
ISyE 2028 – Basic Statistical Methods - Fall 2015
Bonus Project: “Big” Data Analytics
Proposal (or Final Report)

Kevin Breslin

Problem Description:

There is an old adage in baseball that says “walks will kill you,” meaning that walking batters will greatly hinder a team’s chance of winning games. It has been theorized and taught by many baseball professionals that 80% of walked batters will score, serving as a deterrent for pitchers to walk batters and give them a free pass to reach base. I plan to analyze data from Georgia Tech Baseball to see if the percentage of runs scored against them from walks in the 2015 season is indeed equal to or within a reasonable range of 80%.

Data Source:

I plan on collecting the number of walks by Georgia Tech pitchers that resulted in runs for the opposing team in the 2015 season, along with possibly other telling statistics such as the percentage of overall runs scored against Georgia Tech that resulted from walks. To collect the data, I will use statistics provided by Georgia Tech Baseball. I will also use statistics from <http://www.ramblinwreck.com/sports/m-basebl/stats/2014-2015/teamstat.html>, where I can get box scores from each game. I can compile this data into a spreadsheet and use R to find statistical values to analyze the findings.

Method:

I plan to find a five number summary of the percentage of walked batters who score per game as well as construct a histogram of these findings for the entire 2015 season in order to better visualize the data. I also plan on conducting a two-tailed 95% confidence interval, with a null hypothesis of the mean percentage being 80% and an alternative hypothesis of the mean being unequal to 80%. While the data will not be representative of the mean percentage throughout the game of baseball at all levels, I expect the percentage for Georgia Tech Baseball to be somewhat below the 80% mark, closer to 60%.