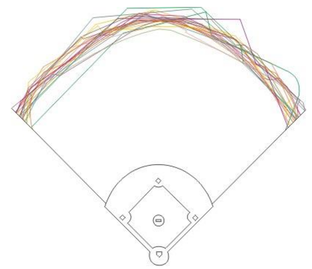
Do Field Dimensions Affect Baseball Outcomes?

Stoltenberg R., Williams K., Harris A.

CALIFORNIA STATE UNIVERSITY, CHICO

MATH 615, Statistical Methods for Graduate Research



Background

Major League Baseball fields have mathematically consistent bases, pitcher's mounds, batter's boxes, foul lines, and infields. However, the outfield fence dimensions have always been field specific. These small variations in wall distance and height can be the difference between a home run or an out, a win or a loss. The dimensions of ballparks outfield and foul ground are rarely identical, with different playing field area, distance of walls, height of walls, and even distinctive angles and wall materials.

Data

The data population used in this analysis is field specific pitching stats accumulated over the past 15 years at all 30 Major League Fields, excluding those built in the last 15 years. To do this analysis, all 30 fields were designated a single number to rank them as a long or short field. This was done by a five-number ranking system taking into account five outfield data points: left field line, left center field gap, center field, right center field gap, and right field line. This 5-number ranking system is representative of field's total outfield area (see "Fields" data below). The analyzed data was sourced online from ESPN.com's archive.

Table 1: Field Score Data

League	Stadium Name:	LFL	LCF	CF	RCF	RFL	TOTAL SCORE
American	Fenway Park	310	365	390	375	302	1742
American	Tropicana Field	315	370	404	370	322	1781
National	AT&T Park	339	375	399	365	309	1787
National	Wrigley Field	355	365	400	368	353	1841
American	Kauffman Stadium	330	387	410	387	330	1844
National	Coors Field	347	385	415	375	350	1872

Project Goals

1. To explore the relationship between outfield fence dimensions and earned runs (ER), batting average against (BAA), doubles (X2B), Triples (X3B), home runs (HR), on base plus slugging (OBP), and total bases (TB).
2. To explore the relationship between year and each of these seven outcomes (ER, X2B, X3B, HR, OBP and TB).

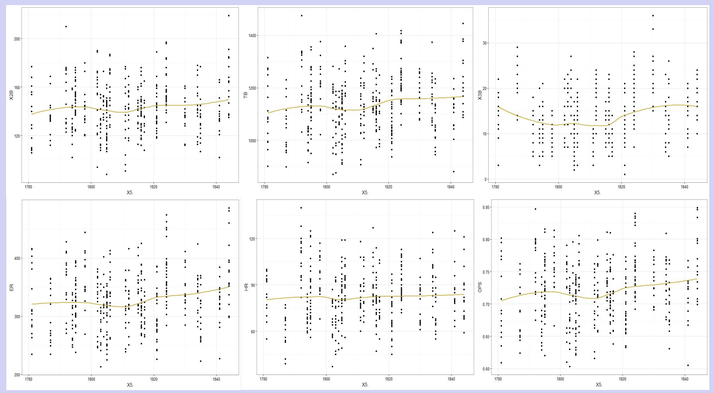
Methods

Data were derived from the ESPN.com archive. Data consolidation and conversion was done in Microsoft Excel. Multiple linear regression was performed for each outcome (BAA, ER, X2B, X3B, OPS, and TB with both field size (rank) and year as predictors. All statistical analyses were run using RStudio (R, 3.2.2). The data used for these analyses did not include the two outlier fields: Boston and Colorado. These two fields had the longest and shortest 5 number outfield dimension scores. Denver's Coors Field which was built at an elevation of 5200 ft. above sea level. Higher elevation leads to thinner air and more carry on balls in play. Boston was also identified as an outlier due to Fenway Park's "Green Monster" left field wall that rises 37 feet above the outfield. This feature contributes to variable offensive statistics.

Table 2: Descriptive Statistics

Variable	Mean	SD	Range	Max	Min
5#	1811	23.1	130	1872	1742
BAA	0.25	0.02	0.091	0.29	0.2
ER	327	48.9	273	486	213
X3B	13.4	5.7	35	36	1
X2B	143	20.3	131	219	88
HR	82.2	17.4	103	140	37
OPS	0.72	0.05	0.246	0.85	0.6
TB	1137	109	605	1475	870

Figure 1: Field rank versus outcome variables (ER, BAA, X2B, X3B, TB, OPS)



Main Results

Regression analysis between outcome variables and year showed correlation meaning that offensive stats have been increasing as year increased. After controlling for year, there was a significant effect of field size on 5 of 7 outcome variables. Figure 1 above shows the direct relationship between the outcome variables and the 5 number dimensions ranking including Lowess trend lines. Table 3 depicts the results of the multiple regression analyses between field dimension and outcomes while controlling for year. Because a 1 unit increase in field dimension is not large enough to be informative, we examined the effect that a change of 63 feet (difference between largest and smallest fields studied) has on the 7 outcomes. These results allow for conclusion inferences between long and short fields.

On Base Plus Slugging showed a significant relationship to field dimensions. An addition of 0.023 (0.0058, 0.04) between long and short fields studied accounts for nearly a ¼ more bases per hit. This shows that more extra base hits occur in larger parks. OPS had a range between 0.85 and 0.6.

The most significant variable tested was Earned Runs. Over an 81 game stretch, a 25 (7.37, 42.59) run difference is very substantial. This outcome makes sense because larger fields have more area to be defended allowing for more runs to be scored.

Home Runs did not have a strong correlation with the field dimension which means that less home runs are hit at larger fields.

Table 3: Multiple Linear Regression Summary

	Outcome Variables	β Point Estimate	β Confidence Interval	p-value
BAA	Batting Average Against	(0.256, 0.264) 0.008	(0.0035, 0.0146)	0.0015
ER	Earned Runs Allowed	(707, 732) 25	(7.37, 42.59)	0.00557
X3B	Triples Allowed	(119, 123) 4	(2.07, 6.36)	0.00125
X2B	Doubles Allowed	(151, 157) 6	(-2.09, 12.79)	0.158
HR	Home Runs Allowed	(45, 46) 1	(-4.98, 8.13)	0.637
OPS	On Base % + Slugging	(0.652, 0.675) 0.023	(0.0058, 0.04)	0.00894
TB	Total Bases Allowed	(1373, 1422) 49	(8.63, 88.83)	0.0174

Conclusions

According to the results in Table 3 above, we can be 95% confident that a field 63 feet longer in outfield dimension will on average, produce between 0.0035 and 0.0146 higher Batting Averages. This effect would only produce between 3.5 and 14.6 more hits per 100 at bats which is an insignificant increase to our data which had a standard deviation of 0.02.

We can be 95% confident that on average, the larger field will increase Earned Runs between 7.37 and 42.59. This outcome produced a wide interval compared to a standard deviation of 48.9. This relationship shows the largest practical significance. Scoring between 7 and 42 more runs over an 81 game stretch would highly affect game outcomes. We can be 95% confident that on average, there will be between 2.07 and 6.36 more triples hit at the larger field than the smallest. This was a strong relationship compared to the range of triples hit of 35 (Table 2).

While field dimensions was proved to have a statistical and practical significance on about half of variables tested, more analysis would need to be performed to remove influence from other possible lurking variables.

References

http://www.espn.com/mlb/

http://mlb.mlb.com/assets/images/2/8/6/73652286/cuts/Dimension3_o0vp5p8x_ilya8856.jpg

R Studio version 3.2.2