Vince Laird

INFERENTIAL STATISTICS
DATA SCIENCE - CAREER TRACK

Introduction

Using the data from the 354 MLB baseball games from April 2016 as a data set, I wanted to look at opening odds and closing odds for teams that were considered to be the favorite. If a team opened as a favorite, were you better off betting immediately or waiting until right before the game started?

Reviewing the data

The Jupyter Notebook used for this project can be found here:

https://github.com/vincelaird/springboard/blob/master/capstone/Inferential%20Statistics.ipynb

For this analysis, a new dataframe was created with one row per game. Opening odds and closing odds were contained on each row. If the favorite odds became worse, the number followed by the three-character string representing the team would be more negative. Example: PIT-110 at open becomes PIT-120 at close.

If the team that was considered the favorite was different from when the odds opened vs. when they closed, this would indicate that the underdog odds became worse. Also, games where the opening odds did not change at all would be ignored. Overall, the data showed the following:

Total games: 354
Games where odds did not change from open to close: 16
Games where the favorite was different from open to close: 43
Games where the favorite odds became worse from open to close: 151
Games where the underdog odds became worse from open to close: (338 - 151 - 43) = 144

The total count of games where the favorite odds became worse was 151, vs. underdog odds becoming worse which was 187.

If our null hypothesis was that favorites and underdogs each should have an equal chance of receiving worse odds, when comparing opening and closing numbers, we should not have seen such a disparity. The probability of observing a sample like this is 2.5%, if our null hypothesis is correct.

Conclusion

Though the adage is to bet favorites early, in the case of MLB baseball opening month, the underdog odds actually became worse when we compared opening odds to closing odds: you'd be better off betting dogs early, as their odds seem likely to become worse.

While somewhat counterintuitive, this could be caused by expectations for teams heading into the season - perhaps the poor teams didn't quite play as poorly as expected, and the strong teams didn't quite play as well as expected. An interesting follow-up might be to see if this holds true throughout the season, vs. only the opening month.