

On Performance Metrics of Seeded and Non-Seeded March Madness Teams

STATS 406

Dinko Imsirovic

Introduction

One of the most exciting and quintessential American sporting events is the NCAA Men's National Basketball Championship, dubbed March Madness, for its consistently unpredictable outcomes. The success of a team's season is often defined not by number of games won or conference success, but performance in March Madness. March Madness features 60 seeded teams along with 8 teams vying for 4 play-in spots, designated as the *First Four*¹. Finding a way to maximize your potential for being seeded can be the difference between staying home or cementing your school in collegiate athletics lore, such as 16 seeded UMBC upsetting 1 seeded Virginia - the first 1-16 upset in the Men's NCAA Tournament². One method of doing so is comparing the performance metrics of teams that earned a seed against those that haven't. The focus of this analysis is to determine if the distributions of certain metrics, namely 3 point field goal percentage and opponents' assist rate, for seeded teams are in fact different than the distributions of those for non-seeded teams.

One motivation for this analysis can be to serve a coaching staff for a team that historically rides a thin line between qualifying and not (a bubble team). If the distribution of a given metric can be shown to be different for seeded versus non-seeded teams, then that statistic can be focused on in their scheme for the season. For example, a coaching staff for a team that just missed qualifying is meeting to discuss what they need to change this forthcoming season. They have ideas and data from last season but it takes time and is difficult to quantify if those metrics could make a discernible difference. Results from this analysis can more concretely guide staff on improving aspects of their scheme and performance. The staff could recommend increasing tempo (i.e. playing with a faster pace) or improving their 3 point field goal percentage and analyzing the distributions of those statistics in seeded and non-seeded teams can reveal if seeded teams do have improved performance in those areas, on average.

This analysis is divided into *Methods, Simulations, Analysis, and Discussion*. Methods will describe the methodologies by which we will analyze these distributions and simulations will justify those choices using generated data, demonstrating how the methods work when assumptions do and do not hold. Analysis then applies those methodologies to the actual data and provides interpretations for those results and Discussion summarizes the results and discusses implications of the study and potential improvements.

Data

The data used in this analysis is gathered from kenpom.com³, a statistical archive of basketball data whose accuracy of ranking's in predicting game outcomes has been noted by popular newspapers and blogs such as FiveThirtyEight and the Wall Street Journal. Two sets of 10 .csv files for the 2010-2019 seasons were collected and merged into a final data set: First, a collection of files recording pre-March Madness tempo and offensive/defensive efficiency statistics and the resulting seed (if applicable) were concatenated into a single data set. Additionally, a collection of other miscellaneous summary statistics such as 2 point field goal percentage, opposing team 3 point field goal percentage, etc., were concatenated into another data set. It is not explicitly stated if this second data set includes March Madness games or not but it does seem like a reasonable assumption to make. These two data sets were then merged by team name and season to form the final data set, containing 3492 observations of 46 variables. Data sources and cleaning scripts are deferred to supplemental materials.

The distribution of seed counts is wise to examine to ensure there are no glaring errors with the data source. This data is displayed in the following table:

Seed_value	Total
1	40
2	40
3	40
4	40
5	40
6	40
7	40
8	40
9	40
10	40
11	53
12	43
13	41
14	41
15	40
16	59
NA	2815

Considering there are 10 years worth of tournament data, with 4 of each seed in each year (4 regions each containing their own independent set of seeds 1-16), an initial assumption was made that there would be 40 listings of each seed. This is not the case, however. This discrepancy is due to the First Four listed as the seeds they are competing for, regardless if they won their qualifying math. For example, in 2018 there were 4 teams vying for 2 16 seeds and 4 teams vying for 2 11 seeds but in 2013, there were 4 teams vying for 2 16

seeds, 2 teams vying for 1 11 seed, and 2 teams vying for 1 13 seed. These inconsistencies are represented in the data, and are left in the data since there is no way to automatically distinguish which teams actually won their First Four game, and removing all instances of at-large seeds would diminish the quality of analysis.

The variables of focus in this analysis are 3 point field goal percentage and opponents' assist rate. For clarity, 3 point field goal percentage is defined as the ratio of made 3 point shots to attempted 3 point shots and opponents' assist rate is the percentage of baskets made by the opponent that were assisted, in other words, baskets that were made after a successful pass in that same possession. These variables were selected with 3 goals in mind: To analyze

- One variable that is offensive and another that is defensive in nature
- Variables that are simple to interpret and begin addressing, and
- Variables that have significant correlation with seeding

We anticipate that the distribution of 3 point field goal percentage for seeded teams is right-shifted and opponents' assist rate is left-shifted when compared to non-seeded teams. In other words, we anticipate in general, seeded teams have a higher 3 point field goal percentage and lower opponents' assist rate compared to non-seeded teams. The distributions of these two variables are shown below:

Opponents' Assist Rate of Mens DI Basketball Teams

2010 – 2019 Seasons

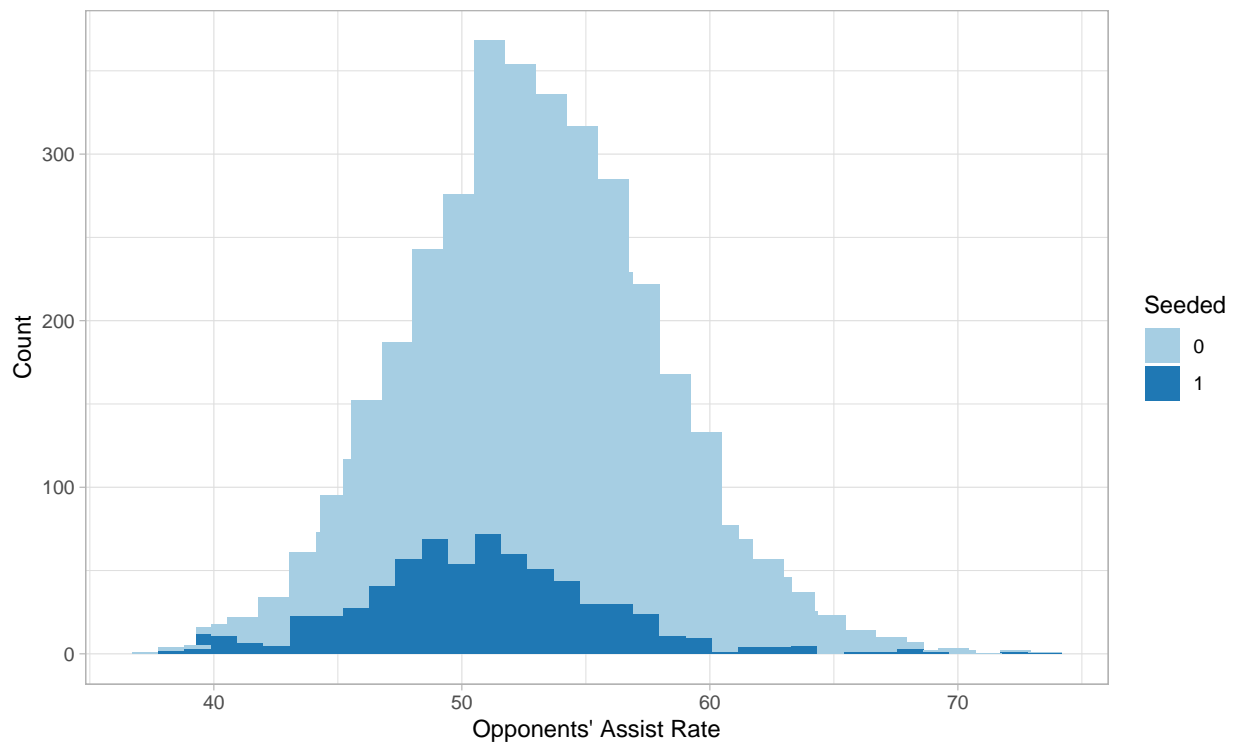


Figure 1

3 Point FG Percentage of Mens DI Basketball Teams 2010 – 2019 Seasons

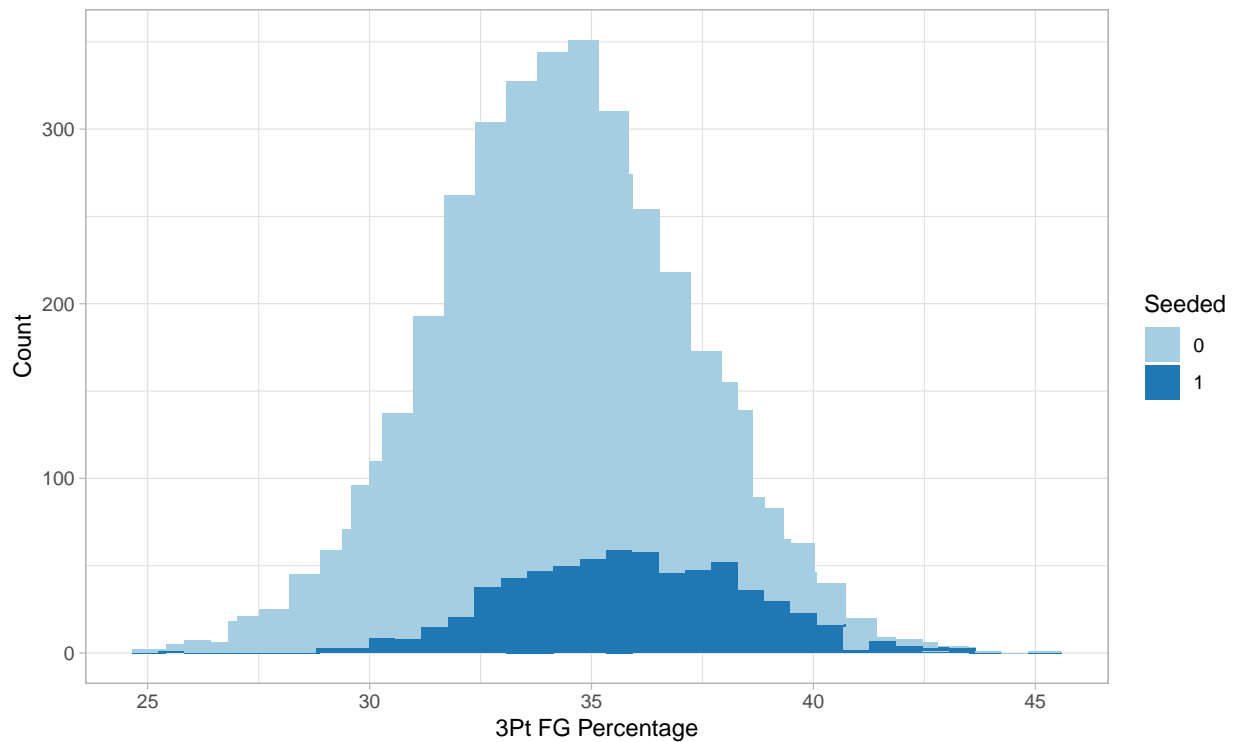


Figure 2

We can see that the distribution of both of these metrics are approximately normal and when colored by seeded versus non-seeded, there does seem to be a shift in distributions according to our anticipation. This matches intuition: if an opposing team has a lower assist rate, then pass opportunities are being minimized and the defense is more effective. More effective defenses are naturally common in the national tournament. This also holds for higher 3 point field goal percentage: tournament teams are likely to be better 3 point shooters than non-tournament teams.

Methods

This analysis features three key methods discussed below: Variable selection, 2-sample permutation tests for difference of means, and stratified bootstrap confidence intervals for estimates of these differences.

Variable Selection

Variables were initially selected heuristically from a set that seemed more offensive or defensive in nature. The Kendall Rank Correlation test was then performed to calculate the correlation between those variables and seeded/non-seeded. This method was chosen over the typical Pearson Correlation test because that is typically used as a measure of correlation between numerical variables whereas one variable in this test is categorical and Kendall's Tau is used to measure ordinal association between two measurements. We can consider seeded/non-seeded to be ordinal since it is certainly more desirable to be seeded than not. Kendall's Tau is a non-parametric test and does not rely on any assumptions on the joint distribution of the variables in question nor their marginal distributions.

Permutation Test

To test if the distributions of 3 point field goal percentage and opponents' assist rate are in fact different for seeded versus non-seeded teams, we perform a 2-sample permutation test to test if the difference of means is significant. Our null hypothesis is $H_0 : F = G$ and our alternate hypothesis is $H_a : F \neq G$ and our test statistic can be calculated as $T(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^n Z_i w_i - \frac{1}{n} \sum_{i=1}^n (1 - Z_i) w_i$. To create a null distribution for our test statistic, we shuffle the labels, seeded versus non-seeded, of our data, compute the difference of means for the two groups in this shuffling, and replicate this many times. Then we compute our observed test statistic and corresponding p-value. The general intuition behind this test is that under our null distribution, the labels, Z , are non informative, so if we permute Z , there should be no difference in our data. This is also known as Welch's permutational t-test.

Stratified Bootstrapped Confidence Interval

To estimate the true difference in means for our two distributions, we will utilize a 95% stratified bootstrapped confidence interval. This type of bootstrapped confidence interval is useful when the data is independent but not identically distributed. We make an assumption of independence and the goal of our permutation test is to show that our data is not identically distributed, therefore a stratified bootstrap is appropriate. Bootstrap samples are created with replacement for both groups and the difference of mean is computed. This is then replicated many times to create our basic bootstrapped confidence interval.

Simulations

To further justify our methods, we will simulate different scenarios and analyze the operating characteristics of these methods. Namely, we will look at Type I Error and Power for the 2-sample permutational test. By construction, permutation tests have size no greater than level. In other words, they will not reject the null hypothesis more than $100 \times \alpha\%$ of the time. However, the power of our test can be poor if we select poor test statistics. By demonstrating our permutation test has high power and size no greater than level when simulated data is similar to the data, we can then properly justify our use of the 2-sample permutation test.

Size

To estimate the size of our permutation test, we first generate a distribution of null test statistics that satisfies the null hypothesis. We then repeatedly generate data that also does satisfy our null hypothesis and compute the proportion of time we incorrectly reject the null hypothesis. To simulate data similar to the actual data, we generate 3500 random variables $\propto N(50, 5)$ and 3500 labels that are split approximately 80/20. This is because the distribution of opponents' assist rate is has an approximately $N(50, 5)$ distribution and the distribution of non-seeded to seeded teams is approximately 80/20. We then permute those labels and compute the difference of means for the labels, replicated 5000 times, serving as our null distribution for the test statistic. To then test size, we repeatedly generate a distribution of test statistics identical to that of our null distribution and gather the proportion of time we incorrectly reject the null hypothesis at $\alpha = 0.05$ out of 1000 replications. We then compute a 95% confidence interval for the estimated size of our test, shown below:

```
## [1] 0.04609504 0.07656049
```

The 95% confidence interval for the estimated size of our permutation test is (0.0461, 0.0766). Because our confidence interval contains 0, we can reject the null hypothesis that size is greater than level. In other words, if the null hypothesis is true, then our method will incorrectly reject less than 5% of the time. This specifically addresses opponents' assist rate and not 3 point field goal percentage but we would like to paint broad strokes with our simulations in the interest of brevity of analysis.

Power

To investigate power, we generate samples of data representative of 3 point field goal percentage ($\propto N(37, 10)$) where one subset has an induced mean shift and the other does not. Specifically, we generate 80% of our data with a shift, simulating seeded teams that do in fact have higher mean 3 point field goal percentage and the remaining 20% are untouched. We then calculate what proportion of the time we reject the null hypothesis when it is in fact not true with differing shift values and gather estimates of power in those respective situations. Our corresponding power curve is shown below:

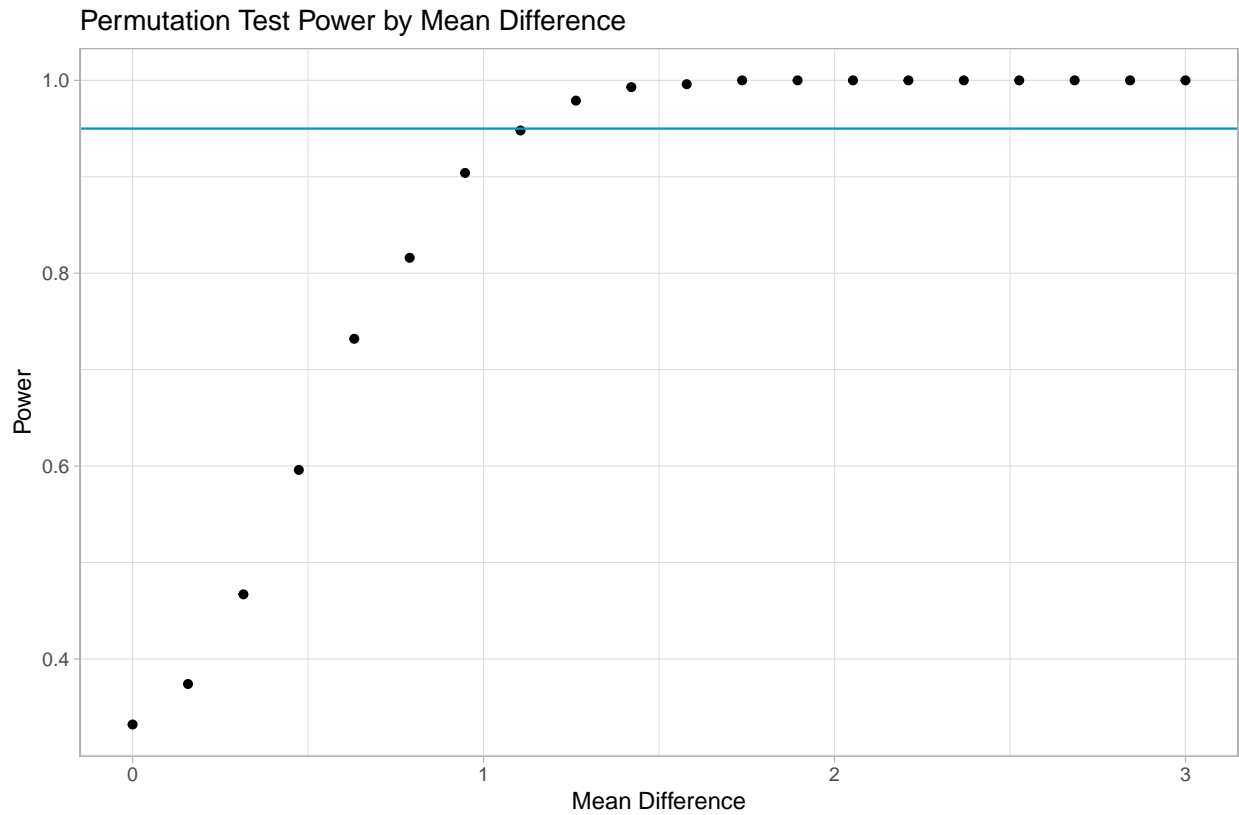


Figure 3

We can see from the graph above that a mean difference of approximately 1.2 will ensure a power of at least 95%. In other words, if our observed mean difference in 3 point field goal percentage is at least 1.2, then the probability of rejecting the null hypothesis when it is not true is at least 95%.

Analysis

We first evaluate the statistical significance of the correlation between seeding value and opponents' assist rate and 3 point field goal percentage using Kendall's Tau.

```
##
## Kendall's rank correlation tau
##
## data:  ncaa_data$Seeded and ncaa_data$OppARate
## z = -11.966, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## -0.1653798

##
## Kendall's rank correlation tau
##
## data:  ncaa_data$Seeded and ncaa_data$FG3Pct
## z = 14.995, p-value < 2.2e-16
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
##      tau
## 0.2072549
```

Our test statistic values for opponents' assist rate and 3 point field goal percentage are -11.966 and 14.995, with p-values of approximately 0 for both. This shows us that these metrics are statistically dependent and is promising to show that the distributions are different for seeded and non-seeded teams.

We now perform our 2-sample permutation test for difference of means for seeded versus non-seeded teams. We can perform this test with confidence of accuracy due to the results of its operating characteristics demonstrated in *Simulations*. Our null distribution of test statistics was created by calculating mean differences of permuted labels for each metric 10,000 times.

Null Distribution of Mean Difference Test Statistics
3 Point Field Goal Percentage

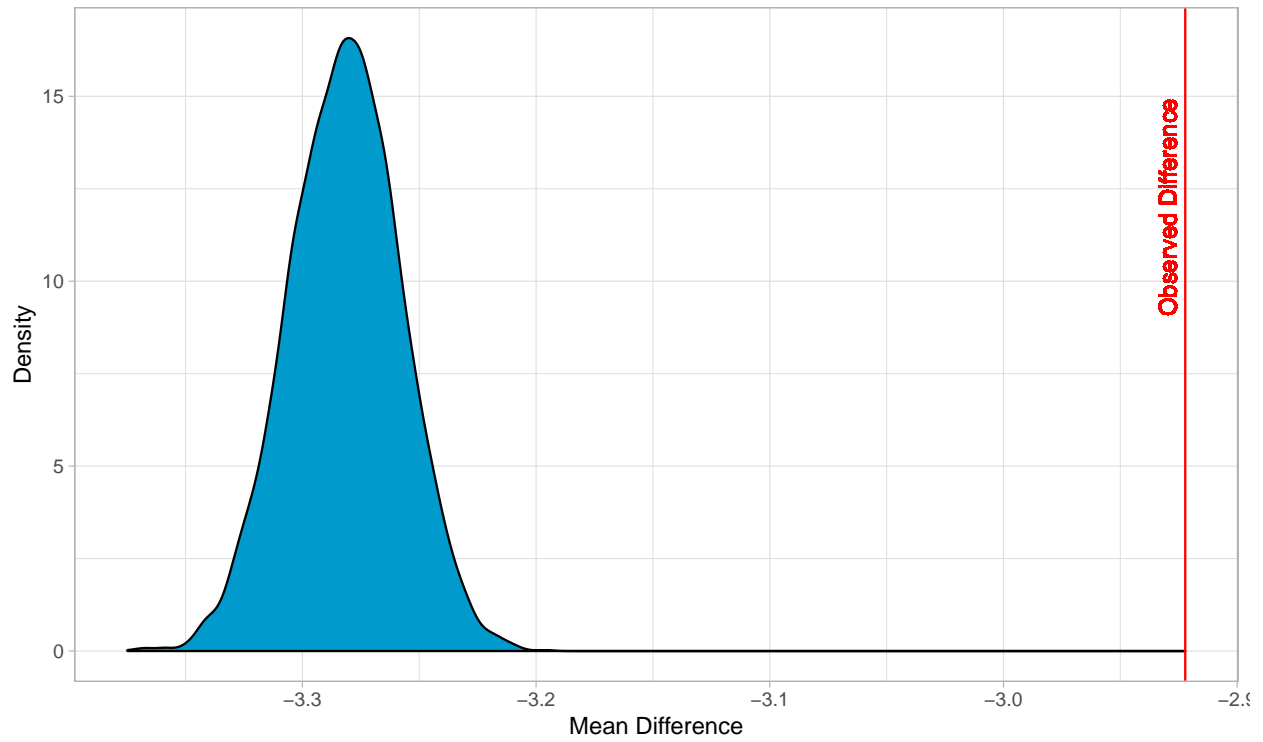


Figure 4

Null Distribution of Mean Difference Test Statistics
Opponents' Assist Rate

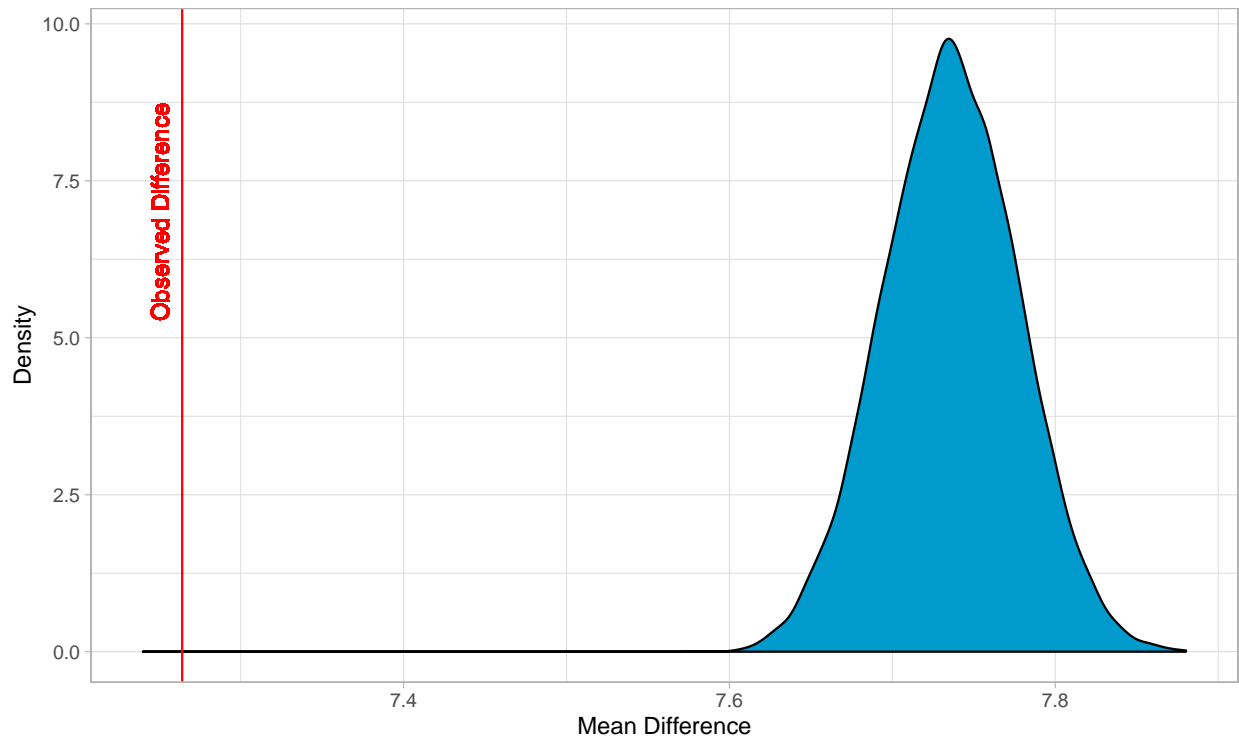


Figure 5

```
## [1] 0
```

```
## [1] 0
```

The graphs above show the distributions of null test statistics and our observed difference. We can clearly see that the observed difference is a significantly extreme value in both instances - the mean 3 point field goal percentage for non-seeded teams is significantly lower than non-seeded teams and opponents' assist rate is significantly lower for seeded teams than non-seeded teams. This is confirmed with calculated p-values of near 0 for variables. In other words, if the distribution of mean differences of opponents' assist rate and field goal percentage for seeded and non seeded teams matched our null distributions, then the probability of observed a difference as extreme as ours is approximately 0% in both cases.

Because our data is shown to come from two different distributions, we can calculate our 95% basic stratified bootstrapped confidence interval with 1000 replications for the estimated difference of means for both 3 point field goal percentage and opponents' assist rate.

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = three_pt_ci, type = "basic")
##
## Intervals :
## Level      Basic
## 95%      (-11.526, -0.253 )
## Calculations and Intervals on Original Scale

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = opp_assist_ci, type = "basic")
##
## Intervals :
## Level      Basic
## 95%      ( 4.939, 24.881 )
## Calculations and Intervals on Original Scale
```

These results show that with 95% confidence, we estimate the true difference of means for 3 point field goal percentage and opponents' assist rate for seeded and non-seeded teams to be within (-11.526, -0.253) and (4.939, 24.881), respectively. Because neither of these intervals contain 0, we can confidently conclude that the mean difference of 3 point field goal percentage and opponents' assist rate for seeded and non-seeded teams is not 0 and therefore the distributions for these metrics are indeed different for seeded and non-seeded teams.

Discussion

Our analysis was focused on determining if the distributions of 3 point field goal percentage and opponents' assist rate were indeed different for seeded versus non-seeded teams in March Madness, using data from the 2010-2019 NCAA basketball regular seasons. We first described our methodology of 2-sample permutation tests with a difference of mean test statistic and a stratified bootstrapped confidence interval for the estimate of the true difference of mean for the two metrics between seeded and non-seeded teams. Within our *Simulations* section, we analyzed the operating characteristics of our permutation test using simulated data similar to our actual data and demonstrated that it does indeed have size no greater than level. In other words, that the probability of rejecting our null hypothesis when it is in fact true is less than $\alpha = 5\%$. We also demonstrated that if the observed difference in mean is at least 1.2, then we achieve a power of at least 95%. In other words, if the observed difference in mean is at least 1.2, then the probability of rejecting our null hypothesis when our alternative hypothesis is true is at least 95%. These results subsequently gave us confidence in the accuracy of our methodology, justifying our choice of methods. Our analysis then showcased an extreme value for our observed test statistic in both variables and a corresponding p-value of approximately 0. This allowed us to conclude that the distributions of these variables are indeed different for seeded versus non-seeded teams. We then provided a 95% stratified bootstrapped confidence interval for the estimate of true difference in mean in both metrics for seeded versus non-seeded teams. Neither of these confidence intervals included 0, further cementing our confidence in our conclusion.

The implications of this analysis can be quite useful for collegiate basketball coaching staffs focused on qualifying for March Madness. Staff members may not be interested specifically in 3 point field goal percentage or opponents' assist rate, as it may seem nearly trivial that teams who qualify for the tourney are better 3 point shooters than those who didn't. However, the general concept of this analysis provides a framework for analyzing other potentially more specific, focused metrics that coaching staffs may wonder if they are indeed different for seeded versus non-seeded teams.

This could be further improved by a more rigorous simulations section that generates data similar to both variables, and not just one variable per simulation, in tandem with stronger tests for independence before performing the stratified bootstrapped confidence interval. Additional machine learning methodologies could also be applied to estimate the true difference in means for these metrics, ranging from supervised regression techniques or hierarchical Bayesian models to advanced unsupervised methods.

References

1. NCAA.com. “How the Field of 68 Teams Is Picked for March Madness.” *NCAA.com*, NCAA.com, 17 Aug. 2020, www.ncaa.com/news/basketball-men/article/2020-08-17/how-field-68-teams-picked-march-madness.
2. Wolken, D. “UMBC Stuns Virginia To Make NCAA Tournament History As First No. 16 Seed Beat No. 1 Seed.”, *Usatoday.com*, 16 March 2018 www.usatoday.com/story/sports/ncaab/16-seed-umbc-stuns-virginia-make-ncaa-tournament-history/434445002.
3. Pomeroy, Ken. “Efficiency and Tempo, Miscellaneous Team Stats.” *kenpom.com* 2020, www.kenpom.com/summary.php.