EPRS 8530 Project 1 – Cover Page

1. Kayla Myers, Early Childhood Education and Mathematics
2. Relationship between Major League Baseball teams' 2014 winning percentage, average home game attendance, and median player salary
3. General interest
4. Descriptive statistics, Correlation, Frequency Tables, Graphs (Histograms & Scatterplots)
5. Microsoft Excel 2013
6. Sources: mlb.com/standings; espn.go.com/mlb/attendance; usatoday.com

In this paper, I use measures of descriptive statistics and correlation to analyze the relationships between MLB teams' winning percentages, average home game attendance, and median player salary for the 2014 baseball season.

Raw Data

$N = 30$

MLB Team	Win Percent	Avg Home Attendance	Median Player Salary
Los Angeles Angels	0.605	38,221	900,000
Washington Nationals	0.593	31,844	2,850,000
Baltimore Orioles	0.593	30,805	1,000,000
Los Angeles Dodgers	0.58	46,695	4,300,000
St Louis Cardinals	0.556	43,711	767,500
Detroit Tigers	0.556	36,014	1,775,000
Kansas City Royals	0.549	24,154	1,500,000
San Francisco Giants	0.543	41,588	4,000,000
Pittsburgh Pirates	0.543	30,155	1,200,000
Oakland Athletics	0.543	25,045	775,000
Seattle Mariners	0.537	25,485	750,000
Cleveland Indians	0.525	18,428	975,000
New York Yankees	0.519	42,520	3,684,426
Toronto Blue Jays	0.512	29,327	3,000,000
Milwaukee Brewers	0.506	34,535	850,000
Atlanta Braves	0.488	29,065	1,142,857
New York Mets	0.488	26,860	1,000,000
San Diego Padres	0.475	27,103	1,980,000
Miami Marlins	0.475	21,386	1,000,000
Tampa Bay Rays	0.475	17,857	1,475,000
Cincinnati Reds	0.469	30,576	1,200,000
Chicago Cubs	0.451	32,742	1,622,857
Philadelphia Phillies	0.451	29,924	1,587,500
Chicago White Sox	0.451	20,896	1,500,000
Boston Red Sox	0.438	36,494	3,100,000
Minnesota Twins	0.432	27,785	1,000,000
Houston Astros	0.432	21,627	507,600
Texas Rangers	0.414	33,564	1,375,000
Colorado Rockies	0.407	33,090	2,025,000
Arizona Diamondbacks	0.395	25,601	1,937,500

More Money, More Problems: An Analysis of MLB Teams' 2014 Winning Percentages, Average Home Game Attendance, and Median Player Salary

Kayla Myers
EPRS 8530 Quantitative Methods I
October 13, 2014

Description

As a baseball fan, I was interested in the relationship between money, fan support, and team performance. The politics surrounding ideas of salary caps and unfair advantages prompted this inquiry. Is a higher payroll an advantage? Does a higher payroll produce a higher game-day turn out from the fans? Do either of those affect the team's performance? This paper analyzes the relationships between all 30 Major League Baseball Teams' 2014 winning percentages, average attendance for home games in 2014, and median player salary amounts in 2014. Only regular season games were considered.

Methods

I gathered data from 3 separate online sources. The first source, mlb.com/standings, provided the 2014 regular season standings. From this data, I collected each team's winning percentage (N = 30). This number is calculated by dividing the team's total wins by total games in 2014. The second source, espn.go.com/mlb/attendance, provided the 2014 regular season average game attendance for each team. From this data, I collected each team's average home game attendance (N = 30). The third source, usatoday.com/sports/mlb/salaries/2014/team/all, provided salary information for all MLB teams in 2014. From this data, I collected the median player salary amounts in 2014 for each team (N = 30). With these three sets of data, I created frequency tables and histograms for attendance and salary, which tell a story by themselves. I used descriptive statistics to analyze each of the three data sets, including mean, standard deviation, median, mode, and range. Then, I used correlation (Pearson's r and r^2) along with scatterplots to illustrate relationships between each of the three variables. Each of these graphs and tables are included in the results. From these correlation measures, I draw conclusions about relationships, which I include in the discussion.

Results

Table 1
Average Home Game Attendance Frequency (N = 30).

Limits	***F***
17,000-20,000	2
20,001-23,000	3
23,001-26,000	4
26,001-29,000	3
29,001-32,000	7
32,001-35,000	4
35,001-38,000	2
38,001-41,000	1
41,001-44,000	3
44,001-47,000	1

Table 2
Median Player Salary Frequency (N = 30).

Limits	***F***
500,000-880,000	5
880,001-1,260,000	9
1,260,001-1,640,000	6
1,640,001-2,020,000	3
2,020,001-2,400,000	1
2,400,001-2,780,000	0
2,780,001-3,160,000	3
3,160,001-3,540,000	0
3,540,001-3,920,000	1
3,920,001-4,300,000	2

Table 3
Descriptive Statistics (*N* = 30).

Category	*Mean*	*SD*	*Median*	*Mode*	*Range*
Winning Percentage	0.5	0.06	0.497	0.543	0.21
Average Attendance	30,436.57	7,391.06	30,039.50	n/a	28,838
Median Player Salary	1,692,674.67	1,019,935.47	1,425,000	1,000,000	3,792,400

Figure 1
Average Home Game Attendance Histogram (*N* = 30).

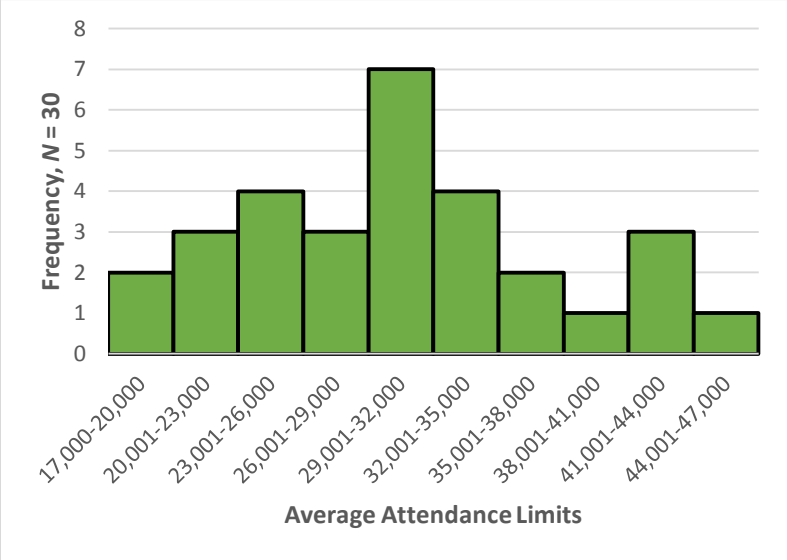


Figure 2
Median Player Salary Histogram (*N* = 30).

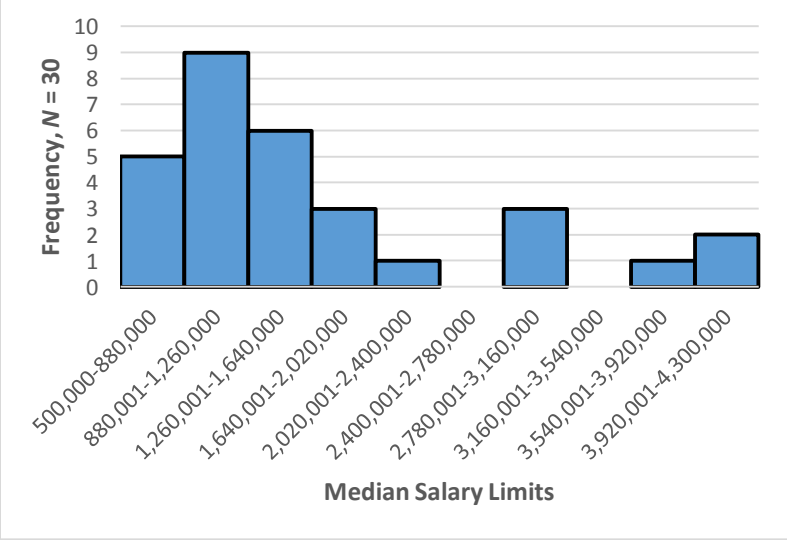


Table 4
Correlation Matrix (*N* = 30).

Category	Winning Percentage	Average Attendance	Median Player Salary
Winning Percentage			
Pearson Correlation	1		
Common Variance	1		
Average Attendance			
Pearson Correlation	0.352	1	
Common Variance	0.124	1	
Median Player Salary			
Pearson Correlation	0.126	0.561	1
Common Variance	0.016	0.315	1

Figure 3
Scatterplot: Winning Percentage and Average Home Game Attendance (*N* = 30).

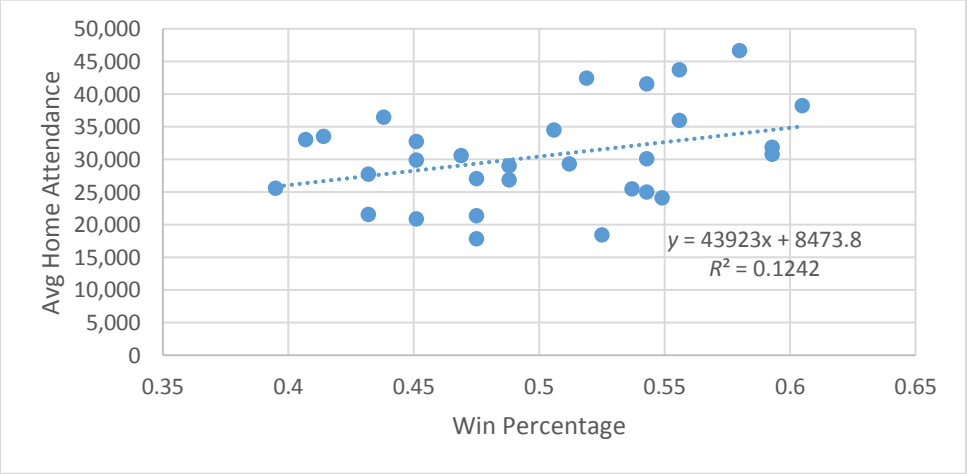


Figure 4
Scatterplot: Winning Percentage and Median Player Salary (*N* = 30).

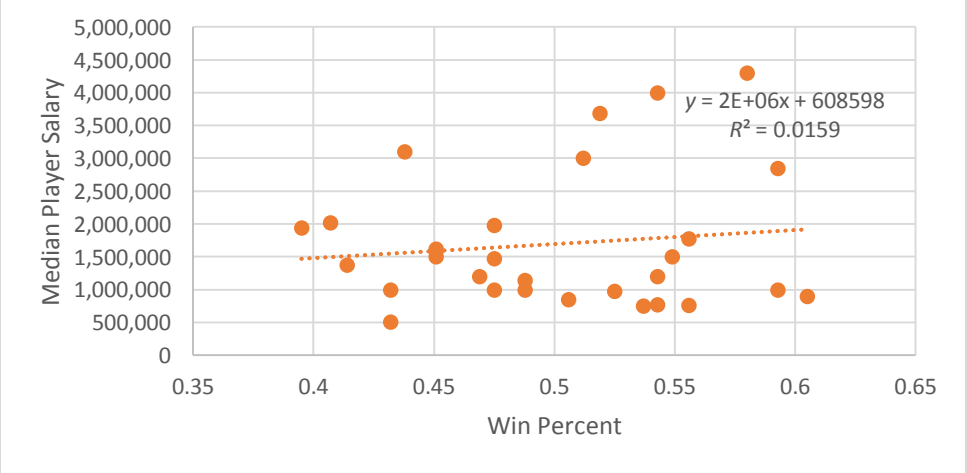
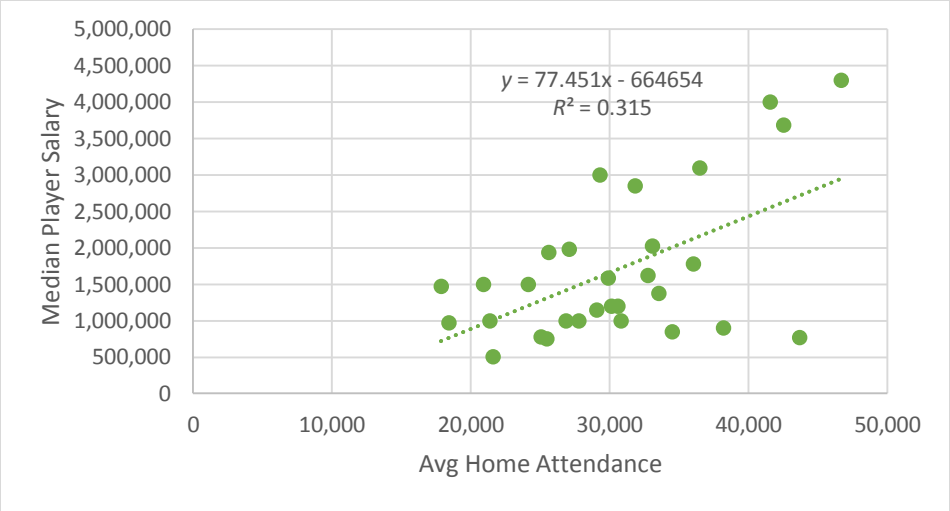


Figure 5
Scatterplot: Average Home Game Attendance and Median Player Salary (*N* = 30).



Discussion

The descriptive statistics I found for attendance and salary were surprising. I had no idea just how vast the range was for these statistics. Different ball parks have different seating capacities, but the range of attendance is greater than the range of available seating. The graph for attendance is relatively normal. In the future, I would be interested to look at the percentage for each ball park. I chose to use median instead of mean player salary, expecting that the median would be a truer representation of the data, but even the median produced a positively skewed graph. The range of salaries is more than alarming, it's nauseating.

The correlational statistics were just as surprising. Each of the correlations was positive, but none were as strong as expected. The strongest correlation was between attendance and salary (r = 0.561). This does not imply that a higher payroll causes higher ticket sales, nor does it imply that higher ticket sales cause higher payroll, but it does mean there is a relationship. There was a slight correlation between attendance and win percentage (r = 0.352). Again, I expected this correlation to be stronger, since a team would expect to win more games if the attendance is higher. This may imply that a certain population of super-fans attend games for their favorite team regardless of win percentage, while others are more likely to go to games when the team is doing well.

The weakest correlation was between salary and win percentage (r = 0.126)! This does not imply causality (or a lack thereof), but it does imply that there is very little relationship between the team's performance and how much players are paid. This, in my humble opinion, is a big step in the direction of discrediting arguments for implementing salary caps for Major League Baseball Organizations. As these results indicate, a team having more money on payroll does not necessarily mean the team will win more games. This interesting result has implication for future research, possibly considering more variables of player salary or winning records over longer periods of time.

Go Braves!