Investigating Data Mythology in Major League Baseball

Danielle Hoffman

Dr. David Agard and Dr. Joseph Nolan

Department of Mathematics and Statistics, Northern Kentucky University

NKU NORTHERN KENTUCKY UNIVERSITY

Introduction:

Baseball statistics have been collected for over a century and are used to analyze players' abilities. This study explores the relationship between physical characteristics, career success measures and player demographics. Research questions include:

- Is there a dependence between handedness and position?
- Does handedness affect a player's career longevity, batting success or pitching success?
- Is height a factor in a player's position?

Data:

The data for this study was found on three internet sources: Baseball America[1], Baseball Reference[2], and Sean Lahman's Baseball Database[3]. Data from the 1901 to 2013 seasons were used in the analyses.

The data only includes players who have played at least 100 games, batters with 2,500 at bats, Starting Pitchers with 100 games played and Relief Pitchers with 200 games played. Pitchers who started more than 25% of their games were categorized as Starting Pitchers.

- The player demographic variables include fielding position, height, throwing hand and batting hand.
- The variables relevant to hitters include home runs (HR), batting average, and runs batted in (RBI).
- The variables relevant to pitchers include wins and earned run average (ERA).

Methods:

- A T-test was used to determine if the proportion of left-handed baseball players is greater than the general population proportion of 10%.
- For positions with left-handed players, a Chi-Square Test of Independence was utilized to determine if there is a dependence between position and handedness.
- Data were analyzed using a One-Way Analysis Of Variance (ANOVA) to compare players' career batting success and batting handedness.
- A Two-Way ANOVA examines the effect of throwing handedness and type of Pitcher on career pitching success.
- T-Tests and ANOVA test assess the effect of handedness on career longevity.
- An ANOVA test was performed to compare the average heights of players and their primary position.

Data Results:
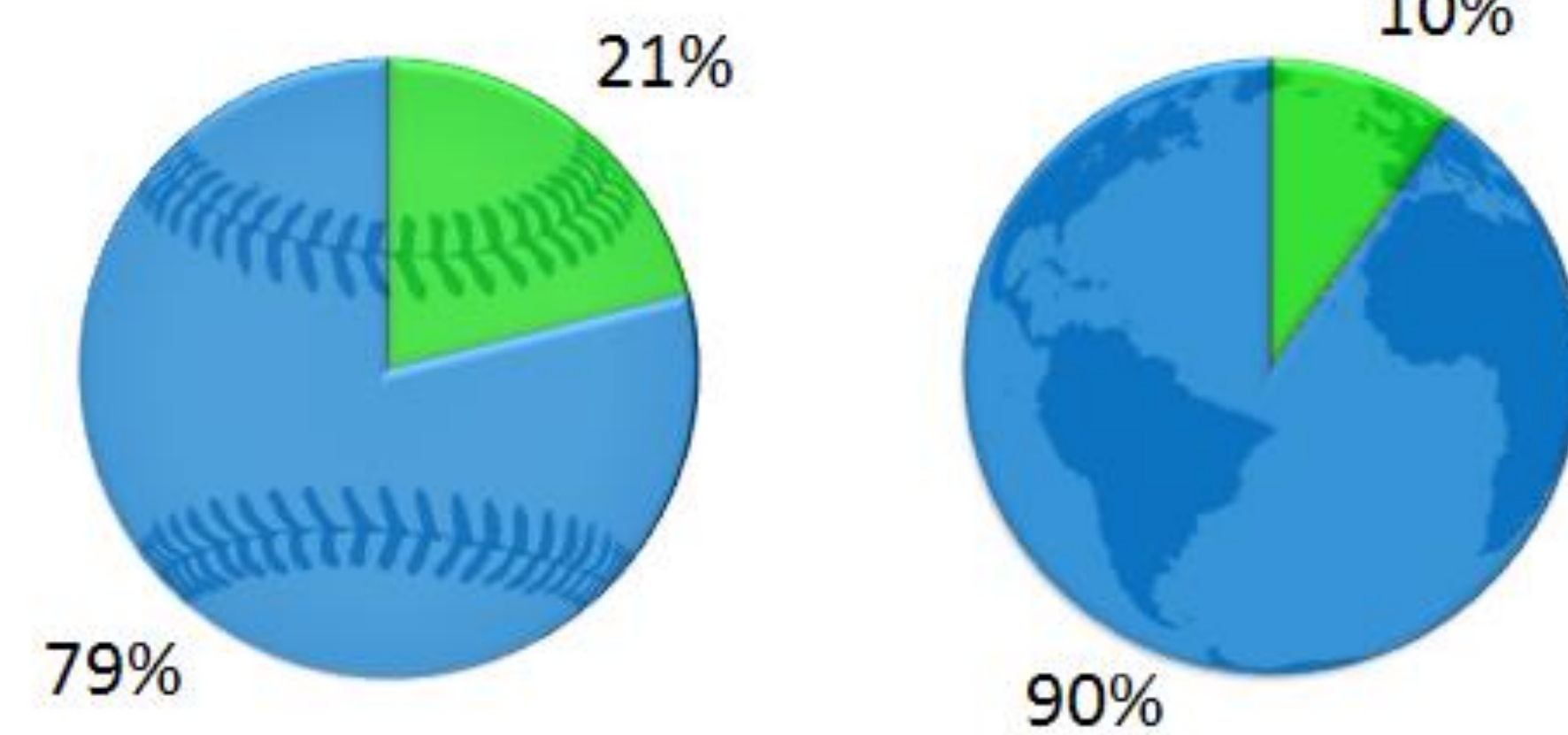
Figure 1: Pie charts illustrating handedness in Major League Baseball versus the General Population. There are no left-handed players at 2B, 3B, SS or C. (95% CI: 0.204060-0.218376, p-value < 0.001)

Figure 2: A bar chart of career ERA for Pitchers by handedness and type of Pitcher. A significant difference in career ERA was found. Right-handed, Starting Pitchers have higher career ERA than right-handed, Relief Pitchers. (F = 4.79, p-value = 0.029)
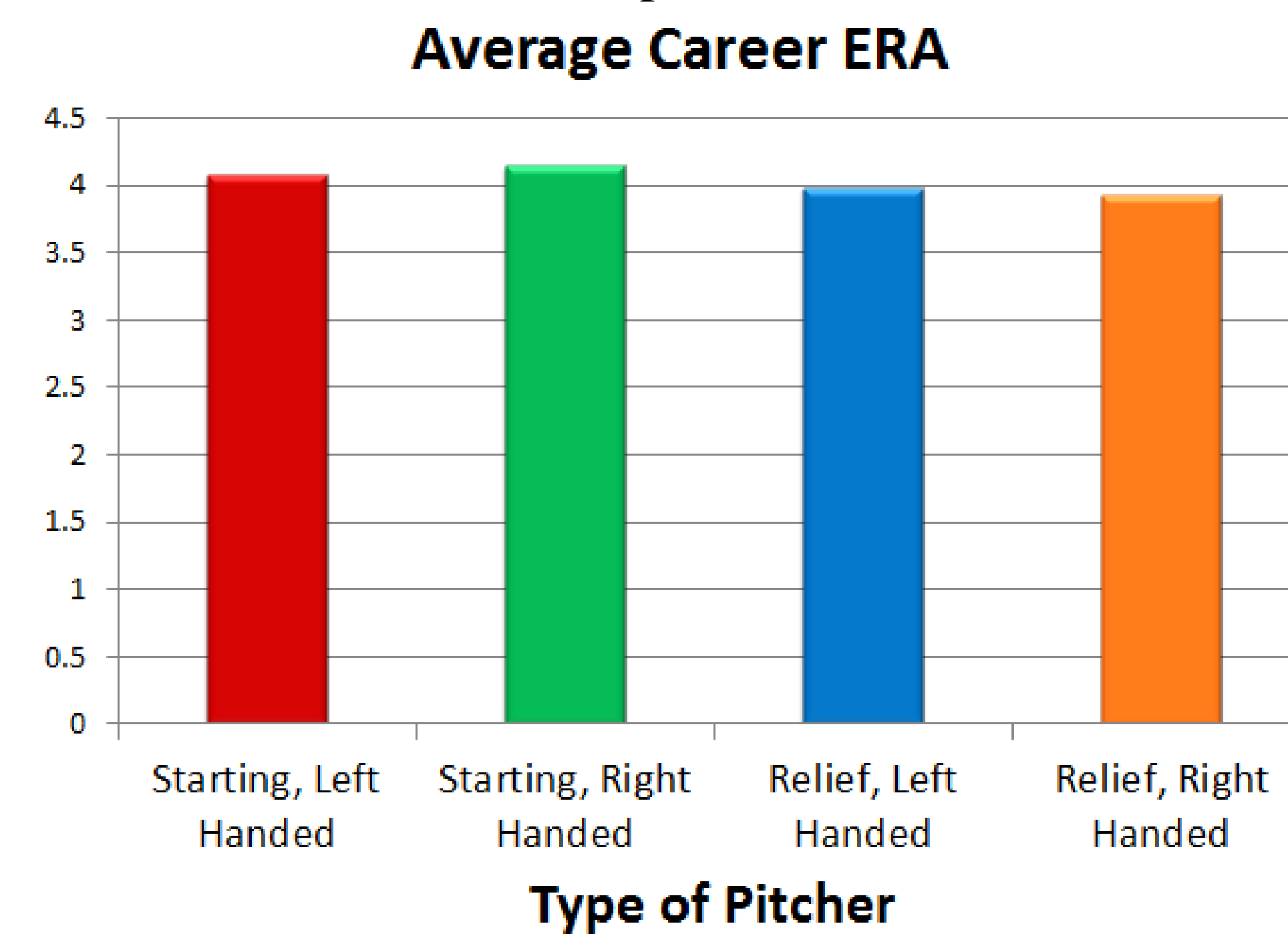
Figure 3: Bar charts showing career batting average, average number of HR and average number of RBI. ANOVA indicated left-handed batters have a higher career batting average. Switch hitters average lower amounts of career HR and RBI than left and right-handed batters.
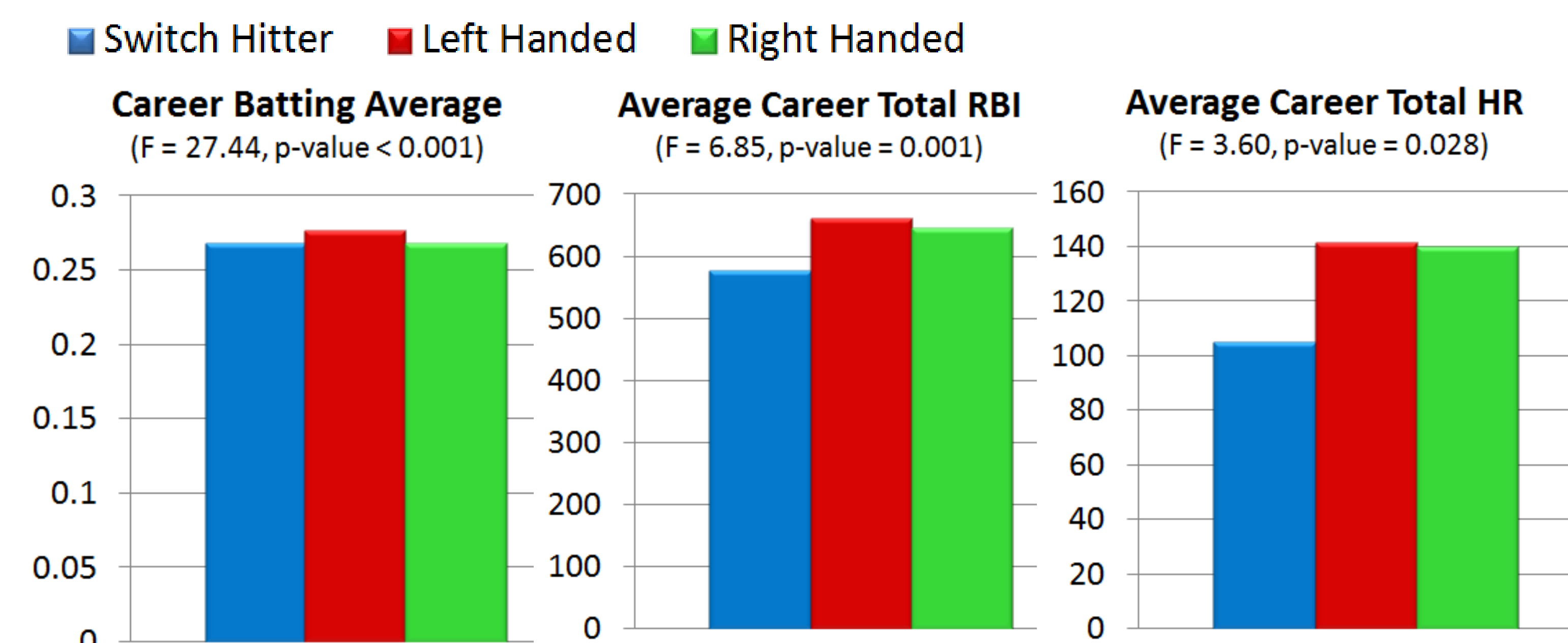

Table 1: This is a contingency table of handedness and position. Included are observed and expected counts. The percent of left-handed 1B is higher than P or OF. (χ^2 = 117.536, p-value < 0.001)

	Left Handed	Right Handed	Total
1B	321	391	712
	203.8	508.2	
OF	539	1702	2241
	641.4	1599.6	
P	1750	4416	6166
	1764.8	4401.2	
Total	2610	6509	9119

	Left	Right	Both
Hitters (batting hand)	1449.50	1401.70	1468.90
Starting Pitcher	284.62	269.91	
Relief Pitcher	449.10	416.51	

Table 2: A table displaying the average number of games played for hitters, starting pitchers or relief pitchers' careers. Left-handed Relief Pitchers appear in more games than right-handed Relief Pitchers. (t = 3.97, p-value = 0.047) No significant difference between handedness and number of games played exists for hitters or Starting Pitchers.
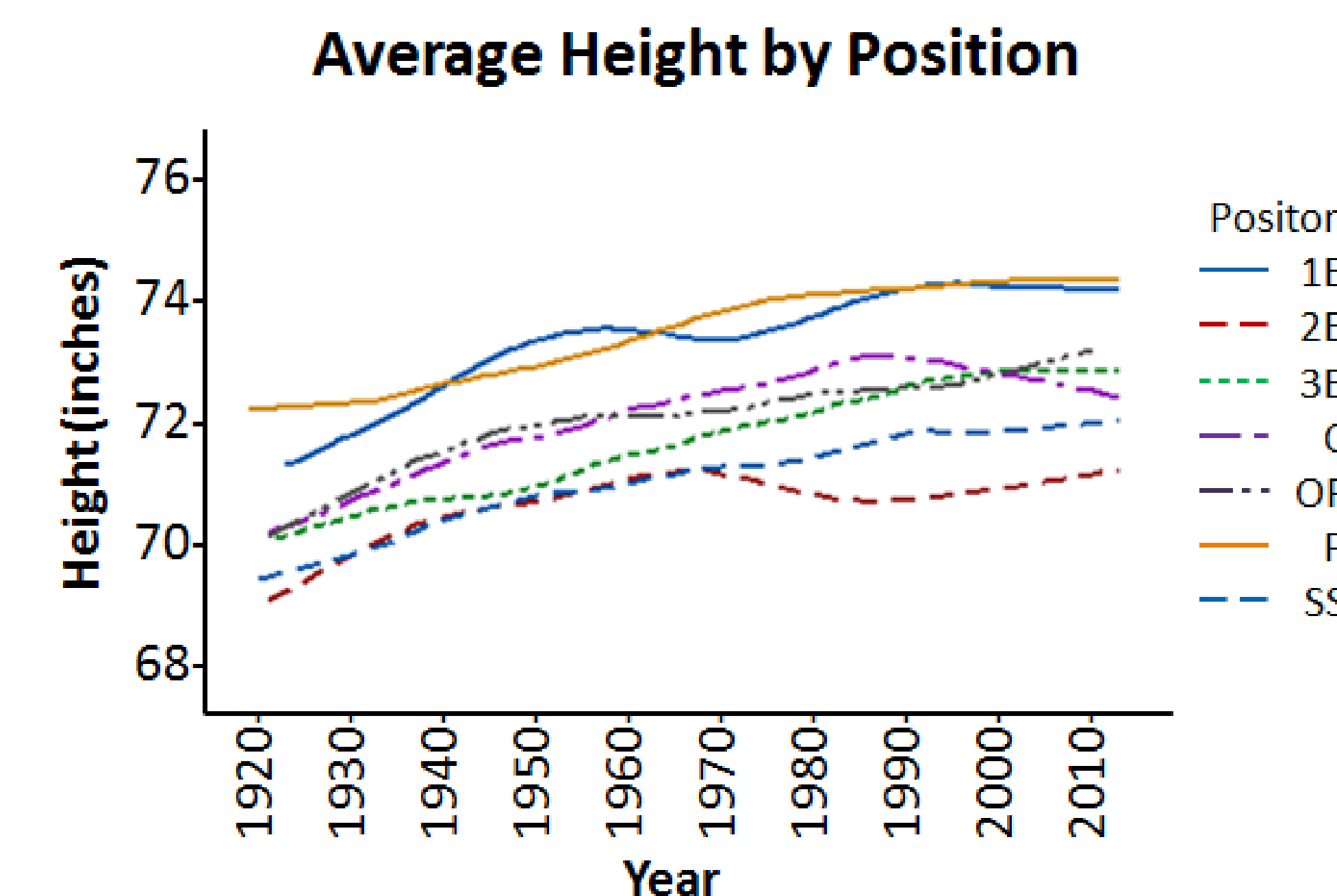
Figure 4: A time plot of average heights of players by their primary position. Here we find a significant difference of height by position. In particular, P and 1B are taller than any other position. (F = 470.39, p-value < 0.001)

Conclusions:

- The percentage of left-handed athletes in Major League Baseball is higher than the general population. (Figure 1)
- Left-handed athletes only play at some positions on the field: First Base, Pitcher and Outfield.
- Among positions with at least one left-handed player, there is a dependence between a player's position and the hand they throw with. The percentage of left-handed First Basemen is higher than Pitchers or Outfield. (Table 1)
- On average, right-handed Starting Pitchers have a higher career Earned Run Average (ERA) than right-handed Relief Pitchers. (Figure 2)
- Based on the number of games played in a career, the average length of a Relief Pitcher's career is longer for left-handed pitchers than for right-handed pitchers. (Table 2)
- Left-handed batters have a higher career batting average than right-handed and switch handed batters. Also, career average Home Runs (HR) and Runs Batted In (RBI) are lowest for switch hitters. (Figure 3)
- On average, Pitchers and First Basemen are taller than any other position. (Figure 4)

Future Work:

Ideas to look at in the future:

- Create a Minor League Baseball database
- Conduct similar analyses of Minor League data and compare to this study's results
- Determine the probability of a player in the Minor Leagues making it to the Major Leagues based on the player characteristics

References:

1. *Baseball America*. The Enthusiast Network, 2014. Web. 5 Aug. 2014.
2. *Baseball-Reference.com*. Sports Reference LLC, 5 Aug. 2014. Web. 5 Aug. 2014.
3. Lahman, Sean. *Lahman's Baseball Database*. 14 Feb. 2014. *SeanLahman.com*. Web. 19 May 2014.

Acknowledgements: