Matthew Blackley

Dr. Fellingham, Dr. Page, Dr. Fisher

August 2023 – April 2024

Background:

It is commonly believed that certain players perform better in the playoffs rather than the regular season and vice versa. For example, Michael Jordan and Lebron James are commonly considered playoff performers. Steph Curry and Karl Malone are considered regular season performers. We wanted to confirm this belief statistically.

Goal:

The goal of this research is to confirm a common belief that certain players in the NBA are playoff or regular season performers.

Previous Work:

A previous master's student did a lot of the heavy lifting when it came to the code for web scraping. They created a variable called "adjusted VORP" and coded to calculate and scrape it from Basketball Reference.

My Work:

I edited the scraping code to split the data into regular and post season while also changing it to be in according to the new scraping policies that were put into effect by the website. I did this on the statistics server as it took a lot of time. Additionally, I chose to only include certain teams from each decade as it was going to take way too long to scrape all the seasons and every team. After this, I chose specific players to look into: Michael Jordan, Scottie Pippin, Stephen Curry, Lebron James, Karl Malone, Kobe Bryant, and John Stockton. There are a few others, but these were the major ones. After this, team adjusted VORP was accounted for by taking the mean of the rest of the team and subtracting that to the specific players adjusted VORP for a specific year. This was done to regular and post season. After this, a plot was made to see if there was a difference. However, the results seemed insignificant, so we ran a two-sample t-test to see what would happen and most of the results came back as significant. Because of this, we decided that

this is a classic problem of small sample size and we just needed to do more work to conclude any real results.

Future Work:

In the future, we want to scrape over a player's career rather than a few specific seasons. Additionally, we want to use possibly a season average rather than a team average because VORP is supposed to already account for the team's performance. Lastly, we wanted to possible use another metric, or create our own, to measure performance to see if the new metrics work better than adjusted VORP.

Code Locations:

- Pictures: This is where I kept most of the graphs that I used for my presentation and weekly update
- John: This is where I did most of the work for specific players. This is where you can find the code to scrape for one specific player, clean the data, and do the same analyses that I did.
- Linear: This is where I tried doing some linear regression, but we switched paths.
- Merging: This is where I was merging data files.
- Old_data: This is a lot of old data that I ended up not using.
- Old_stuff: This is a lot of R files that I created trying to find the correct way to web scrape, but I ended up using the stuff in John and server.
- Server: This is a copied folder of all the code and data that I had on the stats server to do all of the web scraping.
- Stats: This is where all of my models and graphing is done. There is two R files.