Jason Liebmann and Mingxian Gao (Jimmy)

December 15th, 2017

CIS 105 Final Project Write Up

**Methodology:**

The main goal of this project is to explore the boxscores on the 2016-2017 basketball season to statistically validate whether or not there is home field advantage, as well as what teams have statistically significant home field advantages. In order to do this, we tested three main hypotheses:

* The first tested whether the home spread and away spread for the entire NBA come from the same underlying population or not
* The second tested if each of the specific team difference in home and away spread was statistically significant from the league average
* The third tested whether the home and away spread for any given team came from the same underlying distribution or different distributions

In addition to testing three hypotheses, we looked at other statistics that may contribute to the difference between the Pacers’ home and away spread since they had the greatest difference of any team. Finally, two models for spread was built upon factors that predict this statistic after running the correlation between the spread and variables in the dataset as well as plotting a scatterplot of the statistics in the dataset to see if there is significant linear correlation. We built a model for Indiana Pacers because it has the highest difference between the home and away spread and also constructed a model for the league in general.

**Findings:**

In terms of the hypotheses, we found that the home spread came from an underlying population with a higher mean than the population distribution where the away spread is generated from for the NBA as a whole. We also found that none of the home field advantages for a specific team were statistically significant from the average difference in home and away spread for the league. Finally, we found that 22 out of 30 teams had home spreads that came from a greater mean of the population distribution than the distribution where the away spread is drawn from.

We also found that field goals made, turnovers, and blocks are three other statistics that may contribute to the difference between the home and away spread of Indiana Pacers. In terms of the model, we found that there are four statistics that significantly predict the spread of Indiana Pacers: defensive rebound of the team, field goal made of the opponent team, assist of the opponent team and the field goal percentage of the team. When we were building a model for the league, in addition to the four factors that were mentioned, we found that two extra statistics that predict the spread: the defensive rebound of the opponent team and the field percentage of the opponent team.