Each year, the National Basketball Association (NBA) holds a draft, where prospective basketball players are able to be chosen to join one of the 30 professional teams across the United States and Canada.

In order to be eligible for the draft, a player must be at least 19 years old and out of high school for at least one year. Prior to 2006, this rule was not in effect, and players could be drafted during/right out of high school.

The draft is comprised of 60 players and takes place over two rounds of 30 selections. Teams pick players in an order based on performance from the previous season, with teams that performed poorly getting earlier picks in order to create a seemingly more level playing field. It's important to note that there weren't always 30 players selected in each round, the number made its way up to 30 as more teams were added into the NBA.

The data you’ll be using comes from the dataset **nba\_draft.csv** and focuses on players that were selected in the first round of the NBA draft between the years 1990-2021, and they are divided based on what number in the first round they were selected (1-10, 11-20, or 21-30). For each player, there are a range of different statistics from their careers in the NBA, some of which are ongoing, we will be focusing on the average number of minutes played per game.

The graphs below show a density plot corresponding to the minutes per game based on players who were selected in Round 1 of the NBA draft. The dashed line represents the groups mean. The box plots represent the variance for the minutes played for players who were selected in the first round.



1. Based on this density plot and the box plots above, can we determine if there is a significant difference between the average number of minutes played per game based on a player’s pick in the first round of the draft? What conclusions can you draw upon first glance of the two data visualizations? Are there any concerns?

We could determine if there is a significant difference between the average number of minutes played for each group of picks in the NBA draft. The box plots show that the variance between the three groups seems to be relatively similar, which is necessary for an ANOVA test. There are limited extreme outliers which is good when looking at the distribution and variance, nothing that causes concern. The distribution of the groups seems to follow a relatively normal shape and have means somewhat close together. The assumptions for a one-way ANOVA test seem to be met.

1. Assuming an ANOVA test is most likely appropriate for our data, write and interpret in context the null and alternative hypotheses we'll be using to determine if the average number of minutes played per game is differs based on the position in which the player is picked in the first round.

Null Hypothesis:

 (Where is the population mean for that group)

The mean number of minutes played in the NBA for players who were picked 1-10, 11-20, and 21-30 in the NBA draft are all equal.

Alternative Hypothesis:

For at least one group, the mean number of minutes played in the NBA is not equal. (There is at least one group of draft picks that played a different mean number of minutes in the NBA.)

1. Using the statistics for each category provided below, fill in the ANOVA table and use it to answer the following questions.



k = number of groups

n = total sample size

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **SS** | **MS** | **F-statistic** | **P-value** |
| Pick in Draft | (k-1) = 3-1= 2 | 13839 | SSG/(k-1)= 13839/2= 6919.5 | MSG/MSE= 6919.5/50.1= 138.1 | 0.000 |
| Error | (n-k)= 926 – 3= 923 | SST = SSG + SSE60091=13839+SSESSE = 46252 | SSE/(n-k)= 46252/923= 50.1  |
| Total | (n-1)= 926-1= 925 | 60091 |

1. Is there a significant difference in the mean number of minutes played per game based on when a player was selected in the draft?

The p-value given in the ANOVA table is meaning there is strong evidence that there is a difference among the average number of minutes played between players selected 1-10, 11-20, and 21-30 in the first round of the draft.

1. Between which groups are we *most likely* to see a significant difference. Between which groups are we *least likely* to see a significant difference in minutes played. Provide evidence.

From looking at the means in the density plot and box plot, we are most likely to find a difference between those selected 1-10 and those selected 21-30 because there is the largest difference in mean number of minutes played. We are least likely to find a difference between those selected 11-20 and 21-30 because the difference in mean is the smallest.