# Sean McLean

# ALY 6010

# Final Project –

# Milestone 2

**Questions to consider**

Questions to explore from the initial data set that pertains to the college basketball rankings from the 2020-2021 season.

1. How different does the final poll done from the Associated Press of the 2020-2021 season look from the first poll that was created? Is it the same teams in the top 25 or is it different?

2. From the 15 teams that were ranked in both the first poll and final ranking, did the means of the team's rankings between the two polls change from the first to final polls?

**Hypothesis**

Hypotheses established based off the questions:

Null: There was no mean difference between the first poll and final poll using a two-sample test for the 15 schools ranked in both.

Alternate: There was a mean difference between the first poll and final poll using a two-sample test for the 15 schools ranked in both.

**Findings**

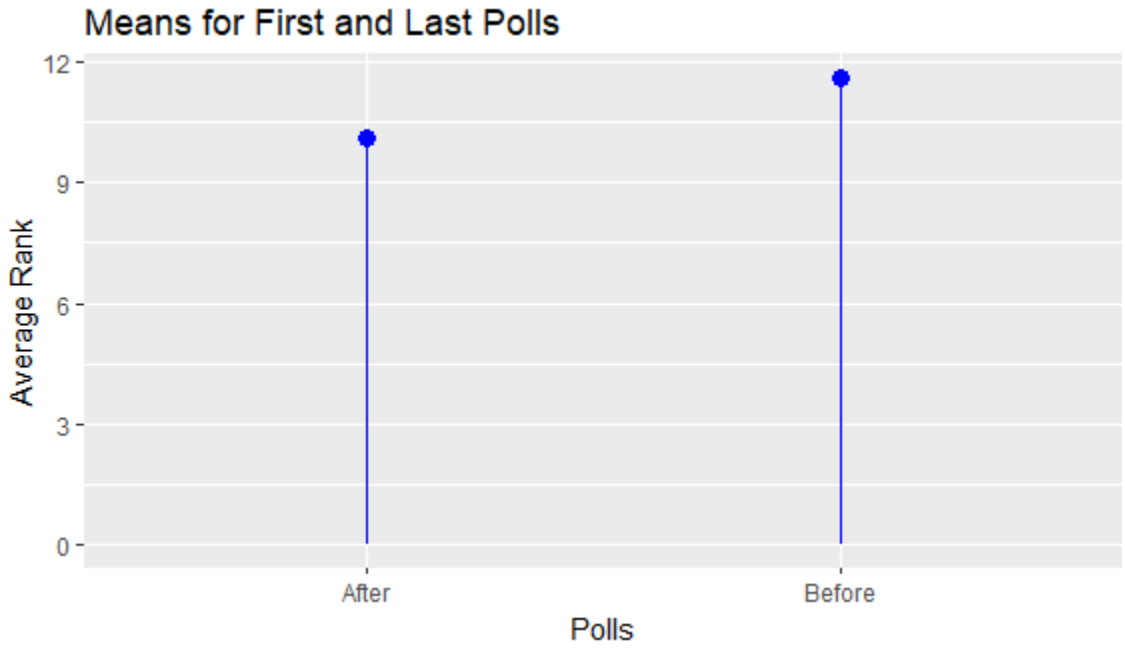
Research and analysis were done using the data from the first part of the final milestone project and how the questions were formulated. The data for the first and final polls conducted by the Associated Press was extracted to locate the top 25 rankings and see what schools were ranked in both. The 15 schools that were ranked all season or for a portion of the season are Gonzaga, Baylor, Villanova, Virginia, Iowa, Kansas, Illinois, Crieghton, Texas Tech, West Virginia, Houston, Texas, Florida State, Ohio State, and Michigan. The school's rankings from both polls were then put into new variables that were used for t-tests.

With the usage of several t-tests that used one sample and two sample testing and also at different significance levels, the tests consistently showed high p-values and low t-statistics in each test conducted. The mean ranking of the 15 schools in the final poll was 10.13, about 1.5 points higher than the mean ranking of those schools in the first poll of the season. While this shows improvement as the season goes on, the high p-value suggests that this small difference is not enough to reject to the null hypothesis because there was nearly no difference in the means. Traditionally the preseason poll is not a reflection of how the final poll will look, with several factors like injuries and other schools that are better than anticipated will alter the polls over the

course of a season.

**Conclusion**

I expected that there wouldn't be too much of a difference in the means because there are usually about 10 to 15 schools that are in the polls all season, and that consistency in the top 25 rankings gave me the impression that the means could be pretty similar. The closeness of the means is shown visually in the dot plot shown in the report on the next page. I conclude from the testing done that because the means are very close that there is not enough sufficient evidence to reject the null hypothesis.




References:

Chat GPT. (2023, December 2nd). Default (GPT 3.5).

Masseycre. (2021, November 13th). College Football/Basketball/Baseball Rankings. Kaggle. Retrieved on November 11, 2023 from https://www.kaggle.com/datasets/masseyratings/rankings?select=cb2021.csv