

---

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

---

**Abstract:**

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, which would give them a free pass to reach base. I analyzed data from Georgia Tech Baseball to see if the runs scored against them from walks in the 2015 season is indeed equal to or within a reasonable range of 80%. I have been surprised that many believe the rate to be as high as 80% ever since I first heard the proposition years ago. I believe that after analyzing data, I will find a rate around 60%, which would be significantly lower than the perceived level of 80%.

**Method:**

I collected my data by clicking on each game listed on <http://www.ramblinwreck.com/sports/m-basebl/sched/geot-m-basebl-sched.html> and then clicking on the Gametracker link to observe the play-by-play information of each game. From there, I could gather who pitched for Georgia Tech in each game, how many walks each pitcher surrendered, and how many of those walks came around the bases to score. Also, after contacting WREK radio, who broadcasts the games, I was also able to check my findings with statistics provided by Wiley Ballard, who is the radio voice of GT Baseball. I compiled my data into an Excel spreadsheet, where I recorded statistics such as walks per GT pitcher each game and percentage of walks that scored each game.

Once I recorded all of the data and found totals and percentages of desired categories, I used a z-score to conduct a two-tailed, 95% confidence interval for a population proportion. I also constructed two histograms.

The first depicts the frequency of the percentage of walks that scored each game. In order to highlight the importance of preventing as many runs from scoring as possible, the second histogram shows the margin of loss of Georgia Tech's closest games in 2015. I thought it would be interesting to break down my findings by pitcher as well, so I included a chart of all of Georgia Tech's pitchers who played in 2015, ranked in order of highest to lowest percentage of walked batters who scored.

### **Calculations:**

Parameter of interest: The percentage of walked runners who score,  $p$ .

Hypothesis:

$$H_0: p = 0.80$$

$$H_1: p \neq 0.80$$

Test statistic:

$$Z_0 = \left( Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Point estimate of proportion parameter:  $\hat{p} = 0.312$  (See Interesting Statistics below for a calculation of  $\hat{p}$ )

95% Confidence Interval:  $\alpha = 0.05$ ,  $Z_{\alpha/2} = z_{0.025} = 1.96$

$$n = 250$$

Rejection region: Reject  $H_0$  if  $p$  falls outside of  $\hat{p} \pm Z_0$

Computations:

Confidence Interval:

$$\left( \hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$
$$\left( 0.312 - 1.96 \sqrt{\frac{0.312(1-0.312)}{250}}, 0.312 + 1.96 \sqrt{\frac{0.312(1-0.312)}{250}} \right)$$

$$(0.312 - 0.0574324817, 0.312 + 0.0574324817)$$

$$= (0.2545675183, 0.3694324817)$$

Conclusion:

Since 0.80 does not fall within the interval (0.2545675183, 0.3694324817), I can reject the null hypothesis and conclude that the true proportion of walked runners who score p is not equal to 0.80. According to the gathered data, I am 95% confident that the true proportion p is within the range (0.2545675183, 0.3694324817), meaning that between about 25% and 37% of runners who reach base from being walked circle the bases to score.

Histograms:

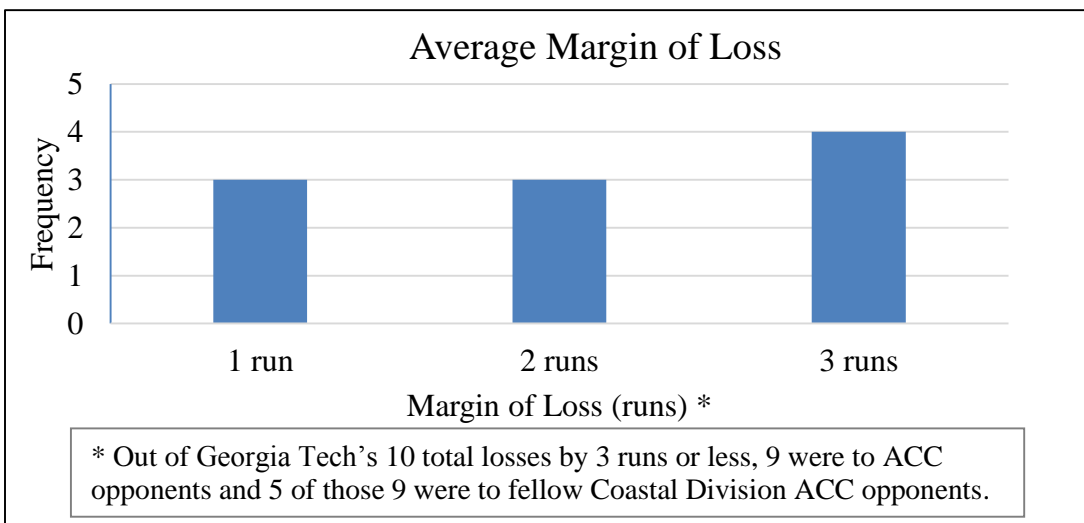
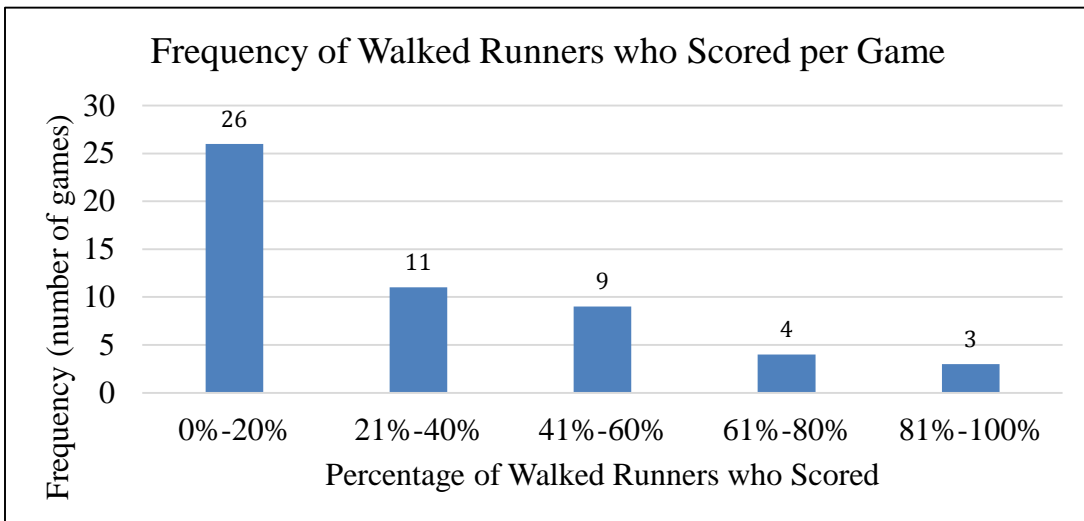


Chart:

<b>By the Pitcher:</b>			
<b>Pitcher</b>	<b>Total Walks that Scored/Pitcher</b>	<b>Total Walks/Pitcher</b>	<b>Total Walks Turned Runs/Pitcher %</b>
Pope	1	2	50%
Gooden	5	11	45%
Schniederjans	3	7	43%
Pitts	12	32	38%
Gorst	9	26	35%
Shelton	6	17	35%
Stanton	8	23	35%
Craport	1	3	33.33%
J. Wiseman	2	6	33%
Phillips	4	12	33%
Datoc	2	7	29%
Gold	7	25	28%
King	7	25	28%
Ryan	4	15	27%
Parr	7	34	21%
P. Wiseman	0	5	0%

<sup>1</sup> It is important to note that the total number of innings pitched varied greatly among the different pitchers, so Total Walks that Scored/Pitcher and Total Walks/Pitcher are not fully representative of performance.

Interesting Statistics:

Total Games: 55      Total Walks: 250      Total Walks that Scored: 78

Total Walks that Scored / Total Walks:  $\hat{p} = 0.312$  (Therefore, 1 Walk = 0.312 Runs)

Walks per Game: 4.545      Runs from Walks per Game: 1.418

Total Runs against GT in 2015: 306      Runs against GT resulting from Walks (%): 25.49%

1 Less Walk per Game (55 Less Walks; 195 Total Walks on the Season)\*: Saves 17.16 runs

1/3 Less Walks Overall (83 Less Walks; 167 Total Walks on the Season)\*: Saves 25.896 runs

Average Margin of Loss: 5.391 runs

\* Assuming that there would be an even correlation of 1 less walk equaling 0.312 less runs

## **Conclusion:**

After analyzing the data, it seems unlikely that 80% of walked runners round the bases to score, at least based off of Georgia Tech Baseball's 2015 season. I was even much too high with my prediction of 60%. The data suggests that the true rate is around 31.2%, give or take about 5.7%. This is definitely lower than I expected, but it is important to keep in mind that this is a miniscule sample size compared to one that would be more indicative of the true proportion in the entire game of baseball. The sample size is perfectly representative for evaluating Georgia Tech's 2015 season, but it is a biased sample when used to generalize the true population proportion for all of baseball since it draws from one season of one collegiate team and does not include statistics from any professional teams. To find a more representative sample, I could use statistics from teams around Major League Baseball and find the rate over the course of several seasons to keep bias from skewing the data too much. Despite my analysis not being representative of all of baseball, it can still be used for real-life application.

Even though it may seem like preventing just few walks over the course of an entire season might have a negligible difference, there can actually be important changes that come from keeping runners from getting free passes. For example, Georgia Tech made the 2015 ACC tournament by 1 game on the final day of the regular season. Since they were such a high seed, they were faced with playing the low-seeded University of Virginia, the eventual College World Series champions, in the opening round and Tech was eliminated after losing their first game 11-0. If they had prevented just a few more walks that turned into runs against ACC opponents and won a few more games, they could have had a better seed entering the tournament and would have therefore played a team ranked not as high as UVA, increasing their chances of lasting longer in the tournament.