interpretation of this variable revealed for every additional 1% on average a team allowed their opponent to shoot from the floor they increased their final ranking by .5824. Next, the Opponents' Points per Game variable had a P-value of 3.8894 E-06 and a coefficient of .5551. This revealed that for every additional point a team allowed per game on average, their ranking increased by .5551. Next, the Campus Support variable had a P-value of 1.0766 E-10 and a coefficient of -.0007. This interpretation revealed that for every 1 additional person to attend a home game on average, the team's final ranking decreased by -.0007.

The final independent variable that was significant in the model was Three Point Field Goal Percentage. This variable had a P-value of .0401 and a coefficient of .0991. This meant that for each additional 1% a team shot from behind the three point line during the year, their final ranking increased by .0991. This coefficient seems to be opposite of what it should be. It seems as though if a team increased their three point shooting percentage, they would be a better team overall and decrease their ranking rather than increasing it. The reasons for the coefficient actually being positive may be one of two things. First, multicollinearity may be present. This means that the Three Point Field Goal Percentage Variable may be highly correlated with another variable or variables. The overall model may still retain its ability to accurately represent the outcome of the final rankings, but the accuracy of the individual variable(s) may be skewed. A second possibility for the reason the variable has a positive coefficient could be found in simple logic relating to basketball. Perhaps a better shooting three point shooting team stresses three point shooting more than a team with a lower percentage. This could lead to more of an emphasis on shooting three point field goals rather than interior scoring. An emphasis on interior scoring typically leads to more free throw attempts, higher percentage shots, and more points. This is especially important when teams

are playing away from their home court where three point field goal percentage typically decreases.

The next step was to input the data into Minitab and run a stepwise regression. As mentioned, a stepwise regression eliminates any independent variables that are not significant to the model and keeps the significant variables.

After running the stepwise regression Minitab provided an output to be analyzed. As the stepwise regression got through the various steps or levels, there were 6 significant independent variables revealed. The 6 variables included Campus Support, Turnovers per Game, Points per Game, Three Point Field Goal Percentage, Opponents' Field Goal Percentage and Opponents' Points per Game. The following variables and their respective P-values are listed as follows: Campus Support, P-value less than .001; Turnovers per Game, P-value less than .001; Points per Game, P-value less than .001; Three Point Field Goal Percentage, P-value of .018; Opponents' Field Goal Percentage, P-value less than .001; and Opponents' Points per Game, P-value less than .001.

After running a stepwise regression and determining the significant independent variables, it was necessary to run a regression analysis in Minitab. This regression analysis again included all of the 6 previous years of data, but only included the 6 significant variables previously mentioned. The P-values for each independent variable remained the same as they were in the stepwise regression. The overall model had a P-value less than .001. Again this revealed that the overall model was significant. The coefficients for each of the 6 significant independent variables are listed as follows: Campus Support, coefficient of -.0006; Turnovers per Game, coefficient of 1.3082; Opponents' Field Goal Percentage, coefficient of .4656; Opponents' Points per Game, coefficient of .4282; Three Point Field Goal Percentage, coefficient of .1091; and Points per Game, coefficient of -.6919.

After viewing the output for the regression analysis it was important to interpret the meaning of each coefficient. For the Campus Support variable, each additional fan averaged during home games lowered a team's final ranking by .0006. Next, the Turnovers per Game variable revealed that each additional turnover averaged increased a team's final ranking by 1.3082.

Next, the Opponents' Field Goal Percentage displayed that when a team allowed their opponent to shoot an additional 1% from the field their ranking increased by .4656. The Opponents' Points per Game variable revealed that for each additional point a team allowed their opponent to average scoring throughout the season increased the team's ranking by .4282. Next, the Three Point Field Goal Percentage variable revealed for each additional 1% a team shot from behind the three point line, a team's ranking was increased by .1091. Finally, the Points per Game variable showed that for each additional point a team averaged throughout the year, their final ranking was decreased by .6919.

Collective Data Conclusions

After running the combined data from the 6 most recent seasons, beginning with 2003-2004 and ending with 2008-2009, there were several notable findings. First, the original ordinary least squares regression run in Excel revealed a significant overall model. It also revealed 5 significant independent variables; Points per Game, Three Point Field Goal Percentage, Opponents' Field Goal Percentage, Opponents' Points per Game and Campus Support. After reviewing this output, it was decided a stepwise regression would be beneficial.

The stepwise regression determined there were 6 significant variables. The variables included Points per Game, Three Point Field Goal Percentage, Opponents' Field Goal Percentage, Opponents' Points per Game, Campus Support and Turnovers per Game. All of the P-values were low with the highest one being Three Point Field Goal Percentage at .018. Another assuring result was that the 5 significant variables from the Excel ordinary least squares regression were all included in the stepwise regression.

Finally, a regression analysis was run in Minitab including only the significant independent variables as determined by the stepwise regression. The coefficients of these were previously listed and interpreted. Again, all of the coefficients seemed to be going in the right direction except for the Three Point Field Goal Percentage. Again, the reason that an increase in three point percentage may in fact lead to a increase in a team's final ranking may either be due to multicollinearity or an emphasis on three point shooting rather than interior scoring.

Current Year Predictions

After finding the coefficients associated with each of the 6 significant independent variables, those coefficients were assigned to each team's statistics for the current 2009-2010 season. The only variable that had to be left out was the Campus Support variable as this data was unavailable from the NCAA or individual school websites. After the coefficients were attached to each team's statistics, a ranking was determined. The team with the smallest ranking was the team that was determined by the equation to finish in first place.

As the formula was input, it was clear that there would not be rankings beginning with 1 and ending with 64. This was expected. This, however, was not a problem. The only thing that needed to be done was each output from the equation using the coefficients needed to be sorted in order. The team with the smallest output from the equation, BYU, was the team that the model determined as its champion. The bottom 32 teams were the teams that the model determined would lose in the first round.

CONCLUSIONS

In summary, a significant overall model for each season, excluding the 2005-2006 season, was developed. In addition an overall model using all of the data entries from the 6 most recent seasons, 2003-2004 to the 2008-2009 season, was also proved to be significant. Within this model 6 significant independent variables were included. The stepwise regression helped eliminate any variables that lacked significance. The model's coefficients were then applied to the current year's, 2009-2010, input data.

As expected in any March Madness, there were various upsets. The eventual champion, Duke, was ranked 3rd in the statistical model. Kansas, the 2nd ranked team in the model, lost in the second round of the tournament. For a complete listing of the model rankings refer to the attached Excel file named "rankings 09-10." As mentioned, the overall model was significant when it was run through Excel and Minitab. For many reasons, discussed below, it is very difficult to predict the actual outcome of the March Madness tournament.

Although various factors and statistics can be considered, the March Madness tournament may be one of the most difficult sporting events to predict. First the tournament is dealing

with college student-athletes, not professional athletes. The performance of collegiate student-athletes is much less predictable than the performance of a professional basketball player. Second, injuries are very difficult to bring into the equation. Purdue, which received a final ranking of 11 by the model of this study, lost their best player due to a torn ACL just prior to the beginning of the tournament. Finally, unlike most professional sports, the tournament is a single elimination tournament, hence the nickname "March Madness." The single elimination play allows for the "Cinderella" teams that seem to stand no chance on paper and in a model to advance far into the tournament.

RECOMMENDATIONS FOR FURTHER STUDY

For future study, I would first advise that the input data should not be stretched out to too many years. Although the data would include more entries which is good for running a regression, the data may become less and less relevant. The way the game is coached and played evolves over time. As a general example, the game is much faster and more athletic with less skilled jump shooters than it was 20 years ago. Although this example is over a 20 year period, there are many more subtle differences that occur even on a yearly basis. Rule changes occur, including an extension of the three point line an additional 9 inches in 2007.

For future study I would also advise a look into more non-traditional statistics. I included 2 non-traditional statistics in this study including Experience and Campus Support. Campus Support was recognized as a significant independent variable. It is likely that there are more of these measures out there to be discovered and included for a more accurate model.

APPENDICES

Appendix A- 2003-2004 Output

Regression Statistics	
Multiple R	0.679687743
R Square	0.461975428
Adjusted R Square	0.322089039
Standard Error	8.8564716
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	3367.504915	259.0388396	3.302504494	0.001142229
Residual	50	3921.85446	78.4370892		
Total	63	7289.359375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-53.96407914	41.70101927	-1.294070986	0.201586412	-137.7230397
points per game	-0.377128619	0.712878928	-0.529021976	0.599131046	-1.808988058
field goal percentage	-0.793022872	0.854268818	-0.928306002	0.357708745	-2.508872257
three point field goal percentage	0.133187223	0.101285142	1.314972963	0.194518235	-0.070249967
free throw percentage	0.491859919	0.34929584	1.408147086	0.165275193	-0.209721409
rebounds per game	0.123193706	0.529927741	0.232472649	0.817119888	-0.941197466
assists per game	1.11402378	0.894996193	1.244724603	0.219035797	-0.683628942
blocks per game	0.070608899	1.301990615	0.054231496	0.956966841	-2.544516162
steals per game	0.238060323	1.386838355	0.171656864	0.864400165	-2.547486436
turnovers per game	3.610424053	1.221657822	2.955348043	0.004753673	1.156652152
opponent's field goal percentage per game	1.838569123	0.983003152	1.870359338	0.067293271	-0.135850775
opponents points per game	-0.637686172	0.75442872	-0.845257021	0.401994114	-2.153000822
Experience	0.054899814	0.180448313	0.304241218	0.762206772	-0.307541281
Campus Support	-0.000361025	0.000242338	-1.489757052	0.142569574	-0.000847776

Appendix B- 2004-2005 Output

Regression Statistics	
Multiple R	0.584035368
R Square	0.341097311
Adjusted R Square	0.169782611
Standard Error	9.800998415
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	2486.380878	191.2600676	1.99105688	0.041412603
Residual	50	4802.978497	96.05956993		
Total	63	7289.359375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	53.08888979	49.90160691	1.063871348	0.292498065	-47.14143547
points per game	-0.426573724	0.812167151	-0.52522898	0.601745481	-2.057859423
field goal percentage	1.076892482	0.814237535	1.322577792	0.191993602	-0.558551707
three point field goal percentage	0.078383046	0.111668784	0.701924413	0.485981031	-0.145910304
free throw percentage	-1.078082487	0.483634455	-2.229126721	0.03032868	-2.049490859
rebounds per game	-0.079762717	0.667861949	-0.119429946	0.905413315	-1.421202894
assists per game	-0.778039551	1.034990578	-0.751735878	0.455734026	-2.856879266
blocks per game	-0.178526884	1.253231397	-0.142453249	0.8872945	-2.695716176
steals per game	-0.759396013	1.437308531	-0.528345861	0.59959669	-3.646315103
turnovers per game	0.558274893	1.376666427	0.405526628	0.686820555	-2.206840948
opponents field goal percentage per game	0.025259634	1.058890806	0.023854805	0.981063374	-2.101585102
opponents points per game	0.623586876	0.811755831	0.76819513	0.445984379	-1.006872663
Experience	-0.042502926	0.217935378	-0.195025361	0.846163427	-0.480239007
Campus Support	-0.000495843	0.000307707	-1.611411509	0.113385524	-0.001113892

Appendix C 2005-2006 Output

Regression Statistics	
Multiple R	0.500377896
R Square	0.250378039
Adjusted R Square	0.055476329
Standard Error	10.75195752
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	1930.629847	148.5099883	1.284637466	0.253448909
Residual	50	5780.229528	115.6045906		
Total	63	7710.859375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	109.6905674	51.49965071	2.129928376	0.038122327	6.250476764
points per game	-0.969602674	0.695023605	-1.39506438	0.169162802	-2.36559864
field goal percentage	-0.496239152	0.951175914	-0.521711226	0.604174925	-2.406732162
three point field goal percentage	0.065077221	0.169980033	0.382852147	0.703452201	-0.276337717
free throw percentage	-0.314800787	0.487353345	-0.645939522	0.521272227	-1.293678769
rebound margin per game	0.028411601	0.494208052	0.05748915	0.954384686	-0.964234465
assists per game	-0.138236084	1.185472876	-0.116608391	0.907637509	-2.519328385
blocks per game	-1.284691168	1.349014865	-0.952318022	0.345516903	-3.994267213
steals per game	-0.246299514	1.357236076	-0.181471388	0.856731258	-2.972388347
turnovers per game	-0.128765065	0.20144439	-0.639208992	0.525604736	-0.533378021
opponents field goal percentage per game	-0.89719471	1.047124092	-0.856817942	0.395633577	-3.000405305
opponents points per game	1.226918232	0.662278683	1.852570926	0.069851086	-0.103307624
Experience	-0.005300399	0.233539885	-0.022695903	0.981983178	-0.474379054
Campus Support	-0.000461006	0.000306963	-1.501828171	0.139431046	-0.001077559

Appendix D 2006-2007 Output

Regression Statistics	
Multiple R	0.640728749
R Square	0.410533329
Adjusted R Square	0.257271995
Standard Error	9.891058917
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	3406.785175	262.0603981	2.678649061	0.006251924
Residual	50	4891.652325	97.8330465		
Total	63	8298.4375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	25.21641336	48.72334639	0.517542723	0.60705965	-72.64730606
points per game	-0.504855361	0.374912714	-1.346594399	0.18418358	-1.257889693
field goal percentage	-0.513974323	0.680069411	-0.755767449	0.45333456	-1.879933908
three point field goal percentage	-0.148746551	0.20000754	-0.743704718	0.460535801	-0.550473509
free throw percentage	0.213053225	0.377303853	0.564672805	0.574821056	-0.544783853
rebound margin per game	0.900589938	0.510922114	1.762675587	0.084066733	-0.12562731
assists per game	-1.220718758	1.022423869	-1.193945872	0.238132983	-3.274317496
blocks per game	-1.011666393	0.853446183	-1.185389792	0.241466646	-2.725863465
steals per game	-0.043871082	0.897314614	-0.048891527	0.961200522	-1.846180491
turnovers per game	1.32748341	0.878933651	1.510334037	0.137252585	-0.437906749
opponents field goal percentage per game	0.486444345	0.790256573	0.615552418	0.540981889	-1.100832664
opponents points per game	0.558097084	0.452896525	1.232283872	0.223606481	-0.351572341
Experience	-0.062133022	0.224413144	-0.276868906	0.78302217	-0.512880078
Campus Support	-0.000615192	0.000265861	-2.31396414	0.024818506	-0.00114919

Appendix E 2007-2008 Output

Regression Statistics	
Multiple R	0.73619341
R Square	0.54198073
Adjusted R Square	0.41460365
Standard Error	8.13664383
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	3995.40576	307.3389	5.029086	1.43096E-05
Residual	51	3376.45361	66.20497		
Total	64	7371.85938			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-28.232824	46.2594702	-0.61031	0.544364	-121.1025831
points per game	-1.4885105	0.53319945	-2.79166	0.007362	-2.558953032
field goal percentage	0.86056716	0.73086739	1.17746	0.244478	-0.606710319
three point field goal percentage	0.21243505	0.15588562	1.362762	0.178945	-0.10051839
free throw percentage	-0.0217471	0.30193177	-0.07203	0.942862	-0.627900448
rebounds margin per game	0.88008968	0.49298402	1.78523	0.080171	-0.109617008
assists per game	-0.3645784	0.8612855	-0.4233	0.673859	-2.093681115
blocks per game	-0.4830255	1.13513115	-0.42552	0.672245	-2.761896348
steals per game	0.77082354	1.06976384	0.720555	0.474473	-1.376816928
turnovers per game	0	0	65535	#NUM!	0
opponent's points per game	1.851509	0.45134534	4.1022	0.000147	0.945395452
opponents field goal percentage	-0.2777274	0.89301278	-0.311	0.757068	-2.070525352
Experience	0.42222775	0.22674286	1.862144	0.068349	-0.032977528
Campus Support	-0.0008824	0.00021196	-4.16308	0.000121	-0.001307919

Appendix F- 2008-2009 Output

Regression Statistics	
Multiple R	0.697227556
R Square	0.486126265
Adjusted R Square	0.352519094
Standard Error	8.74269429
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	3615.374203	278.1057	3.638474	0.000467325
Residual	50	3821.735172	76.4347		
Total	63	7437.109375			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-21.98653708	47.9008616	-0.459	0.648223	-118.198247
points per game	-1.328536905	0.614385152	-2.16238	0.035398	-2.56256578
field goal percentage	0.730951869	0.816496912	0.895229	0.374954	-0.90903041
three point field goal percentage (minimum 5 made per game)	0.031668509	0.09875891	0.320665	0.749801	-0.1666946
free throw percentage	0.606010969	0.375887012	1.612216	0.11321	-0.1489803
rebound margin per game	-0.217854331	0.467792429	-0.46571	0.643447	-1.15744306
assists per game	0.157089231	0.926270023	0.169593	0.866014	-1.70337883
blocks per game	1.183426096	1.181456186	1.001667	0.321327	-1.18959844
steals per game	-0.53025531	1.378362951	-0.3847	0.702092	-3.29877872
turnovers per game	0.636513329	1.428832033	0.445478	0.657897	-2.23338021
opponent's field goal percentage per game	-0.232610737	0.957545572	-0.24292	0.809058	-2.15589758
opponents points per game	1.039206794	0.715200053	1.45303	0.152463	-0.39731476
Experience	0.136300869	0.216834231	0.628595	0.532476	-0.29922349
Campus Support	-0.000890216	0.000260897	-3.41214	0.001285	-0.00141424

Appendix G- 2009-2010 Output

Regression Statistics	
Multiple R	0.674266
R Square	0.454635
Adjusted R Square	0.339269
Standard Error	14.98528
Observations	64

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	11	9734.3861	884.9442	3.940815	0.00035278
Residual	52	11677.051	224.5587		
Total	63	21411.438			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-65.2086	84.632238	-0.77049	0.444494	-235.03562
points per game	-0.47448	1.4248629	-0.333	0.740472	-3.333678
field goal percentage	0.462193	1.3416191	0.344504	0.731857	-2.2299628
three point field goal percentage	-0.07496	0.2267791	-0.33054	0.742323	-0.5300242
free throw percentage	-0.10818	0.7254229	-0.14913	0.882029	-1.5638492
rebounds per game	-0.6706	1.2680515	-0.52884	0.599166	-3.2151286
assists per game	-1.2921	1.3555595	-0.95319	0.344904	-4.0122334
blocks per game	-6.20201	2.323675	-2.66905	0.010123	-10.864807
steals per game	2.723116	2.7936705	0.974745	0.3342	-2.8827936
turnovers per game	4.28733	2.3388457	1.833097	0.072516	-0.4059069
opponent's field goal percentage per game	1.966306	1.7576945	1.118685	0.268416	-1.5607658
opponents points per game	0.144118	1.4616666	0.098598	0.921836	-2.7889305

Appendix H- Collective Output (Excel)

Regression Statistics	
Multiple R	0.513790655
R Square	0.263980837
Adjusted R Square	0.238120704
Standard Error	9.41580056
Observations	384

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	11765.20518	905.0157832	10.20802327	1.91256E-18
Residual	370	32803.20107	88.65730018		
Total	383	44568.40625			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	12.71692769	17.34731351	0.733077643	0.463975259	-21.39476235
points per game	-0.767395456	0.173939873	-4.41184325	1.34601E-05	-1.109430149
field goal percentage	0.331952033	0.279877293	1.186062754	0.236358597	-0.218397591
three point field goal percentage	0.099153426	0.048135783	2.059869378	0.040110004	0.004499408
free throw percentage	-0.059390075	0.146266372	-0.406040529	0.684947544	-0.347007702
rebound margin per game	0.135890293	0.18244061	0.744846739	0.456837214	-0.222860214
assists per game	-0.251327907	0.348856308	-0.720433891	0.471712778	-0.937317599
blocks per game	0.289702655	0.424141611	0.683032853	0.495013535	-0.544327766
steals per game	-0.136636759	0.416222163	-0.328278432	0.742886848	-0.955094407
turnovers per game	0.058851539	0.150766397	0.390349177	0.696502995	-0.23761492
opponent's field goal percentage per game	0.58243503	0.123192554	4.727842786	3.23113E-06	0.340189668
opponents points per game	0.55510716	0.118416269	4.687760955	3.88944E-06	0.322253869
Experience	0.016162126	0.078805692	0.205088304	0.837615913	-0.138801081
Campus Support	-0.000653316	9.82931E-05	-6.646616739	1.07664E-10	-0.0008466

Appendix I- Minitab Output (Stepwise and Regression Analysis)

Results for: march_madness.MTW

Stepwise Regression: team rank versus points per g, field goal p, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is team rank on 13 predictors, with N = 384

Step	1	2	3	4
Constant	31.31	14.91	37.46	35.45
Campus Support	-0.00085	-0.00083	-0.00074	-0.00074
T-Value	-9.14	-8.95	-7.90	-7.89
P-Value	0.000	0.000	0.000	0.000
turnovers per game		1.18	1.37	1.46
T-Value		3.27	3.83	4.09
P-Value		0.001	0.000	0.000
points per game			-0.353	-0.395
T-Value			-3.71	-4.10
P-Value			0.000	0.000
three point field goal percentage				0.111
T-Value				2.37
P-Value				0.018
opp fgp				
T-Value				
P-Value				
opponents points per game				
T-Value				
P-Value				
S	9.78	9.66	9.50	9.44
R-Sq	18.05	20.31	23.12	24.25
R-Sq(adj)	17.84	19.89	22.51	23.44
Mallows Cp	47.1	37.4	24.9	21.0

Step	5	6
Constant	32.87	11.13
Campus Support	-0.00073	-0.00063
T-Value	-7.80	-6.78
P-Value	0.000	0.000
turnovers per game	1.48	1.31
T-Value	4.16	3.73
P-Value	0.000	0.000
points per game	-0.414	-0.692
T-Value	-4.27	-6.10
P-Value	0.000	0.000
three point field goal percentage	0.105	0.109
T-Value	2.23	2.38
P-Value	0.027	0.018
opp fgp	0.084	0.466
T-Value	1.60	4.65
P-Value	0.110	0.000
opponents points per game		0.428
T-Value		4.44
P-Value		0.000
S	9.42	9.19
R-Sq	24.76	28.52
R-Sq(adj)	23.76	27.38
Mallows Cp	20.4	2.9

Regression Analysis: team rank versus Campus Suppo, turnovers pe, ...

The regression equation is

team rank = 11.1 - 0.000633 Campus Support + 1.31 turnovers per game

```
            + 0.466 opp fgp + 0.428 opponents points per game
            + 0.109 three point field goal percneta - 0.692 points per game

Predictor                                Coef     SE Coef      T      P
Constant                                11.128       9.241   1.20  0.229
Campus Support                    -0.00063315  0.00009337  -6.78  0.000
turnovers per game                     1.3082      0.3509   3.73  0.000
opp fgp                                0.4656      0.1001   4.65  0.000
opponents points per game             0.42824     0.09652   4.44  0.000
three point field goal percentage      0.10907     0.04583   2.38  0.018
points per game                       -0.6919      0.1134  -6.10  0.000

S = 9.19377   R-Sq = 28.5%   R-Sq(adj) = 27.4%

Analysis of Variance
Source           DF       SS      MS      F      P
Regression        6  12616.0  2102.7  24.88  0.000
Residual Error  374  31612.5    84.5
Total           380  44228.5

Source                            DF  Seq SS
Campus Support                     1  7984.5
turnovers per game                 1   999.2
opp fgp                            1   139.7
opponents points per game          1   162.9
three point field goal percentage   1   182.2
points per game                    1  3147.6

Unusual Observations
      Campus
Obs  Support  team rank     Fit  SE Fit  Residual  St Resid
  8    13383     33.000  14.587   1.113    18.413      2.02R
 96     8060     33.000  30.159   2.270     2.841      0.32 X
 98     2293     34.000  22.301   2.173    11.699      1.31 X
122     9838      9.000  27.234   1.383   -18.234     -2.01R
164    17530     34.000  15.254   1.026    18.746      2.05R
167     8421      5.000  23.375   0.633   -18.375     -2.00R
171     7743      9.000  30.869   1.206   -21.869     -2.40R
176    10010     33.000  35.361   3.131    -2.361     -0.27 X
179     5798      9.000  33.027   1.172   -24.027     -2.63R
205    10851      1.000  19.929   0.970   -18.929     -2.07R
206     4533      2.000  24.503   1.031   -22.503     -2.46R
278    18746     34.000  11.144   1.442    22.856      2.52R
301    22978     33.000   7.782   2.018    25.218      2.81R
315     8510      5.000  24.828   1.147   -19.828     -2.17R
319     9683      5.000  24.854   0.699   -19.854     -2.17R
342    19443     33.000  13.665   1.200    19.335      2.12R
366     1770     33.000  35.091   2.329    -2.091     -0.24 X
R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large leverage.
```

REFERENCES

Deddens, J.A., and K. Steenland. "Effect of Travel and Rest on Performance of Professional Basketball Players." *National Institute for Occupational Safety and Health* 20.5 (1997): 366-369. Web. 28 Apr 2009. <http://cat.inist.fr/?aModele=afficheN&cpsidt=2752126>.

Dirks, Kurt T. "Trust in Leadership and Team Performance: Evidence from NCAA Basketball." *Journal of Applied Psychology* 85.6 (2000): 1004-1012. Web. 17 Sep 2009. <http://web.ebscohost.com/ehost/detail?vid=1&hid=107&sid=27c0524f-386c-4152-b01e-410f816f020f%40sessionmgr111&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#fn5#db=pdh&AN=apl-85-6-1004>.

Esters, Irvin G., and Richard F. Uttenbach. "Utility of Team Indices for Predicting End of Season Ranking in Two National Polls." *Journal of Sport Behavior* 18.3 (1995): 216-224. Web. 10 Nov 2009. <http://web.ebscohost.com/ehost/detail?vid=1&hid=111&sid=b4852609-bbd5-4fcd-96fd-b0a84449e7b5%40sessionmgr114&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=aph&AN=9510063904#db=aph&AN=9510063904>.

Holbrook, Brett C. , Kathryn L. Schenk, and Neil C. Schwertman. "More Probability Models for the NCAA Regional Basketball Tournaments." *American Statistician* 50.1 (1996): 34-38. Web. 17 Sep 2009. <http://web.ebscohost.com/ehost/detail?vid=1&hid=8&sid=3a978f11-d307-4c86-bff4-fbd97fbd4e17%40sessionmgr13&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=aph&AN=9604164615#db=aph&AN=9604164615>.

"Men's Basketball Statistics." *National Collegiate Athletic Association.* National Collegiate Athletic Association, 14 Mar 2010. Web. 8 Mar 2010. <http://www.ncaa.org/wps/portal/ncaahome?WCM_GLOBAL_CONTEXT=/ncaa/ncaa/sports+and+championship/general+information/stats/m+basketball/index.html>.

Onwuegbuzie, Anthony J. "Factors Associated with Success Among NBA Teams." *Sport Journal* 3.2 (2000): web. Web. 12 Oct 2009. <http://www.thesportjournal.org/article/factors-associated-success-among-nba-teams>.

Schwertman, Neil C., and Thomas A. McCready. "Probability Models for the NCAA Regional Basketball Tournaments." *American Statistician* 45.1 (1991): 35-39. Web. 25 April 2009. <http://web.ebscohost.com/ehost/detail?vid=1&hid=103&sid=2da5b55c-906f-415c-8d7a-f8e1ff275ce5%40sessionmgr111&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#db=aph&AN=9605213308>.