**The NBA Draft**

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.

**Playing Basketball**

The goal in basketball is to score as many points as possible by throwing the ball through the other team's hoop. The two netted baskets, positioned on opposite ends of a rectangular court, are normally 10 feet high, and have a backboard behind them.

In the middle of the court there is a half-court line that divides the two sides. On both sides, surrounding the hoops, there is an arch called the three-point line. Within the three-point line stands the free-throw line, where a player would shoot from should there be a foul called.

During play, each team has five players on the court. The game starts with the referee tossing the ball in the air in between a player from each team, while each player tries to swat the ball to a member of their team (called a jump ball).

**Scoring**

Players can shoot towards the hoop from any point on the court. A different number of points is rewarded based on where the player is standing when they release the ball. The three shots are explained below.

- Field Goal: Worth 2 points, scored by shooting within the three-point line

- Three Pointer: Worth 3 points, scored by shooting outside the three-point line.

- Free Throw: Worth one point, taken from the free throw line after a foul.

Players move the ball up and down the court by dribbling, passing the ball to teammates, or shooting the ball. At the end of the game, which is played in four twelve-minute quarters, the team with the most points win.

A diagram of a pick in a draft pick

Description automatically generatedThe 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.

A graph of a diagram

Description automatically generated

1. Based on this density plot and the box plots above, do you feel as though ANOVA is appropriate? What conclusions can you draw upon first glance of the two data visualizations? Are there any concerns?

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. The distribution of the 10-20 and 20-30 groups seem to follow a relatively normal shape and have means somewhat close together. The 1-10 group is skewed slightly to the right which could be slightly concerning. A one-way ANOVA does seem appropriate, the conditions seem met and there could be a significant difference among the groups.

1. Now that we've determined ANOVA is most likely appropriate, write and interpret in context the null and alternative hypotheses we'll be using for the ANOVA test.

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, 10-20, and 20-30 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. A screenshot of a white and black text

   Description automatically generated 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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 | 2x10^-16 |
| Error | (n-k)  = 926 – 3  = 923 | SST = SSG + SSE  = 60091-13839  = 46252 | SSE/(n-k)  = 46252/923  = 50.1 |
| Total | (n-1)  = 926-1  = 925 | 60091 |

1. What F-Statistic did you find as a result of the ANOVA test? What does it indicate about the average number of minutes played based on a player’s pick category in the draft? What would an F-statistic close to 1 indicate? Explain.

The F-statistic is 138.1. This is a large f-statistic and indicates that the variability between groups is much larger than the variability within groups. This is because the F-statistic is a ratio of the mean square of groups (the variability between groups) and the mean square errors (the variability within groups). A large F-statistic means that the numerator (variability between) is much larger than the denominator (the variability within). The larger the statistic, it means the variation between is more than you would expect to see just from random chance. A f-statistic close to 1 would suggest that the variation between the groups is approximately equal to the variation within the groups and that you can’t conclude a difference between the group means (the null hypothesis is true).

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?

We can conclude that there is sufficient evidence to reject the null hypothesis that the mean number of minutes played per game is the same regardless of when a player was selected in the first round of the draft. The p-value given in the ANOVA table is 2x10^(-16) meaning there is strong evidence to reject the null hypothesis.

1. Using the ANOVA table, interpret the sum of squares for the groups. What does it imply about the data?

SSG = 13839 This number represents the total variability between the three groups and measures how far apart the mean minutes per game played between the players that were picked 1-10, 10-20, and 20-30. Since it is large, it suggests that there is a relatively large difference between the averages of the three groups.

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 20-30 because there is the largest difference in mean number of minutes played. We are least likely to find a difference between those selected 10-20 and 20-30 because the difference in mean is the smallest.