Association of Statistics & Success in the Big East

Abstract

This study aims to discover which statistics are associated with being a good college basketball team in the Big East. Data from Sports Reference was collected, and then we apply a backward stepwise regression model to determine variables that significantly impact rank. It turns out that conference wins, field goal percentage, and points are the most important factors when deciding which team is going to be successful in the Big East.

Introduction

Background and Significance

Historically, college basketball has different conferences to divide up teams into groups. Teams are selected into these conferences depending on the location of the schools. These conferences are put in to have a structural schedule, which gives them about half of their schedule, for each team to play for the upcoming year. During conference play the teams are ranked not only through the entire NCAA, but the conference as well. As college basketball fans and students at Xavier University, our focus is on the Big East. We are looking to investigate who is the best team in the Big East.

Explanation of the data set

The dataset contains yearly (1979 - 2021) College Basketball Big East Conference data for every active Big East season. The data was collected from Sports Reference then cleaned for data analysis.

Question

Which statistics are associated with being a good college basketball team in the Big East conference?

Methodology

First, we ran the model using all variables besides school, ID, and year. Then, we removed the variable with the highest p-value. We continued to run the model ~removing one variable at a time~ until all variables remaining had a p-value less than 0.05. From running a backward stepwise regression model, we found that conference wins, away losses, offensive rebounds, total rebounds, and personal fouls have statistical significance to rank. Then, we ran a backward stepwise regression model using rank as the y-variable for the new Big East (2013-2021), the Big East football-playing schools (2005-2012), and the old Big East (1980-2004). From these

models, we will compare the results to attempt to find out which variables are associated with being a good college basketball team in the Big East.

The Big East Conference was formed in 1979; however, the conference split in 2013. From 2013 to now, the Big East conference consists of the non-football playing schools. These schools include Butler, Creighton, DePaul, Georgetown, Marquette, Providence, St. John's, Seton Hall, Villanova and Xavier. Therefore, we created three linear regression models for each era of the Big East Conference:

- the new Big East (2013-2021),
- the Big East football-playing schools (2005-2012), and
- the old Big East (1980-2004).

Using a backward stepwise approach, we created three linear regression models for each era of the Big East (the old Big East, the new Big East, and the Big East football-playing schools) with rank being the dependent variable. Within each model, we ran the regression using all variables in the data set, excluding school, ID, and year. For a variable to be considered significant, it needs to have a p-value less than 0.05. Upon running the initial model, many variables were considered to be insignificant. Therefore, we removed the variable with the highest p-value first, and then we continued to run the regression (removing one variable at a time based on the variable with the maximum p-value) until each variable in the model had a p-value less than 0.05.

Results

To make our final conclusion, we will compare the results to attempt to find out which variables are associated with being a good college basketball team in the Big East. The significant variables in the old Big East model include conference wins, conference losses, field goal attempts, field goal percentage, free throws, total rebounds, and points. See attached. Conference wins, field goal attempts, field goal percentage, and free throws have negative slopes; therefore, an increase in either conference wins, field goal attempts, free throws, and/or field goal percentage decreases rank. The rest have positive slopes; so, an increase in either conference losses, total rebounds, and/or points increases rank. The significant variables in the "football"

Big East model include conference wins, away losses, field goal percentage, 3-pointer attempts, and points. See attached. Conference wins, field goal percentage, and 3-pointer attempts have negative slopes; therefore, an increase in either conference wins, field goal percentage, and/or 3-pointer attempts decreases rank. The rest have positive slopes; so, an increase in away losses and points increases rank. The significant variables in the new Big East model include: conference wins, field goal percentage, X3_pointer_attempts, free-throw attempts, offensive rating, free-throw percentage, offensive rebounds, turnovers, and points. See attached. Conference wins, offensive ratings, and turnovers have negative slopes; so, an increase in either conference wins, offensive ratings, and/or turnovers decreases rank. All else have positive slopes; therefore, an increase in field goal percentage, X3_pointer_attempts, free-throw attempts, free-throw percentage, offensive rebounds, and points increase rank. For all three models, conference wins, field goal percentage, and points are significant variables. This shows that conference wins, field goal percentage, and points are the most important factors when deciding who is a good college basketball team in the Big East.

Conclusion

There are a lot of statistics that are kept for a game of basketball. Statistics are a good way to tell how you are doing in a sport and what you are best at. They can also help to determine where you and your team need to improve. Of course scoring is one of the most important statistics in the game. However, times change. Long gone are the days when a player was judged simply on points, rebounds, assists, blocks, steals and field-goal percentage. For example, our dataset had missing data for a lot of statistics in the old Big East; therefore, various statistics were omitted from our backward stepwise approach when computing our regression model in the old Big East. Possible future work includes running a forward stepwise approach on our dataset in order to see if it matches our backward stepwise approach. Also, it would be helpful to include data from other NCAA basketball conferences (Big 10, ACC, SEC, etc.) to see if the models match from all the conferences. This would expand our original question to: which statistics are associated with being a good college basketball team in the NCAA. In addition to a model with all NCAA basketball conferences, we could compare conference to conference in order to see which statistics are associated with a highly ranked team.

References

Data taken from https://www.kaggle.com/datasets/mattop/college-basketball-big-east-1979-2021

Gaydos, Ryan. (2021). *Big East Conference men's basketball championship history*. Fox News, https://www.foxnews.com/sports/big-east-conference-mens-basketball-championship-history.