By the Numbers

Volume 15, Number 3 | The Newsletter of the SABR Statistical Analysis Committee | August, 2005

Summary

Academic Research: Batting Last and Seeing the Ball

Charlie Pavitt

The author reviews two recent academic studies – one on the advantage of batting last, and the other on whether successful hitters actually see the ball as being bigger.

This is one of a series of reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at its new location *www.udel.edu/communication/pavitt/biblioexplan.htm* . Use it for your research, and let me know what is missing.

Steven R. Bray, Jeff Obara, and Matt Kwan, Batting Last as a Home Advantage Factor in Men's NCAA Tournament Baseball, Journal of Sports Sciences, Volume 23 Number 7, July 2005, pp. 681-686

This is another stab at determining the factors behind home field advantage. Along with crowd influence, travel, and familiarity with home field, it has been proposed that the very fact of batting last can lead to an advantage to the home team. It follows that a test of this proposal might be possible if a context exists in which the team batting last is not characteristically the home team. Bray, Obara, and Kwan examine such a context: NCAA tournament baseball, in which teams often play one another at a neutral site, and where the home team does not necessarily come to bat last. Using data from four-team tournaments at Divisions I, II, and III from 1999 through 2003, Bray et al. first found that neutral teams won exactly half the time batting both first and last. In addition, they noted that that home teams actually won more often batting first (67.3%) than last (59.7%); it follows that visiting teams did so also (40.3% versus 32.7%). Thus, there is no support for the hypothesis that batting last in and of itself is advantageous. One additional finding relevant to a slightly different issue was that home teams did not have a higher probability of winning a tied game or staging a come-from-behind victory in the bottom of the ninth inning than visiting teams.

In this issue

Jessica K. Witt and Dennis R. Profitt, See the Ball, Hit the Ball: Apparent Ball Size Is Correlated With Batting Average, Psychological Science, Volume 16 Issue 12, December 2005, p. 937

In my book, this one is a lark on the surface but with some significant implications. Players often say that the ball looks bigger to them when they are getting hits than when they are not, so Witt and Profitt decided to see if this were true. They recruited 47 softball players from Charlottesville, Virginia who had just completed a game or two, showed them a poster displaying eight black circles ranging in size from 9 to 11.8 centimeters, and asked them to select the circle that best corresponded to the size of a softball. Finally, the players reported their at bats and hits for the day. There was a .27 rankorder correlation between batting average and judged size of the ball. So, indeed, the ball tended to look bigger for the batters who had gotten hits than for those who had not.

I am pretty convinced that what we have here is another example of the deep-seated human need to make sense of the world, and attempt explanations for the often-inexplicable. Although we will never know for sure, our research evidence strongly suggests that batting streaks and slumps are probably the result of random processes, but we as ballplayers find it hard to accept that, and so strive to come up with a reason. Attribution theories in social psychology describe how we often perform some action and then come to accept an explanation for the action after the fact. "I got a lot of hits today, because I was seeing the ball well." Witt and Profitt note that causality cannot be determined for this issue without measuring the perceived size of the ball before the games, but this misses the point that better hitters might judge the ball as bigger in the first place. In my book, we must consider the whole thing to be an optical illusion until someone, somehow, can demonstrate that streaks and slumps are "real."

Charlie Pavitt, 812 Carter Road, Rockville, MD, 20852, chazzq@udel.edu ♦

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, either by e-mail or on CD. 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).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

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
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca

The Declining Influence of Good Pitchers

Bill Deane

Why is there more offense these days than in the past? Many theories have been suggested, but one obvious one is seldom mentioned – that offense is up because teams have been giving their best pitchers fewer and fewer innings. Here, the author shows evidence for a substantial change in the way managers assign their staff workload.

Many recent BTN submissions leave me scratching my head and skipping to the conclusions. Maybe that's a reflection on me – but I've been a member of SABR and its Statistical Analysis Committee since 1982, and I got a 750 on my math SATs, so I suspect that if I don't get it, there are probably many others who don't, either. So, at the risk of lowering BTN's intelligence quotient, and earning the snickers of my colleagues, I'll attempt to reach a broader audience – and encourage others to do the same – with some simple, back-of-the-envelope studies.

I'll start with something I did recently on pitcher usage. There are a lot of theories – some provably false – on why there is more offense in today's game than in yesteryear's. One I don't hear mentioned very often is the fact that more and more innings are shifted from a team's best pitchers to its worst ones.

To illustrate this, I went through Neft & Cohen's *Sports Encyclopedia Baseball*. I compared the seasons of 1962 and 1999 – the former because it was the first year that all teams played 162 games, and the latter simply because it was the latest in my edition of the book.

For each team in each year, I ranked pitchers by innings pitched. Then, I averaged the number of frames tossed by the #1 pitcher on each team, by the #2 pitcher, and so on. Following are the results:

Pitcher	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Others	Total
1962	252	220	197	157	133	108	91	77	64	51	100	1450
1999	216	192	173	145	112	92	79	73	66	58	234	1440
Difference	-36	-28	-24	-12	-21	-16	-12	-4	+2	+7	+134	

In other words, the average staff ace in 1962 hurled 252 innings; by 1999, the number was down to 216, a decrease of 36 innings. That pattern holds through the top eight pitchers on the staff: each one's frames were down in 1999 in comparison with '62, totaling 153 innings pared from the best eight pitchers on each staff, to be divided between its worst ones. Isn't that going to help the hitters?

Some random observations:

- One might guess that a big part of the shift would have been the move from four-man to five-man rotations. But the percentage drop in innings of the #5 pitchers is actually greater than that of any of the other spots!
- The average staff pitched ten more innings in 1962 than in '99. My first guess was that there were more tie games in 1962, but in fact there were only three ties all that year (one in '99), so that's not it. In any case, it's not enough to skew the data much.
- In 1999, the Royals had only four pitchers with more than 75 innings, and the Brewers had only three who worked more than 93.

Bill Deane, PO Box 47, Fly Creek, NY, 13337, 607-547-5786, DizDeane@USAdatanet.net ♦

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis - that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any - I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics - confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
David Kaplan	dkaplan@UDel.Edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
Duke Rankin	RankinD@montevallo.edu	Statistics
John Stryker	johns@mcfeely.interaccess.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

The Pythagorean Projection and the Standard Deviation of Runs

Ray Ciccolella

Sometimes, the Pythagorean Projection gives a less accurate estimate of winning percentage. Here, the author shows that the discrepancy is related to the standard deviation of run scoring – a team with a high SD will fail to live up to its projection, while a team with a low SD will overshoot.

Bill James' Pythagorean Formula is a cornerstone of sabermetric analysis. It has been utilized, analyzed, and validated numerous times. In this article I examine one reason why actual wins may differ from those predicted by this useful tool.

If a team has a significantly different winning percentage than predicted by the Pythagorean Formula, it essentially means that the team's distribution of runs scored and allowed relative to their average per game figures differs from the typical distribution. This different distribution of runs leads to a different winning percentage than otherwise expected. For example a good rule of thumb is that it takes 10 runs to change a win into a loss (see "BTN" February 2004). But depending upon the allocation of those 10 runs across various games, the impact could be as much as five extra wins (change five one-run losses into one-run wins) or as few as zero additional victories.

To check for a relationship between a team's distribution of runs scored and allowed and the difference between their actual win total and Pythagorean projection, I performed the steps listed below for each team.

1. Calculate runs scored per game;
2. Find the standard deviation of runs scored per game;
3. Divide the result in Step 1 by the result in Step 2 to calculate the *"Runs Scored Ratio"*;
4. Calculate runs allowed per game;
5. Find the standard deviation of runs allowed per game;
6. Divide the result in Step 4 by the result in Step 5 to calculate the *"Runs Allowed Ratio"*;
7. Calculate the wins predicted by the Pythagorean Formula;
8. Compare the actual wins to predicted wins.

Below is the data for the 2003 Kansas City Royals as an example. The 1.58 for the runs scored ration is 5.16 (runs scored per game) divided by 3.27 (standard deviation of runs scored per game). The 1.42 for runs allowed ratio is 5.35 (runs allowed per game) divided by 3.77 (standard deviation of runs allowed).

Games	Runs Scored (R)	Runs Allowed (RA)	R/game	RA/game	SD of R	SD of RA	Runs Scored Ratio	Runs Allowed Ratio	Wins	Pythagorean Wins	Delta
162	836	867	5.16	5.35	3.27	3.77	1.58	1.42	83	78	5

I used the ratio described above, rather than the standard deviation alone, because there was a reasonable correlation between runs scored or allowed per game and the respective standard deviations. The r-squared values were 0.58 for runs scored and 0.48 for runs allowed. Since I was checking for the relationship between a team's distribution of runs and their averages per game, I believe that the ratio helps reduce the effects of the correlations.

Table 1 – Averages and Total Difference in Actual and Predicted Wins

Predicted Wins vs. Actual Wins	Runs Scored Std Dev	Runs Allowed Std Dev	Runs Scored per Game	Runs Allowed per Game	Runs Scored Ratio	Runs Allowed Ratio
-14	3.24	3.23	4.87	4.87	1.51	1.51

For my data set I used the 1999 to 2003 seasons for a total of 150 teams. Table 1 below shows the averages for these 150 teams. The 150 teams won a cumulative 14 fewer games than predicted (rounding to whole numbers caused the difference from zero), while the average ratio for both runs scored and allowed was 1.51 (4.87/3.24 and 4.87/3.23).

For both the runs scored and allowed ratio I created 3 categories (high, medium, and low) defined as:

```
More than 1.55             = High
Between 1.45 and 1.55      = Medium
Less than 1.45             = Low
```

I then categorized each team as High, Medium, or Low for both their runs scored ratio and runs allowed ratio giving me a 3 by 3 matrix and resulting in a total of 9 groups. For each of these groups I calculated the average per team difference between actual wins and predicted wins with the results shown in Table 2 below.

Table 2 – Actual versus Pythagorean Wins per Team Averages by Runs Scored and Allowed Ratios

		Runs Allowed Ratio (horizontal axis)			
		More than 1.55	1.45 to 1.55	Less than 1.45	Total
Runs Scored Ratio (vertical axis)	More than 1.55	(1.3)	1.3	3.6	1.3
	1.45 to 1.55	(1.4)	(1.0)	0.9	(0.6)
	Less than 1.45	(2.3)	(1.3)	0.4	(1.0)
	Total	(1.6)	(0.4)	1.7	(0.1)

The numbers in brackets indicate winning fewer games than predicted by the Pythagorean Formula. I had 15 to 19 teams in each of the nine groups.

A high runs scored ratio correlated with exceeding the predicted number of wins. For runs allowed, it was a low ratio, rather than high, that correlated with exceeding predicted wins. Teams with the highest runs scored ratio and lowest runs allowed ratio exceeding their predicted wins by an average of 3.6 wins per team. Teams on the other extreme missed their predicted number of wins by an average of 2.3 wins per team.

So what causes these results? I think it is related to how often a team wins when it scores or allows a certain number of runs. As an example, Table 3 shows for the 2004 season how often a team wins when it scores a given number of runs.

Teams that scored only one run went 35-401 for a winning percentage of .080, while teams that scored two runs went 112-476, or .190.

Table 3 – 2004 Season Winning Percentage by Runs Scored

Runs Scored	Occurrences	Wins	Losses	Pct
0	251	0	251	.000
1	436	35	401	.080
2	588	112	476	.190
3	612	203	409	.332
4	650	310	340	.477
5	562	341	221	.607
6	467	324	143	.694
7	394	291	103	.739
8	254	211	43	.831
9	209	189	20	.904
10	161	147	14	.913
11	115	111	4	.965
12+	157	154	3	.981
Total	4856	2428	2428	.500

I found that a high ratio of average runs scored per game relative to the standard deviation was correlated with exceeding the Pythagorean Projection. A simple example relates that finding to the data in Table 3. Assume a team averages 5 runs a game in one of six different combinations as shown below.

```
Combination 1:      Score 5 Runs in 162 games
Combination 2:      Score 4 Runs in  81 games,  6 runs in 81 games
Combination 3:      Score 3 Runs in  81 games,  7 runs in 81 games
Combination 4:      Score 2 Runs in  81 games,  8 runs in 81 games
```

```
Combination 5:      Score 1 Run  in  81 games,  9 runs in 81 games
Combination 6:      Score 0 Runs in  81 games, 10 runs in 81 games
```

Each combination yields a different winning percentage as shown in Table 4 below and utilizing the values in Table 3.

The results here are exactly in line with the data shown in Table 2 -- as the standard deviation decreases, the winning percentage increases, even though the average number of runs scored per game is unchanged. A similar analysis for runs allowed would yield Table 4 in reverse; the team with the highest standard deviation would have the best winning percentage.

Intuitively, these results make sense, because not all runs are created equal. Some runs have more marginal value than others. For example, a team has a winning percentage of .904 when they score nine runs and .913 when they score ten runs, an increase of 0.009 for the one extra run. But adding a run to move a team from three runs scored to four has 16 times the impact, since the winning percentage increases .145 (.477 minus .332).

Summary

Table 4 – Winning Percentage by Combination of Runs Scored

Combination #	Winning Pct in 81 Games	Winning Pct in 81 Games	Overall Winning Pct	Avg Runs Scored per Game	SD	Ratio
1 (5, 5)	.607	.607	.607	5	0	--
2 (4 ,6)	.477	.694	.585	5	1	5.00
3 (3, 7)	.332	.739	.535	5	2	2.50
4 (2, 8)	.190	.831	.511	5	3	1.67
5 (1, 9)	.080	.904	.492	5	4	1.25
6 (0,10)	.000	.913	.457	5	5	1.00

This analysis showed that differences between the actual number of wins and the number predicted by the Pythagorean Formula are correlated with the relationship between the average number of runs scored and allowed relative to the standard deviation in game to game runs scored and allowed.

Specifically, teams that had a high ratio of runs scored per game relative to their standard deviation of runs scored collectively won more games than predicted by the Pythagorean Formula. Teams that had a low ratio of average runs allowed to the standard deviation of runs allowed also tended to out perform their Pythagorean prediction. The opposite of these statements is also true.

These findings, though, are really only one level deep. We can see how the variability in runs scored and allowed relative to the average number of runs scored and allowed impacts the actual versus predicted winning percentage. I do not know, however, the causes of differing variability but investigating these causes might be an interesting area for additional research.

In addition I found no evidence that the difference between a team's actual wins and its Pythagorean projection in one season was correlated with the difference in the following season, which I believe is consistent with other research. For my dataset the r-squared was just 0.0023.

While differences from a team's projected number of wins occur all the time and often are correlated with a team's distribution of runs scored and allowed, it could be that these differences are just the result of normal variation.

Ray Ciccolella, rciccolella@austin.rr.com ♦

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

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 (preferably with your snail mail address for our records) 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.

Receive BTN by Internet Subscription

You can help save SABR some money, and me some time, by downloading your copy of *By the Numbers* from the web. BTN is posted to http://www.philbirnbaum.com in .PDF format, which will print to look exactly like the hard copy issue.

To read the .PDF document, you will need a copy of Adobe Acrobat Reader, which can be downloaded from www.adobe.com.

To get on the electronic subscription list, visit http://members.sabr.org, go to "My SABR," and join the Statistical Analysis Committee. You will then be notified via e-mail when the new issue is available for download.

If you don't have internet access, don't worry – you will always be entitled to receive BTN by mail, as usual.

Do Some Teams Face Tougher Pitching?

Phil Birnbaum

Because not every team faces the same opposition pitchers – one team may face lots of aces, while the other sees fewer of those top-end pitchers – some teams will have an easier time of it than others, and have an advantage in the win column because of that. What is the extent of this effect? Is there a meaningful difference in the caliber of opposition pitching faced?

In the 1986 *Baseball Abstract* (Padres comment), Bill James noted that there is some variation in the quality of starting pitching faced. Bill found that in 1984, the Reds faced 94 "front line" starters (defined as an ERA-qualifying pitcher with a better-than average ERA or W-L record) and only 68 "second line" starters. The Padres, on the other hand, had a ratio of 74 to 88. That is, the Reds faced good starters 20 times more than the Padres, which should have added up to a two-and-a-half game advantage in the win column.

Bill found that 1984 may have been an outlying season: there was a spread of 20 starts in the NL West that year, but only a 6-game spread in 1985.

I thought I would revisit the question in a bit more detail, and take a look at all pitchers, not just starters. How much does the typical team gain or lose from the assortment of opposing pitchers it faces?

The Raw Numbers

To find out, I examined every batter/pitcher matchup from 1960 to 1992 (big hug to Retrosheet). For each season, I computed that season's average Component ERA (CERA) for opposing pitchers on each team, weighted by number of plate appearances.

I used CERA because it's a better indicator of pitcher skill than ERA. CERA is based on what a pitcher gives up to batters, rather than how many runs actually score. The relationship between the batting line and the number of runs depends on timing, or "clutch pitching," which is mostly (or perhaps entirely) random. That is, two pitchers could give up exactly the same number of hits, walks, and so on, but one pitcher may nonetheless have a higher ERA than the other. But, really, they should be considered of equal caliber, which is why CERA is more appropriate.

I made two adjustments to the raw CERA. First, I adjusted for park.

Second, if the pitcher had limited action that year (defined as 50 innings pitched or less), I required his CERA to be between two and six. That is, if a pitcher with only 25 IP had an CERA of 1.65, I made it 2.00. If his CERA was 8.26, I adjusted it down to 6.00. The reason is that for pitchers with extreme CERAs and low IP, their CERA probably doesn't reflect their true ability, and without the adjustment, they would skew the results.

Here are the teams facing the toughest pitchers:

Team	Year	Opponent's CERA	League CERA	Difference
TEX	1972	2.99	3.16	-0.17
SDN	1981	3.31	3.48	-0.17
NYN	1962	3.86	4.03	-0.16
OAK	1979	4.06	4.21	-0.15
HOU	1965	3.35	3.50	-0.15
TEX	1985	3.95	4.09	-0.14
CLE	1985	3.95	4.09	-0.14
TEX	1973	3.83	3.95	-0.13
KC1	1964	3.55	3.68	-0.13
CAL	1980	3.94	4.07	-0.12

In the 33 years studied, the largest disadvantage faced by any team was 0.17 points of CERA. That's about 30 runs, or three wins – a significant disadvantage.

Here are the teams that had it the easiest:

Team	Year	Opponent's CERA	League CERA	Difference
LAN	1975	2.90	3.72	0.18
CIN	1972	3.64	3.48	0.16
NYA	1972	3.30	3.16	0.14
TOR	1990	4.06	3.92	0.14
BAL	1961	4.18	4.05	0.14
LAN	1985	3.70	3.57	0.13
DET	1984	4.14	4.01	0.13
NYN	1988	3.43	3.31	0.13
SLN	1989	3.51	3.39	0.12
SLN	1973	3.83	3.71	0.12

The magnitude of the difference is roughly the same – at the extremes, three wins.

Adjusting for Team Pitching

However, not all of this effect is luck. If you check, you'll find that the top chart contains mostly teams with poor pitching, while the second chart is predominantly comprised of teams with good pitching. That's because teams with a strong staff will, by this method, appear to be lucky, simply because they don't have to face their own pitchers!

The 1975 Dodgers had the best pitching in the league, by far, with a park-adjusted 2.85 CERA. The league average was about 3.70. That's a difference of 0.85. Since the Dodgers saved having to face themselves in one-eleventh of games (there were twelve teams in the NL that year), they faced teams with a CERA of about 0.08 greater than average.

So, of the .18 difference, only .10 is luck – the other .08 is expected, from the fact that the rest of the league's pitching is worse than the Dodgers'.

Adjusting the numbers to take the team's pitching into account gives these two updated charts. These are the new unluckiest teams:

Team	Year	Opponent's CERA	League CERA	Difference
HOU	1965	3.35	3.50	-0.15
TEX	1972	2.99	3.14	-0.15
SDN	1981	3.31	3.45	-0.14
TEX	1985	3.95	4.08	-0.13
BAL	1976	3.43	3.56	-0.12
CAL	1980	3.94	4.05	-0.11
NYN	1962	3.86	3.97	-0.10
MON	1977	3.90	4.01	-0.10
OAK	1979	4.06	4.17	-0.10
CAL	1971	3.47	3.57	-0.10

And the luckiest:

Team	Year	Opponent's CERA	League CERA	Difference
CIN	1972	3.64	3.49	0.15
NYA	1972	3.30	3.15	0.15
CIN	1986	3.81	3.68	0.13
TOR	1990	4.06	3.94	0.12
LAN	1975	3.90	3.79	0.11
SLN	1989	3.51	3.41	0.10
CIN	1970	4.09	4.00	0.10
BOS	1971	3.62	3.52	0.10
SLN	1961	4.11	4.02	0.09
DET	1984	4.14	4.04	0.09

As expected, the effect is now less extreme – the top teams are at .15 instead of .18, which is about 25 runs instead of 30.

But these are the luckiest and unluckiest teams, which gives us perhaps an inflated idea of the effect. It's more informative to look at all teams.

The standard deviation of the differences was .043 – about seven runs per season. In any given season, two-thirds of teams would be within seven runs either way, and about one-and-a-half teams out of 30 would show more than 14 runs either way.

But the effect is even smaller than that, for two additional reasons.

Batting Adjustments

First, the CERA of a team's opposition is affected by the quality of its own hitters. Suppose we divide the 1962 NL pitchers into two groups – the ones who faced the Mets a lot, and the ones who faced the Mets only a little, or none. Even if the two groups of pitchers are identical in talent, the first group will have a lower CERA than the second group. Since the first group is more heavily weighted in the Mets' opposition, the Mets will have faced pitchers with lower CERAs – but, really, the lower CERAs don't reflect better pitchers, just pitchers who faced worse opponents!

That is, even as some teams will face better pitchers than others, some teams will face better *batters* than others. When the 1962 Mets appear to have faced better pitchers, it might be simply that those pitchers simply faced worse batters – namely, the 1962 Mets!

Those Mets scored 0.65 runs per game fewer than the league. Suppose half the pitchers in the league accounted for 75% of the Mets' plate appearances, while the others faced the Mets only 25% of the time.

0.65 runs per game is 104 runs. The first group of pitchers would account for 75% of those 104 runs, or 78 runs. The second group would account for the remainder, or 26 runs.

The first group therefore "saves" 26 runs more than average (and the second group "saves" 26 runs fewer). There were 13,060 innings pitched by non-Mets pitchers in 1962. Under our assumptions, each group pitched half of those, or 6,530 innings. 26 runs off 6,530 innings is .036 per nine innings, so the first group would appear to be better pitchers by .036 runs per game, and the second group would appear to be worse pitchers by that same .036. In reality, though, the groups have the same level of talent – the difference is simply due to having been lucky enough to face the Mets more often.

Since the Mets will have faced the first group 75% of the time, and the second group 25% of the time, they appear to be facing better pitchers by a total of (75% of .036) plus (25% of negative .036), which works out to .018.

That accounts for about 11% of the Mets' observed difference. The Mets, admittedly, are an extreme case, but many of the top and bottom teams are good and bad teams, respectively. Also, it's possible that our 75%/25% assumption was too conservative, in which case the effect would be even stronger.

Regression to the Mean

Aside from the adjustment we made for pitchers with few innings, we took the pitchers' CERA as a true measure of their ability. That is, if one team faced an average pitcher, but the other team faced a pitcher who was 1.00 runs worse, we assumed that the latter team would gain the equivalent of 1 run for each 27 they faced that pitcher.

That's probably not true. The farther from the mean a pitcher's performance, the more likely some of the difference was caused by luck rather than skill. That is, if you found all pitchers with ERAs above 5 in (say) 1984, you'd find that, on average, their ERAs were lower in 1983 and 1985. This suggests that their ERA in 1984 was partially caused by bad luck – and, that, therefore, when teams faced that pitcher in 1984, they were really facing a more average pitcher than the analysis assumes.

That is, the *talent* of the pitchers facing a team is probably closer to average than the *statistics* of the pitchers facing that team. This means the amount of luck in pitchers faced is less than our analysis suggests.

Conclusions

Taking all this into account, we find that the adjusted standard deviation is only seven runs, and that, for the two reasons above, this is probably an overestimate. It therefore seems reasonable to conclude that the effect of this kind of luck on a team's record is probably fairly small.

Bill's Study

So this study finds that differences between teams are minimal. Bill James' 1986 study suggested that they were more substantial. Why the difference? I believe there are at least two factors.

First, this study used Component ERA, while Bill's used actual ERA. Since ERA has a larger deviation than Component ERA (due to luck), this would amplify any results.

And second, Bill considered any pitcher with more wins than losses to be a front-line starter. Again, since there is a fair bit of luck in W-L records, this would increase any observed effect.

Phil Birnbaum, birnbaum@sympatico.ca ♦