

# Project Part 3

## Project Question

Do players in the 2022 MLB season with at least a 0.275 batting average have an average hit distance greater than the population on average? With this question, I want to explore the possibility of what I will define as “good” hitters (batting average is at least 0.275) achieving further average hit distances. This question could also give insight on how a team should set up their defense based on a hitter’s average hit distance. I am choosing 0.275 as the threshold for “good” hitters as this is higher than what is generally accepted by the MLB community as being an average hitter (0.250 batting average)[2].

## Data Description

The two data sets I will use both come from <https://baseballsavant.mlb.com/>, which is a data collection site using STATCAST powered by GCP. This site is directly maintained by the MLB, therefore making it a primary source. The data contains aggregated hitting statistics for every eligible hitter for the entire 2022 season. The data was directly collected from each player’s game, with an aggregation of each player’s statistics conducted after each game. Since every eligible hitter in the MLB is accounted for in the data, the data is a population, and conclusions can only be made about MLB hitters. Since the data is coming directly from the player’s performance in every game, it makes a great tool to use when conducting my statistical test, as population parameters can be calculated. The data also contains the two variables I want to test, batting average and average hit distance.

## Test Identification

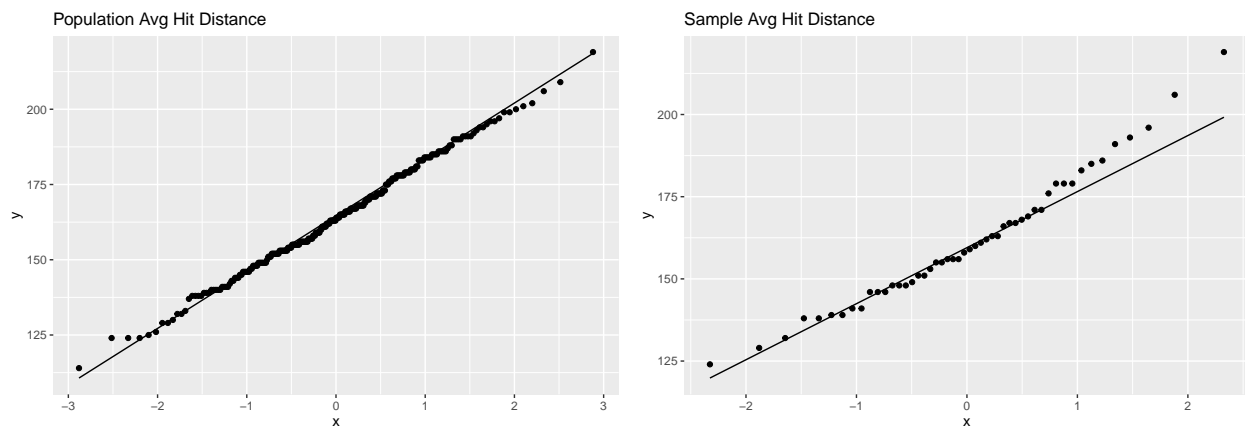
The test I will use is the one-sample t-test. I will be using this test since I am trying to figure out whether an unknown mean parameter is different from a specified value.

## Test Characteristics

The test characteristics to use a one-sample t-test are a population parameter, a sample, and a sample parameter to compare against the population parameter. All three are met for my research question. The population parameter is the mean of average hit distance among the entire population of MLB hitters. The sample is MLB hitters who have a batting average of at least 0.275. The sample parameter is the the mean of average hit distance among the sample of “good” MLB hitters.

## Test Assumptions and Validity of Assumptions

The one-sample t-test requires the data to be approximately normally distributed. The QQ plots below show that both the population and sample for average hit distance is approximately normal, with the majority of points falling on or near the QQ line.



## Test Hypothesis

The null hypothesis is the average hit distance of “good” hitters is equal or less than the average hit distance of all hitters. The alternative hypothesis is the average hit distance of “good” hitters is greater than the average hit distance of all hitters.

## One-Sample T-Test

```
##  
## One Sample t-test  
##
```

```
## data:  sub$avg_distance
## t = -0.97183, df = 49, p-value = 0.832
## alternative hypothesis: true mean is greater than 163.9921
## 95 percent confidence interval:
##  156.4923      Inf
## sample estimates:
## mean of x
##      161.24
```

## Test Result and Conclusions

The p-value obtained from the one-sample t-test is 0.832. Since the p-value obtained is greater than the alpha value of 0.05, we fail to reject the null hypothesis. Based on the statistical test result, we can conclude that there is no evidence that “good” hitters tend to hit the ball farther on average when compared to all MLB hitters in the 2022 season.

## Generalization

This test result can be generalized beyond the data to multiple MLB seasons or across different baseball leagues. Since values are averaged across an entire season and the rules of baseball tend not to change from season to season, it is safe to say that this result will hold for previous MLB seasons. Also, since the MLB is one of the most diverse leagues around the world and contains athletes from a variety of countries, the test result will hold when looking at other top tier baseball leagues around the world.

## References

- 1) <https://baseballsavant.mlb.com/>
- 2) <https://coachingkidz.com/the-ultimate-guide-to-batting-average-in-baseball/>