By the Numbers

Volume 20, Number 1 The Newsletter of the SABR Statistical Analysis Committee February, 2010

Comment

Academic Research: Conference Papers

Charlie Pavitt

The author reviews several sabermetric studies appearing at a 2008 sports statistics symposium.

Ben S. Baumer, Using Simulation to Estimate the Impact of Baserunning Ability in Baseball

Nicholas S. Miceli and Alan D. Huber, "If the Team Doesn't Win, Nobody Wins": A Team-Level Analysis of Pay and Performance Relationships in Major League Baseball

Brad Null, Modeling Baseball Player Ability with a Nested Dirichlet Distribution

David M. Rockoff and Philip A. Yates, Chasing DiMaggio: Streaks in Simulated Seasons Using Non-Constant At-Bats

Jesse Weinstein-Gould, Keeping the Hitter Off Balance: Mixed Strategies in Baseball

All of these studies were presented at the 2008 *Northern California Symposium on Statistics and Operations Research in Sports,* and published in Volume 5 Issue 2 of the *Journal of Quantitative Analysis in Sports*. I have been critical of this journal for publishing too many exercises in statistical sophistication and very few papers with substantive significance. Most of this bunch of papers rises above that level, although none make any critical contributions. The most substantively interesting in my view is Weinstein-Gould, who, using commercially-available data from the 2002 season, examined the degree of variation on selection of first pitch and whether it had an impact on batter performance. Pitch type options were regular, split-finger, and cut fastball, curve, change-up, slider, sinker, knuckleball, and screwball. Both across batters and for the same batter at different plate appearances, 98 percent of the pitchers varied in their first pitch selection; the only exception were pitchers who relied on knuckleballs (not surprising) and split-fingers (to me, surprising). Significantly, first-pitch choice had no effect on batter performance as measured by on-base percentage and weighted on-base-average (a linear weights type formula; see Tango, Lichtman and Dolphin's *The Book*); the author wisely cautions us not to generalize that finding to pitch selection over the course of an at bat.

In this issue

Ben Baumer, who I assume will be reading this, attempted to estimate the impact of base running in all its facets (stolen bases, taking extra bases on balls in play, tagging up on flies, beating throws to first on attempted double plays) on run scoring. Ben compared actual run production from 2005 through 2007 with what each team would have scored if the team had matched average tendencies to both attempt extra bases when the situation allowed for it and, given the attempt, be successful at attaining it. The results provide us with an estimate of how many runs teams gain or lose from baserunning relative to the average. This is different than the usual method of estimating the impact of a strategic move relative to no attempt at all, and, to be honest, is for that reason less informative concerning the practical value of these strategies. Nonetheless, his findings reflect those of other studies implying the minor contribution of baserunning as compared to hitting, pitching, and fielding; compared to one another, the most successful and unsuccessful teams may differ by about 40 runs (ie, 4 wins) a year through base running.

Turning to the relative importance of those four aspects, as I mentioned in my review of Lewis, Lock, and Sexton in the May, 2009 issue of BTN, academic research attempting to estimate it has been almost uniformly terrible. Lewis et al.'s work was somewhat better. Miceli and Huber is another try, and, although their effort is better than most, it is also fundamentally faulty. Rather than haphazardly choose performance indices that are often poor measures of what they are intended to represent, the authors attempted a principled empirical approach, via subjecting a slew of team-level hitting and pitching indices from 1985 to 2001 (downloaded from www.baseball1.com) to factor analyses. Both resulted in three-factor solutions (for hitters, ability to make contact, power hitting, and ballpark factor; for pitchers, power pitching, starter endurance, and walks allowed) accounting for a bit more than 60% of variance. All this implies is that a few factors summarize the data that happens to be on that website, not that these are necessarily the factors that actually determine run scoring and prevention. Miceli used stolen bases to represent baserunning, as caught stealing did not increase accounted-for variance. Their biggest mistake was in trusting the almost quarter-of-a-century-old-and-now-outdated claim from The Hidden Game of Baseball that fielding has little impact and so ignored it. They then regressed winning percentage on these variables. "Pitching" accounted for 60% of the variance, batting 28%, and steals 3%. The former really measures defense in general; an educated guess is that pitching would actually be 40% and fielding 20% if the authors had used better data.

Computing the odds of a 56-game hitting streak has been a popular research effort in the statistical literature, attacked many times from several different directions. Some of the relevant studies have made the simplifying assumption that a batter gets four at bats per game; these efforts cannot be taken seriously. For this reason alone, Rockoff and Yates (2008) did quite a bit better than these two attempts. Their idea was to simulate 1000 seasons of play using actual seasonal game-to-game performance for each of 58 years of Retrosheet data. Out of the 58,000 simulated seasons, a total of 30 (about .005%) included a hitting streak of 56 or more games. Interestingly, Ichiro's 2004 season included 5 of them. Using this data, the authors concluded that the odds of a streak of more than 56 games in any of the 58 seasons in the data set were about 2½ percent.

Null's work is of methodological interest only.

The papers reviewed here can be downloaded at http://www.bepress.com/jqas/vol5/iss2/ .

Charlie Pavitt, chazzq@UDel.Edu ♦

Do Motivated Baseball Players Have Higher Batting Averages?

Paul Scott and Phil Birnbaum

A recent academic study suggested that players who are hitting close to.300 late in the season subsequently hit much better than expected, as a result of expending extra effort in an effort to reach a personal goal. Here, the authors argue that hitters do not, in fact, change their performance level, and that the apparent improvement is due to sampling issues.

Pope and Simonsohn (2010) document that "professional batters are nearly 4 times as likely to end the season with a .300 batting average as they are to end the season with a .299 average. " An article in the New York Times surmised "hitters at .299 or .300 batted a whopping .463 in that final at-bat, demonstrating a motivation to succeed well beyond normal." We argue that this disparity is produced by sampling bias: near the end of the season, many players quit immediately after getting the hit that pushes them over .300. While Pope and Simonson do not claim to find evidence of increased performance, they do not rule out the possibility. They show that hitters in their sample are frequently replaced by pinch hitters; we document that pinch running and sitting out of games are also prominent. Once all these sources of bias are removed, the apparent performance disparity disappears. Moreover, tests with unbiased sampling criteria show nothing unusual about the performance of players on the cusp of .300.

Example

Here we illustrate the source of bias with a simple example.

Suppose every player hits with a .300 average independently across at-bats (ABs) in a season of at most twelve ABs. For simplicity, there is no walking--players either make a hit or an out in each AB. After each of his first eleven ABs, a player may decide to keep going or quit early. Players usually want as many ABs as possible, but each player who hits .300 or higher in at least ten ABs will receive a big bonus.

Each player will take the first ten ABs since he can't get the bonus otherwise. After the tenth AB, players with three hits (a .300 average) will stop rather than risk losing the bonus. No other players stand to lose the bonus in the final two ABs, so the rest will continue through 12 ABs.

If we select final ABs in which players stand to cross the .300 mark, about 23.8% of the sample would consist of players who were 2-for-9 going into their 10th AB and then stopped (these batters had a hit for sure). The other 76.2% would consist of players who were 3-for-11 going into their 12th AB (these batters hit at a .300 rate). Even though players hit with probability .3 in each AB, we can expect to observe a .466 batting average in the sample.[1] The problem is that we sample only the ABs following 2-for-9 starts in which there was a hit, omitting the outs because those hitters continue to twelve ABs.

Thus, if a plate appearance's outcome can affect whether it is a player's last, then selecting final plate appearances will lead to a biased sample.

A biased sample

Using data from Retrosheet.org for the 1975-2008 regular seasons of Major League Baseball, we attempt to replicate Pope and Simonsohn's sample by selecting plate appearances (PAs) satisfying the following criteria by year:

- The PA is the batter's last.
- The batter has at least 200 ABs.

---

[1] This comes from the fact that P(hit in last AB | last AB in sample) =

(P(3rd hit in 10th AB) + P(4th hit in 12th AB)) / (P(3rd hit in 10th AB) + P(3 hits after 11th AB)) = 0.4663

- The batter has a batting average below .2995 going into the PA.
- A hit in the PA would make the batter's average at least .2995.
- The date is September 25 or later.

We use the cutoff .2995 rather than .3000 because recorded batting averages are rounded to three decimal places.

There are 121 PAs in our sample, including 57 hits in 116 ABs for a batting average of .491, similar to Pope and Simonsohn's observations.

We consider three ways in which a batter's season can be stopped early to preserve a .300 average:

- A pinch runner replaces him immediately (13 occurrences in our sample),
- A pinch hitter replaces him the next time his position in the batting order is reached (13 occurrences),
- He sits out his team's remaining games (34 occurrences, 24 distinct from 1 and 2).

Following 50 out of 121 PAs, the batter's season was apparently cut short. Removing these observations drops the sample's batting average from .491 to .299 (20 for 67), matching the batters' earlier performance exactly.

Thus, there is strong evidence that the high batting average is a result of sampling bias alone.

An unbiased sample

To implement Pope and Simonsohn's tests with an unbiased sample, we simply choose PAs based on criteria that are known before the PA's outcome.

We sample all PAs in a team's last two games of the season where the batter was hitting less than .300, but a hit in the current PA would push him over. These highly motivated players hit a combined .299 (155 for 517), almost exactly their season average.

Then, still looking only at teams' last two games, we sample PAs where batters were already at .300 or higher, but would drop below .300 if they made an out. Those players also should have been highly motivated, but their average was .297 (43 for 145), again closely matching their earlier performance.

While hitters in our sample do not increase their batting averages, it might be argued that they did hit better than expected. A batting average of .300 is substantially above average for the league (typically around .260). Therefore, it would be expected that players in our samples would hit less than .300, regressing to the mean somewhat. They regressed only slightly.

However, even for the larger sample of 517 AB, the standard error of observed batting average is about .020. Since the mean batting average is within two standard deviations of .300, the lack of regression to the mean is not significant evidence of increased performance, for players would not be expected to regress fully to the mean.

We find no statistically significant increase in performance for players around the .300 mark, and conclude that sampling bias fully accounts for Pope and Simonsohn's findings.

References

Pope, D., & Simonsohn, M. (2010). Round Numbers as Goals: Evidence from Baseball, SAT Takers, and the Lab. *Psychological Science*, forthcoming. Ungated version found at http://opim.wharton.upenn.edu/~uws/papers/round_inpress.pdf .

Schwarz, A. (2010, October 2). Sniffing .300, Hitters Hunker Down on Last Chances. *The New York Times*. Retrieved from http://www.nytimes.com/2010/10/03/sports/baseball/03hitters.html .

Paul Scott, ptscott@princeton.edu , Phil Birnbaum, birnbaum@sympatico.ca ♦

Long-Term Trends in Major League Baseball

Jim Grant

With charts and commentary, the author takes us through trends in run scoring, batting average, and fielding average over 130+ years of baseball history.

This article offers some long term perspective on what has been happening to some common baseball statistics, observations, and some questions that may prompt some response or additional analyses on the long term.

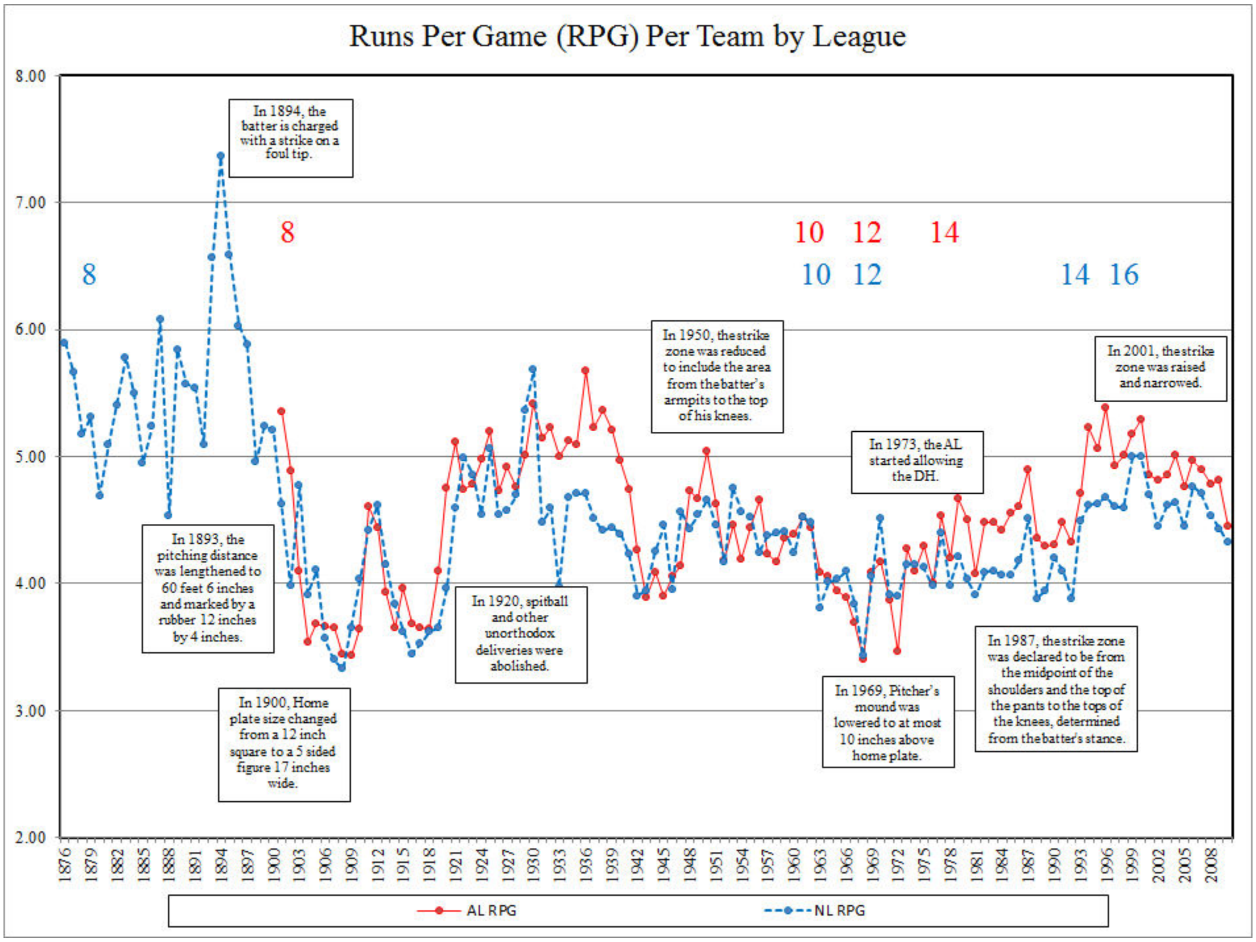
The data that I have used are league summaries by year for the American League (AL) and the National League (NL) since their inception in 1901 and 1876, respectively. The boxes on the charts provide a brief description of key changes in the rules at various times over the last 130+ years.

Runs Per Game (RPG) Per Team

Winning in baseball is achieved by scoring runs. Hence, it seems like runs scored would give us the best indication of what's been happening with the game. Since the number of teams and games per year has changed several times over the years, I've used Runs Per Game (RPG) per team as the subject statistic.

The chart on the next page shows the RPG by year for both leagues. The number of teams in each league at the time the number changed is also shown.

In each chart, the AL is in red and the NL in blue (no political connection is implied). For those reading in monochrome format, the AL is the solid line, and the NL is the dashed line.



Observations

1. The range over the last 130+ years is wide, peaking in 1894 at a level (7.36) more than 100% higher than the low point in the late 1960s (around 3.50). Even in the last 45 years, it has swung from a low of about 3.50 to a high over 5.00.

2. It takes about a 1.00 swing (up or down) in the RPG (which commonly takes over 10 years) before MLB changes something to reverse the trend.

3. Rules changes had a dramatic effect on RPG:

a) In 1893, the pitching distance was increased to the current length of 60' 6". (Previously, it was 50'.) In the two years after that, the RPG rose 44% from 5.10 to 7.36.

b) In 1894, a foul tip was counted as a strike and RPG dropped back down over the next six years to just above 5.00.

c) Apparently not being satisfied that that was low enough, they increased the width of the plate and the RPG dropped to about 3.50 by 1908.

d) In 1920 and 1921, spitballs and other unorthodox pitches were abolished. This raised the RPG another 1.00 and got it above 5.00 again.

e) In 1950, the strike zone was reduced to the area from the batter's armpits to the top of his knees. While is seems logical that this would have increased RPG, instead, it dropped about 1.40 over the next 18 years.

f) Presumably in order to correct for this, MLB owner/leaders lowered the height of the mound to at most 10" above home plate, reversing the declining trend.

g) In 1973, the AL introduced the designated hitter. The average difference between the NL's RPG and the AL's RPG for the 10 years after the introduction was about 0.30, which is probably a good indication of the impact of the designated hitter. This appears to be the impact even to this day.

h) In 1987, the strike zone was changed again. It was declared to be from the midpoint of the shoulders and the top of the pants to the tops of the knees, determined from the batter's stance. Following that change, the RPG rose by about 1.00 over the next 10 years.

i) In 2001, probably in order to standardize a strike zone between the two leagues and to reverse the impact of the "steroid area", the strike zone was raised and narrowed. This has perhaps caused a minor decline (0.20 to 0.30) in the RPG.

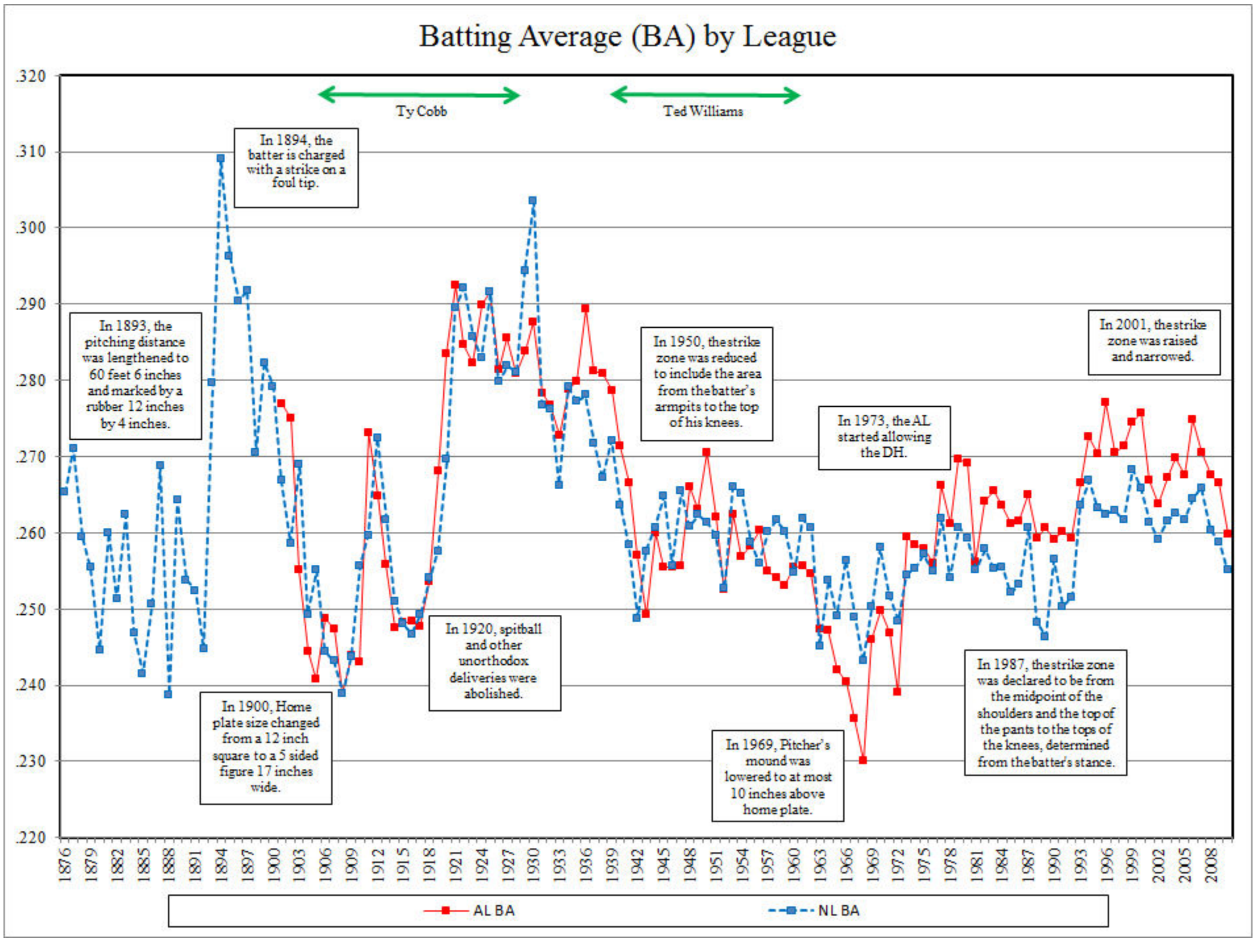
4. It is intriguing that after a change in rules which ends a trend (either up or down), an opposite trend (down or up) occurs that lasts for more than 10 years. (In other words, over the last 130+ years, there were no extended periods when RPG was essentially flat.) It is as though it takes the players several years to adjust to a rules change and they continue adjusting, until the MLB owners/leaders decide to reverse the trend.

5. From the late 1920s through the early 1940s, RPG steadily dropped over 1.00. I have not found an explanation for this. One could try to make the case that during the depression that some great ballplayers had to leave the game to support themselves and their families. However, until demonstrated otherwise, it seems to be reasonable to presume that as many great pitchers left the game as great batters and the depression wouldn't have reduced the RPG.

6. Some people feel that league expansion and the associate dilution of talent had a negative impact on the game. The issue is whether it can be seen in common statistics. On first thought, a reasonable presumption would appear to be that whatever impact the expansion had on offense, it would have also impacted defense. Visually examining the RPG chart reveals that sometimes the RPG increased when expansion occurred and sometimes it declined. Hence, whatever impact expansion had, it was either negligible or more than offset by other factors.

Batting Average (BA)

The chart below shows the BA for the AL and NL.



Observations

1. Again, over the last 130+ years, the range is large, peaking at almost .310 in 1894 and bottoming at about .230 in 1968.

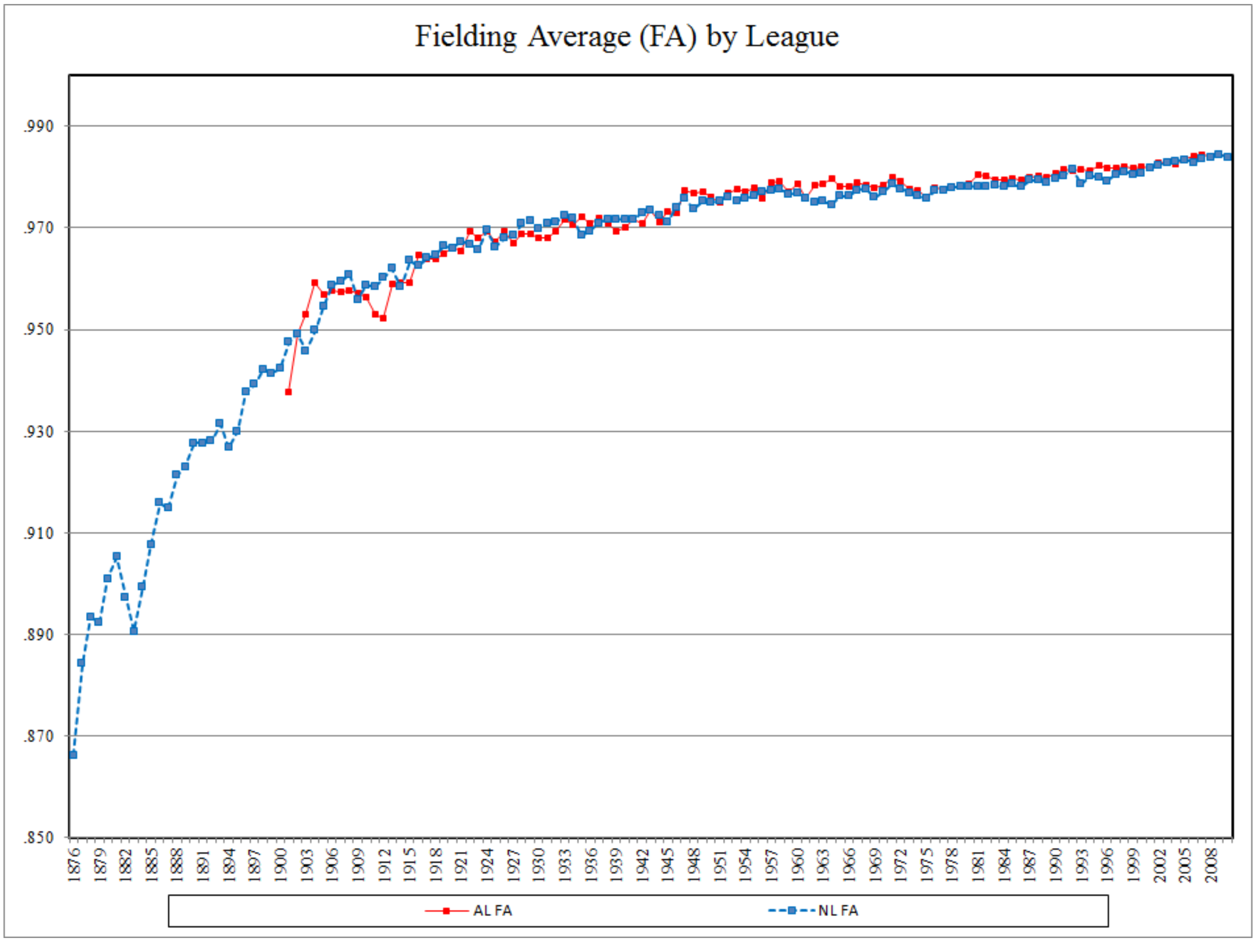
2. The rules changes had a dramatic effect on BA.

3. The BA line tracks the RPG line extremely closely, indicating the high correlation between the two statistics.

4. The impact of the designated hitter on the AL BA was on the order of 13 points, when considering the average difference between the AL batting average and the NL batting average for the 10 years before and the 10 years after the introduction.

Fielding Average (FA)

The chart below shows the trend in fielding average for the AL and NL.



Observations

1. Now we see something entirely different. Fielding performance was increasing dramatically up until the mid-1920s, no doubt due to the improvement in gloves and the tactics of changing defensive positioning, based on who was batting.

2. Since the mid-1920s, FA has been gradually increasing.

3. Since about the mid-1950s, the American League FA has been slightly better than the National League FA. Although small, it has been consistent. In the last 10 years, the gap has been narrowing.

Jim Grant, jwgrant@aol.com ♦

"By the Numbers" mailing list

SABR members who have joined the Statistical Analysis Committee will receive e-mail notification of new issues of BTN, as well as other news concerning this publication.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, preferably by e-mail. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to Phil Birnbaum, at birnbaum@sympatico.ca .