8-2003

Fair-Weather Fans: The Correlation Between Attendance and Winning Percentage

Darren B. Glass
Gettysburg College

Glass, Darren. Fair-Weather Fans: The Correlation Between Attendance and Winning Percentage. The Baseball Research Journal, 32, 81-84.

Fair-Weather Fans: The Correlation Between Attendance and Winning Percentage

Abstract
In Rob Neyer's chapter on San Francisco in his Big Book of Baseball Lineups, he speculates that there aren't really good baseball cities, and that attendance more closely correlates with winning percentage than with any other factor. He also suggests that a statistically minded person look at this. I took the challenge and have been playing with a lot of data.

Keywords
baseball, fair-weather fans, correlation, attendance, performance

Disciplines
Analysis | Mathematics

DARREN GLASS

Fair-Weather Fans

In Rob Neyer's chapter on San Francisco in his *Big Book of Baseball Lineups*, he speculates that there aren't really good baseball cities, and that attendance more closely correlates with winning percentage than with any other factor. He also suggests that a statistically minded person look at this. I took the challenge and have been playing with a lot of data.

METHODOLOGY

I looked at all seasons from 1973 until 2002. In particular, I looked at the correlation coefficients between the following variables:

- Average home attendance per game (ATT)
- Home attendance per game divided by average
- Home attendance over all teams (to normalize for nationwide trends) (ATT/AVG)
- Final place in divisional standings (PLACE)
- Winning Percentage (WIN)

There are a few basic properties of correlation coefficients (CC's). If a CC is equal to zero, then the two variables are uncorrelated, if it is close to one they are close to linearly correlated in a positive way, and if it is close to -1, then there is a strong negative relationship between them.

CORRELATION WITH WINNING PERCENTAGE

To begin with, let us look at the most naïve study: the correlation between winning percentage and home attendance. Over the 30 years between 1973 and 2002, the baseball-wide CC was .464. Table 1 lists teams that can be described as having fair-weather fans—their correlation between winning and attendance is more than 0.2 greater than the baseball-wide average.

Table 1. Teams with correlation coefficients between ATT and WIN greater than 0.2 above baseball average

Atlanta	0.884
Seattle	0.815
New York N	0.786
Cleveland	0.755
Montreal	0.753
Chicago A	0.752
San Francisco	0.673

On the other side of the spectrum are those teams that have correlation coefficients significantly lower than the baseball-wide average. An optimistic interpretation of this would be that the fans stick with the team no matter how badly they are doing (the case of the Red Sox and the Cubs), while a pessimistic interpretation might be that the fans refuse to support the team no matter how good they are. Table 2 lists cities that have correlation coefficients between ATT and WIN more than 0.1 below baseball average.

Table 2. Cities with correlation coefficients between ATT and WIN more than 0.1 below baseball average

St Louis	0.345
Chicago N	0.321
Texas	0.304
Tampa Bay	0.266
Milwaukee	0.234
Arizona	0.142
Pittsburgh	0.131
Los Angeles	0.117
Boston	0.004
Colorado	-0.087
Florida	-0.118
Baltimore	-0.246

The presence of all four of the expansion teams of the 1990s on this list makes sense, as the small sample size is distorted by the first few years in which novelty value runs high and the teams are not likely to be very good.

The most interesting data point on this list to the author is the Orioles, where the fans of Baltimore over the past 30 years actually supported the team significantly more the worse they have been. This is likely due in large part to the draw of the new ballpark at Camden

DARREN GLASS *experienced the phenomenon of fair-weather fans first hand when he was one of the dozen people to go to Atlanta Braves games in the mid-1980s. He is currently an assistant professor of mathematics at Columbia University.*

Yards, and that it has been successful in bringing in fans despite the fact that the Orioles have had losing records in six of the 11 years since it opened.

A slightly less naïve study would try to normalize for the effects on attendance of baseball as a whole. The average attendance at baseball games has nearly doubled over the last 30 years, and all of baseball took a hit in 1995, when the average attendance dropped nearly 6,000 fans per game. Thus, I also computed the CC's between ATT/AVG, a given team's average home attendance divided by the average attendance of baseball games league-wide, and winning percentage. The data did not qualitatively change significantly. The league-wide CC went up to .55.

Table 3. Correlation coefficients between ATT/AVG and WIN

Atlanta	0.925
Cleveland	0.832
Seattle	0.786
Philadelphia	0.753
New York N	0.752
Cincinnati	0.724
San Francisco	0.713
Oakland	0.692
Detroit	0.691
Kansas City	0.677
Minnesota	0.667
New York A	0.598
Tampa Bay	0.596
San Diego	0.573
Los Angeles	0.563
Montreal	0.557
Chicago A	0.541
Pittsburgh	0.539
Boston	0.532
Chicago N	0.520
Houston	0.505
Texas	0.489
St Louis	0.485
Toronto	0.478
Milwaukee	0.433
Anaheim	0.387
Colorado	0.303
Arizona	0.079
Florida	-0.035
Baltimore	-0.092

Statisticians say that a correlation coefficient is statistically significant if it is greater than the value of a certain T-test. While I will not go into the details of this calculation, I will point out that for our sample size of 802 team-seasons, any CC over .116 is statistically significant with probability 99.9%. In particular, our league-wide CC of .55 is extremely significant.

For the individual teams, sample sizes are much smaller. In particular, non-expansion teams have 30 data points, and thus a CC over .570 will be statistically significant 99.9% of the time, a CC over .463 is significant 99% of the time, and a CC over .361 is significant 95% of the time. When expansion teams with even smaller sample sizes are included, the CC's are significant at the 99% level for every team except Milwaukee, Anaheim, Baltimore, Toronto, Tampa Bay, Arizona, Colorado, and Florida.

Of course, the CC is not enough to capture what we are interested in. In particular, if a city's ATT/AVE and WIN were strongly correlated to a line with slope zero, we would view it as much less of a "fair-weather fan" city than a city with a weaker correlation to a line and a very large slope. I also computed the slope of the line given by various linear regressions baseball-wide—the results of a linear regression on ATT/AVG and WIN are ATT/AVG = 2.7525 × (WIN) − .3769. While ATT/AVG is a more meaningful statistic, it is also harder to get a feel for. For this reason we will note that the linear regression between ATT and WIN gives ATT = 63,476 × WIN − 7,710. In other words, by increasing winning percentage by .100 (an improvement of roughly 16 wins per season), a team can expect to boost home attendance by an average of 6,347 fans per game.

Table 4. Slopes from linear regressions between ATT/AVG and WIN

Cleveland	4.543672
Philadelphia	4.290944
Atlanta	3.850382
Cincinnati	3.735552
Los Angeles	3.431718
Seattle	3.328853
San Francisco	3.206009
New York N	3.134074
Kansas City	3.067628
Minnesota	2.862508
Montreal	2.772461
Oakland	2.403002
Chicago A	2.214931
New York A	2.202218
Detroit	2.186652
Houston	2.157452
Toronto	2.114608
Boston	1.920404
Anaheim	1.917665
Colorado	1.888440
San Diego	1.858157
Texas	1.775284
St Louis	1.746337
Chicago N	1.634861
Tampa Bay	1.578699
Pittsburgh	1.374932
Milwaukee	1.304664
Florida	-0.15230
Baltimore	-0.37538
Arizona	-0.99382

A natural question to ask, and one that more than a few people are looking at due to its various political implications, is how new stadiums affect attendance. While I did not investigate this phenomenon in any depth, I will note that if you remove all data points in the data set corresponding to the first two years that a team is in a new city or a new stadium, the baseball-wide CC actually raises by .05.

CORRELATION WITH PLACE FINISHED

It is also natural to wonder if it is not the winning percentage that brings in the fans but being in the hunt of a pennant race. I decided to test this hypothesis by calculating the correlation coefficients between our attendance variables and the place in which a team finished within their division, as well as how many games back they finished. Because the nature of both of these variables changed significantly with the realignment in 1994, I ran the study first looking only at the data from the years 1973-1993. In particular, it was not clear how to best handle the situation with the wild card, and teams that might be in the hunt for the wild card despite being many games out of the division lead (see 2003 Phillies and Marlins, for example). It came as a surprise to the author that including the last decade did not significantly change the results, as seen by the following charts:

1973 to 1993	CC	SLOPE
ATT/AVE and PLACE	-0.5590	-0.1050
ATT and PLACE	-0.4632	-2136.5000
ATT/AVE and GB	-0.5300	-0.0164
ATT and GB	-0.4535	-343.1290

1973 to 2002	CC	SLOPE
ATT/AVE and PLACE	-0.5590	-0.0978
ATT and PLACE	-0.5016	-2491.0100
ATT/AVE and GB	-0.4906	-0.0145
ATT and GB	-0.4131	-334.6898

In all of these examples, CC is negative. This is what we would expect as the "higher" your value of PLACE and GB, the less attendance we might expect to see.

I have not included the team-by-team data, but it is qualitatively very similar to the above team-by-team data, with the teams falling in roughly the same order and with the same significance results. Anyone who is interested in the full data should feel encouraged to email me.

CORRELATION WITH PAST PERFORMANCE

Another question that comes up is how correlated attendance is with past performance. In particular, looking at the correlation between winning percentage (or standings) in year x and attendance in year (x+1). The idea being that the rush of winning the World Series creates new fans (and season ticket holders) no matter how badly the team performs the following year.

However, when one runs the numbers, they are not particularly illuminating. In fact, the CC's one gets from comparing last year's winning percentage and this years ATT/AVG is .492, slightly less than when you compare this year's record with this year's attendance, .551. (See below for the full chart of CC's.) Furthermore, the only teams for which there is a substantial difference in the CC's when you run the study the two ways are Colorado (which can be partially explained by the fact that you had a small data set to begin with and are reducing it even further), Minnesota, Montreal, Pittsburgh, and St Louis. Furthermore, in each of these cases there is a weaker correlation. So while my instincts agreed with what many of you suggested might be an interesting effect, the numbers don't seem to bear it out.

	WIN	PREV WIN	PLACE	PREV PLACE
ATT/AVG	0.5505	0.4926	-0.5016	-0.4651
ATT	0.464	0.4293	-0.4669	-0.4329

One problem in trying to do such a study is that there is a relatively strong correlation between how a team does in year X and how it does in year x+1 (CC = .5 for my data set). Isolating that factor would be hard but not impossible.

CONCLUSIONS

Every one of the tests which I ran seems to indicate that Rob Neyer's hypothesis is correct: attendance at ball games is highly correlated with the winning percentage of the home team. This is certainly true baseball-wide, and is also true for almost every team individually. The exceptions by and large are the expansion teams of the 1990s and the Baltimore Orioles. Furthermore, in almost every permutation of the data, it seems that the fans of Cleveland, Atlanta, and Seattle are especially prone to support their teams more the better they do. We do note, however, that all three of these teams got

new stadiums while the teams were doing especially well—and in the case of the Braves and the Indians this was also at a time when baseball was seeing a drop in attendance nationwide—which likely skews the data somewhat.

FURTHER EXPLORATIONS

I think it would be very interesting to look at attendance in smaller units than seasons. This could take away some of this effect by looking at when in (for example) the 1991 season the fans stopped punishing the Braves and Twins for previous subpar performance and rewarded them for being good.

However, to do this one would have to control for factors such as weekend games (which generally have higher attendance) or superstar players coming through town (which certainly boosts attendance) or the like, factors which one can ignore over the course of a season but which could significantly affect the data when looking at units of individual games or weeks or even months.

Another thing that I would like to do is to try to adjust for ballpark size. The only way I could think of to do this would be to use "percentage of seats filled" as my attendance variable, but this seems to pose more problems than it solves. I certainly like the idea of "rewarding" the Cubs and Red Sox and other teams which could sell more seats if they had the capacity, but I'm not sure if it makes sense to "punish" cities for having large stadia in this way. For example, if Stadium One holds 50,000 people and Stadium Two holds 60,000, I do not think that it makes sense to treat the fact that they both draw 30,000 fans differently. It also seems like a bit of opening Pandora's box as we really don't know how many fans the Red Sox would average if they had an infinitely big stadium. It could be that their attendance would stay the same or it could be that it would quadruple—we have no real way of knowing.

REFERENCES

All data came from www.baseball-reference.com

CHICAGO CUBS NEWS, AUGUST 6, 1937

Cubs fans who visited Wrigley Field on July 4 and 5, really went to town with the ol' feed bag.

During the two days, on which doubleheaders were played, the Wrigley Field concession stands broke all existing records in the dispensing of food and beverage.

Three tons of hot dogs were consumed; 11,000 ham, cheese and hot roast beef sandwiches; 19,200 bars of candy; 16,000 sacks of peanuts; 11,280 bags of popcorn; 600 loaves of bread (two pounds to a loaf); 18,000 packages of ice cream; 19,200 cans of beer; 18 barrels of beer; 6,000 lemonades and 41,000 bottles of pop.

Wrigley Field is becoming one of Chicago's most popular eating places when the Cubs are in town. Many fans attending the games plan to have their lunch at the park, where there is a wide variety of food selections."